

Valuation Engineer Journal

A engenharia do valor: medida, disciplinada, reproduzível.

Volume 1 · Número 1 · Verão de 2026



NESTE NÚMERO

A lacuna entre doutrina e prática na avaliação de imóveis
Residual Constraint Approach (RCA): estrutura e protocolo
earthUI, glmnetUI & mgcvUI: interfaces abertas de modelagem

FUNDAMENTOS (F0–F6) Bem heterogêneo · Espaço de características ·
Função de preços hedônicos · Preço implícito · Regressão hedônica · Variável latente

DECISÕES JUDICIAIS RECENTES Estados Unidos · União Europeia · América Latina

Valuation Engineer Journal

Volume 1 · Número 1

Verão de 2026

ISSN 3143-5033

Editor-chefe William Bert Craytor, SRA

Sumário

Missão	2
Editorial	3
Boas-vindas ao número inaugural	4
Cartas ao editor	6
Um convite à correspondência	7
Artigos	9
A lacuna entre doutrina e prática na avaliação de imóveis	10
Residual Constraint Approach (RCA): estrutura e protocolo	42
earthUI: uma interface interativa para o pacote earth (MARS) . . .	103
glmnetUI: uma interface interativa para o pacote glmnet (lasso e elastic net)	150
mgcvUI: uma interface interativa para o pacote mgcv (modelos aditivos generalizados)	188
Fundamentos	229
Visão geral (F0)	230
Bem heterogêneo (F1)	236
Espaço de características (F2)	244
Função de preços hedônicos (F3)	251
Preço implícito (F4)	257
Regressão hedônica (F5)	264
Variável latente (F6)	272
Decisões judiciais recentes	279
Decisões judiciais recentes – Estados Unidos (R1)	280
Decisões judiciais recentes – União Europeia (R2)	283
Decisões judiciais recentes – América Latina (R3)	289

Missão

O *Valuation Engineer Journal* publica pesquisas e estudos de caso rigorosos e voltados à prática sobre a engenharia do valor: métodos, modelos e medições que ligam decisões técnicas a resultados econômicos. Priorizamos a reprodutibilidade, a transparência dos pressupostos e a clareza da exposição.

Editorial

SEÇÃO: EDITORIAL

Bem-vindos à edição inaugural

*Tradução para o português, elaborada com assistência de IA e revisada pela
redação; a versão em inglês é a autorizada.*

William Bert Craytor, SRA — Editor-chefe

Volume 1, Número 1 · Verão de 2026

Publicado em 1º de julho de 2026

DOI: 10.67330/0tymfa32

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da licença Creative Commons
Atribuição 4.0.

Resumo. A Mensagem do Editor para o Volume 1, Número 1: por que esta revista existe, o que aportam os cinco artigos e as três seções, e um convite à participação.

Palavras-chave: editorial, engenharia de avaliação, edição inaugural

Bem-vindos à edição inaugural de *The Valuation Engineer Journal*.

Esta revista existe por causa de uma lacuna. A doutrina da avaliação há muito descreve a avaliação como a medição disciplinada do comportamento do mercado e, ainda assim, a prática cotidiana continua a apoiar-se em grande medida na seleção manual de comparáveis e em ajustes estimados apenas pelo julgamento profissional. Enquanto isso, as ferramentas estatísticas capazes de medir o que os mercados efetivamente pagam pelas características de um imóvel — a regressão hedônica, os modernos modelos de splines, o aprendizado de máquina com validação cruzada — amadureceram ao longo de décadas com escassa adoção na avaliação em exercício. *Engenheiro de avaliação* é o nome que damos ao profissional que fecha essa lacuna: competência habilitada em avaliação somada à engenharia de dados, à estatística e a protocolo reproduzível.

Os cinco artigos desta edição assentam a base de trabalho. O artigo de abertura examina a própria lacuna entre doutrina e prática e sustenta que as normas da profissão já exigem mais medição do que seus métodos cotidianos entregam. Em segundo lugar — e central para tudo o que esta revista pretende — apresenta-se na íntegra o Residual Constraint Approach (Abordagem da Restrição Residual): um protocolo que permite ao canal de regressão medir aquilo que pode ser medido, disciplina o que não pode ser medido em uma decomposição residual explícita e restringe o total de modo que o julgamento profissional possa refinar o valor, mas nunca inventá-lo. Seguem-se três artigos de software, que documentam o earthUI e suas

interfaces afins para regressão penalizada e modelos aditivos generalizados — o ferramental aberto sobre o qual esta revista se constrói —, consolidados em um único aplicativo pouco antes do fechamento da edição, conforme descrevem as notas do editor em cada artigo.

Três seções acompanham os artigos. *Foundations* (Fundamentos) abre com sete verbetes breves (F0–F6) que traçam a cadeia conceitual desde os bens heterogêneos, passando pelo espaço de características, pela função de preços hedônicos, pelos preços implícitos e pela regressão hedônica, até as variáveis latentes que habitam os resíduos. Cada verbete cita sua linhagem acadêmica, carrega seu próprio DOI e termina com uma formalização em Prolog legível por máquina — o começo de uma base de conhecimento que pretendemos ampliar, edição após edição, até convertê-la em um vocabulário compartilhado e preciso para a área. *Recent Court Decisions* (Decisões judiciais recentes) examina decisões dos Estados Unidos, da União Europeia e da América Latina que incidem sobre a prática avaliatória, porque o ambiente jurídico faz parte do problema de engenharia. *Letters to the Editor* (Cartas ao Editor) inaugura a coluna de correspondência: os métodos concebidos para serem usados são testados de maneiras que seus autores jamais anteciparam, e é aqui que esses achados — e as divergências que suscitam — têm lugar.

Todo o conteúdo desta edição é publicado em acesso aberto sob licença CC BY 4.0, tanto em PDF quanto em HTML, com um DOI da Crossref, e preservado de forma independente no Zenodo. Convidamos a envio de contribuições — artigos, verbetes de *Foundations*, correspondência e discordâncias — de avaliadores, economistas e engenheiros por igual.

William Bert Craytor, SRA
Editor-chefe

Cartas ao editor

SEÇÃO: CARTAS AO EDITOR

Um convite à correspondência

*Tradução para o português, elaborada com assistência de IA e revisada pela
redação; a versão em inglês é a autorizada.*

William Bert Craytor, SRA — Editor-chefe

Volume 1, Número 1 · Verão de 2026

Publicado em 1º de julho de 2026

DOI: 10.67330/r55mw896

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da licença Creative Commons
Atribuição 4.0.

Resumo. O editor inaugura a coluna de cartas: o que esta revista espera receber de seus leitores e como enviá-lo.

Palavras-chave: cartas, correspondência, comentário entre pares

Uma revista que apenas fala é meia revista. Esta coluna existe para a outra metade.

The Valuation Engineer Journal publica métodos feitos para serem usados — e métodos que são usados acabam testados de maneiras que seus autores jamais anteciparam. Se você aplicou o Residual Constraint Approach (Abordagem da Restrição Residual) e encontrou uma etapa que falha no seu mercado, queremos essa carta. Se um verbete de *Foundations* define um termo de modo mais restrito do que a sua prática exige, queremos essa carta. Se uma decisão judicial que resumimos se lê de outra forma para um profissional daquela jurisdição, também queremos essa carta. Divergência, replicação, refinamento e a simples correção estão todos dentro do escopo; e também os relatos breves surgidos da prática — uma técnica que funcionou, um problema de dados resolvido, um fracasso do qual convém advertir os demais.

As regras básicas são poucas. As cartas devem ser assinadas, concisas (uma página é o ideal) e específicas: cite o artigo, a seção ou o verbete a que se dirige e apresente evidências do mesmo modo como o faria em um dossiê de trabalho. Podemos editá-las quanto à extensão e à clareza, sempre com a aprovação do autor. As cartas são publicadas em acesso aberto sob licença CC BY 4.0, como tudo o mais nestas páginas, e aos autores cujo trabalho for comentado será oferecida a possibilidade de responder na mesma edição.

Envie as cartas pelo site da revista (journal.valuation-engineer.com) ou diretamente ao editor. As primeiras cartas de verdade que esta revista receber marcarão o momento em que ela se tornará uma conversa, e não apenas uma publicação — aguardamo-las com interesse.

William Bert Craytor, SRA
Editor-chefe

Artigos

A lacuna entre doutrina e prática na avaliação de imóveis: uma exposição estruturada de funções, limites e tensões

Tradução para o português do artigo publicado originalmente em inglês neste número. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

William Bert Craytor

Volume 1, Número 1 · 21 de maio de 2026

Publicado em 1º de julho de 2026

DOI: [10.67330/x7pbf13](https://doi.org/10.67330/x7pbf13)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da [licença Creative Commons Atribuição 4.0](#).

Resumo. A avaliação de imóveis ocupa uma posição de considerável interesse público; no entanto, sua doutrina oficial —centrada em pareceres de valor de mercado sob condições passadas e atuais— encontra-se em tensão crescente com o modo como os produtos da avaliação são efetivamente utilizados nos circuitos financeiros contemporâneos. Este artigo constrói esse argumento em cinco etapas, cada uma acompanhada de uma base de conhecimento explícita e consultável em Prolog: as duas primeiras estabelecem as funções sociais e econômicas que a avaliação cumpre; a terceira identifica seu principal produto final, um parecer de valor de mercado; e a quarta examina a orientação temporal e metodológica da avaliação diante da previsão quantitativa e financeira, juntamente com o limite regulatório que as separa. Todas convergem para a afirmação central da quinta: no crédito hipotecário, o valor de mercado escalonado por uma razão empréstimo-valor (LTV) funciona como uma proxy implícita do valor futuro de execução hipotecária —um uso orientado ao futuro que a doutrina do valor presente não reconhece—. Muitos avaliadores, por desconhecerem esse papel de avaliação da garantia, não tratam o valor de mercado com a cautela quanto às tendências futuras de valor que ele exige; uma questão que esta exposição pretende evidenciar e abrir a trabalhos posteriores.

Palavras-chave: avaliação de imóveis, valor de mercado, doutrina de avaliação, representação do conhecimento, normas de avaliação, valor de execução hipotecária

Sumário

1 Importância social	2
2 Importância econômica	6
3 A importância das estimativas de valor de mercado como produto final das avaliações	10
4 A orientação temporal da atividade de avaliação	14
5 O valor de mercado × LTV como proxy do valor de execução hipotecária	21
6 Conclusão	27

1 Importância social

As avaliações de bens imóveis e de outros ativos cumprem uma função social e econômica essencial ao ajudar compradores, vendedores, credores, seguradoras, investidores, tribunais, autoridades tributárias e o público em geral a tomar decisões informadas e equitativas sobre a transferência, o financiamento, a tributação e a administração da propriedade e de outras formas de riqueza. Avaliações de alta qualidade contribuem para a transparência do mercado, a confiança pública, a alocação eficiente do capital e a estabilidade das instituições financeiras, ao fornecer pareceres de valor independentes e bem fundamentados, apoiados em evidências, análise e julgamento profissional. (LaCour-Little e Malpezzi 2003; K'Akumu e Larsen 2024; Hendershott e Kane 1995; Lee, Peng e Liao 2020)

Avaliações de baixa qualidade, em contrapartida, podem produzir consequências graves em múltiplos horizontes temporais. No curto prazo, avaliações imprecisas podem levar compradores ou vendedores individuais a sofrer prejuízo financeiro imediato por pagamento excessivo, venda abaixo do valor, tributação excessiva ou decisões de crédito inadequadas. No médio prazo, deficiências sistemáticas de avaliação podem distorcer os sinais de mercado, prejudicar a análise de crédito, estimular comportamentos especulativos, contribuir para tributação ou cobertura securitária iníquas e aumentar a litigiosidade e a desconfiança nas transações. No longo prazo, falhas persistentes na qualidade das avaliações podem minar a confiança nos mercados e nas instituições, contribuir para a instabilidade financeira, a má alocação de recursos, a decadência ou o superinvestimento em bairros e iniquidades sociais mais amplas que afetam famílias, empresas e comunidades. (LaCour-Little e Malpezzi 2003; Mayer e Nothaft 2022)

Como os imóveis e outros ativos frequentemente representam parcela substancial da riqueza pessoal, empresarial e pública, a qualidade da avaliação não é apenas uma preocupação técnica privada, mas uma questão de significativo interesse público, ligada à resiliência econômica, à equidade e ao bem comum de longo prazo. (Mayer e Nothaft 2022)

```
%% appraisal_social.pl
%%
%% A Prolog knowledge base encoding the social and public-interest role
%% of real estate and other asset appraisals. Companion to
%% appraisal_economics.pl, which covers the strictly economic axis.
%%
%% All domain-specific predicates are prefixed with `social_` so this
%% module can be loaded alongside appraisal_economics.pl (and any other
%% companion KBs) without name collisions, and so cross-KB rules can
%% mention both vocabularies unambiguously.
%%
%% Conceptual model:
%%   - Appraisals support social DECISION TYPES.
%%   - Each decision involves identifiable PARTIES.
%%   - Appraisal QUALITY (high | low) drives EFFECTS across time HORIZONS.
%%   - Emphasis: equity, public confidence, public interest.
%%
%% Author: generated for Bert Craytor, ValEngr / RCA context.

:- module(appraisal_social,
    [ social_horizon/1
    , social_party/1
    , social_decision_type/1
    , social_involves/2
    , social_effect/4
    , social_contribution/1
    , social_hallmark/1
    , social_equity_principle/1
    ],
    % DecisionType, Party
    % Quality, Horizon, Domain, Effect
```

```

        , social_public_interest_claim/1
        , social_positive_effect/3
        , social_negative_effect/3
        , social_horizon_summary/2
        , social_consequences_of_quality/2
        , social_parties_affected_by/2
        , social_public_harm/2
        , social_private_harm/2
        , social_appraisal_is_public_interest/0
    ]).

:- discontinuous social_effect/4.
:- discontinuous social_involves/2.

%% -----
%% 1. Time horizons
%% -----

social_horizon(short_term).
social_horizon(medium_term).
social_horizon(long_term).

%% -----
%% 2. Parties whose decisions depend on appraisals
%%
%% The opening sentence enumerates these explicitly.
%% -----

social_party(buyers).
social_party(sellers).
social_party(lenders).
social_party(insurers).
social_party(investors).
social_party(courts).
social_party(taxing_authorities).
social_party(general_public).

%% -----
%% 3. Types of decisions appraisals inform
%% -----

social_decision_type(property_transfer).
social_decision_type(financing).
social_decision_type(taxation).
social_decision_type(stewardship).
social_decision_type(insurance_coverage).
social_decision_type(adjudication).           % courts: condemnation, divorce, estate
social_decision_type(public_administration).

social_involves(property_transfer,    buyers).
social_involves(property_transfer,    sellers).
social_involves(financing,            lenders).
social_involves(financing,            buyers).
social_involves(taxation,              taxing_authorities).
social_involves(taxation,              general_public).
social_involves(stewardship,          investors).
social_involves(stewardship,          general_public).
social_involves(insurance_coverage,   insurers).

```

```

social_involves(insurance_coverage, buyers).
social_involves(adjudication, courts).
social_involves(public_administration, taxing_authorities).
social_involves(public_administration, general_public).

%% -----
%% 4. What high-quality appraisals contribute (second sentence)
%% -----

social_contribution(market_transparency).
social_contribution(public_confidence).
social_contribution(efficient_capital_allocation).
social_contribution(stability_of_financial_institutions).

%% Characteristics of a high-quality appraisal opinion of value.
social_hallmark(independence).
social_hallmark(well_supported).
social_hallmark(grounded_in_evidence).
social_hallmark(grounded_in_analysis).
social_hallmark(grounded_in_professional_judgment).

%% -----
%% 5. Effects of appraisal quality by horizon
%% -----

%% ---- SHORT TERM: high quality ----
social_effect(high, short_term, property_transfer, fair_pricing_for_parties).
social_effect(high, short_term, financing, appropriate_lending_decisions).
social_effect(high, short_term, taxation, accurate_individual_assessment).

%% ---- SHORT TERM: low quality ----
social_effect(low, short_term, property_transfer, buyer_overpayment).
social_effect(low, short_term, property_transfer, seller_underselling).
social_effect(low, short_term, taxation, excessive_individual_taxation).
social_effect(low, short_term, financing,
↳ inappropriate_lending_decisions).
social_effect(low, short_term, property_transfer,
↳ immediate_financial_harm_to_individuals).

%% ---- MEDIUM TERM: high quality ----
social_effect(high, medium_term, markets, clear_market_signals).
social_effect(high, medium_term, financing, sound_credit_underwriting).
social_effect(high, medium_term, taxation, equitable_tax_assessment).
social_effect(high, medium_term, insurance_coverage,
↳ equitable_insurance_coverage).
social_effect(high, medium_term, adjudication, reduced_litigation_pressure).

%% ---- MEDIUM TERM: low quality ----
social_effect(low, medium_term, markets, distorted_market_signals).
social_effect(low, medium_term, financing,
↳ impaired_credit_underwriting).
social_effect(low, medium_term, markets,
↳ encouragement_of_speculation).
social_effect(low, medium_term, taxation, inequitable_taxation).
social_effect(low, medium_term, insurance_coverage,
↳ inequitable_insurance_coverage).
social_effect(low, medium_term, adjudication, increased_litigation).
social_effect(low, medium_term, markets, transactional_distrust).

```

```

%% ---- LONG TERM: high quality ----
social_effect(high, long_term, institutions,    public_confidence_in_institutions).
social_effect(high, long_term, markets,         public_confidence_in_markets).
social_effect(high, long_term, institutions,    financial_stability).
social_effect(high, long_term, communities,     equitable_community_outcomes).
social_effect(high, long_term, public_interest, economic_resilience).
social_effect(high, long_term, public_interest, social_fairness).
social_effect(high, long_term, public_interest, long_term_public_good).

%% ---- LONG TERM: low quality ----
social_effect(low, long_term, markets,          eroded_market_confidence).
social_effect(low, long_term, institutions,     eroded_institutional_confidence).
social_effect(low, long_term, institutions,     financial_instability).
social_effect(low, long_term, markets,          resource_misallocation).
social_effect(low, long_term, communities,     neighborhood_decline).
social_effect(low, long_term, communities,     neighborhood_overinvestment).
social_effect(low, long_term, communities,     social_inequities_for_households).
social_effect(low, long_term, communities,     social_inequities_for_businesses).
social_effect(low, long_term, communities,     social_inequities_for_communities).

%% -----
%% 6. Equity principles and public-interest claims
%% -----

social_equity_principle(informed_decisions).
social_equity_principle(equitable_decisions).
social_equity_principle(fair_treatment_of_parties).
social_equity_principle(independence_of_valuer).

social_public_interest_claim(
    'Real estate and other assets are a substantial portion of personal, corporate,
    ↪ and public wealth.').
social_public_interest_claim(
    'Appraisal quality is not merely a private technical concern.').
social_public_interest_claim(
    'Appraisal quality is tied to economic resilience.').
social_public_interest_claim(
    'Appraisal quality is tied to fairness.').
social_public_interest_claim(
    'Appraisal quality is tied to the long-term public good.').

%% Top-level conclusion of the passage.
social_appraisal_is_public_interest :-
    social_public_interest_claim(_).    % at least one such claim exists

%% -----
%% 7. Derived / query predicates
%% -----

social_positive_effect(Horizon, Domain, Effect) :-
    social_effect(high, Horizon, Domain, Effect).

social_negative_effect(Horizon, Domain, Effect) :-
    social_effect(low, Horizon, Domain, Effect).

social_horizon_summary(Horizon, Effects) :-
    findall(Quality-Effect,

```

```

        social_effect(Quality, Horizon, _Domain, Effect),
        Effects).

social_consequences_of_quality(Quality, List) :-
    findall(consequence(Horizon, Domain, Effect),
            social_effect(Quality, Horizon, Domain, Effect),
            List).

social_parties_affected_by(DecisionType, Parties) :-
    setof(P, social_involves(DecisionType, P), Parties).

%% Harms that fall on the public rather than on transaction principals.
social_public_harm(Horizon, Effect) :-
    social_effect(low, Horizon, Domain, Effect),
    member(Domain, [markets, institutions, communities, public_interest,
                    ↪ taxation]).

%% Private (transaction-level) harms.
social_private_harm(Horizon, Effect) :-
    social_effect(low, Horizon, Domain, Effect),
    member(Domain, [property_transfer, financing, insurance_coverage]).

```

2 Importância econômica

As avaliações de bens imóveis e de outros ativos desempenham um papel econômico fundamental ao sustentar o funcionamento eficiente dos mercados de crédito, as decisões de investimento, os sistemas tributários, a elaboração de demonstrações financeiras, a subscrição de seguros, a administração de espólios, as operações societárias e o planejamento de políticas públicas. Avaliações confiáveis ajudam a estabelecer medidas críveis de garantia e de valor de mercado que permitem ao capital fluir com mais eficiência entre tomadores, credores, investidores, governos e empresas. (LaCour-Little e Malpezzi 2003; Ebrahim e Hussain 2010)

No curto prazo, avaliações precisas reduzem a incerteza das transações, aprimoram as decisões de crédito, facilitam negociações justas e ajudam a evitar perdas financeiras imediatas decorrentes de superavaliação ou subavaliação. No médio prazo, a qualidade das avaliações influencia a estabilidade e a eficiência de mercados mais amplos, ao afetar o risco hipotecário, a alocação de investimentos, a equidade tributária, a adequação das coberturas de seguro e o comportamento de precificação dos participantes do mercado. Avaliações frágeis ou sistematicamente enviesadas nesse período podem contribuir para bolhas de ativos, carteiras de crédito comprometidas, incentivos distorcidos à incorporação e maior risco de inadimplência ou de litígio. (LaCour-Little e Malpezzi 2003; Hendershott e Kane 1992; Griffin e Maturana 2016)

No longo prazo, as consequências econômicas cumulativas da qualidade das avaliações tornam-se ainda mais significativas, influenciando padrões de formação de capital, investimento em infraestrutura, transferência intergeracional de riqueza, receitas públicas, acessibilidade habitacional e a resiliência geral do sistema financeiro. Deficiências persistentes nas práticas de avaliação podem contribuir para a má alocação crônica de capital, a instabilidade bancária, a redução da confiança dos investidores e o uso ineficiente da terra e dos ativos produtivos, ao passo que normas de avaliação sólidas ajudam a sustentar mercados estáveis, um planejamento econômico consistente e o crescimento econômico de longo prazo. (Ghent, Torous e Valkanov 2019; Hendershott e Kane 1992; Ebrahim e Hussain 2010)

```

%% appraisal_economics.pl
%%
%% A Prolog knowledge base encoding the economic role of real estate

```

```

%% and other asset appraisals across short-, medium-, and long-term
%% horizons, including both the consequences of high-quality appraisal
%% practice and the consequences of weak or biased practice.
%%
%% All domain-specific predicates are prefixed with `economic_` so this
%% module can be loaded alongside appraisal_social.pl (and any other
%% companion KBs) without name collisions, and so cross-KB rules can
%% mention both vocabularies unambiguously.
%%
%% Conceptual model:
%%   - Appraisals support a set of economic FUNCTIONS.
%%   - Each function operates over one or more time HORIZONS.
%%   - Appraisal QUALITY (high / low) drives EFFECTS on those functions.
%%   - Effects are POSITIVE (benefits) when quality is high and
%%     NEGATIVE (harms) when quality is low/systematically biased.
%%
%% Author: generated for Bert Craytor, ValEngr / RCA context.

:- module(appraisal_economics,
    [ economic_horizon/1
    , economic_function/1
    , economic_supports/2
    , economic_effect/4
    ↪ Effect
    , economic_stakeholder/1
    , economic_benefits/2
    , economic_positive_effect/3
    , economic_negative_effect/3
    , economic_horizon_summary/2
    , economic_consequences_of_quality/2
    , economic_functions_at_horizon/2
    , economic_stakeholders_affected_by/2
    , economic_reliable_appraisals_improve/1
    , economic_appraisal_quality_matters_at/1
    ]).

:- disjoint economic_effect/4.
:- disjoint economic_supports/2.

%% -----
%% 1. Time horizons
%% -----

economic_horizon(short_term).
economic_horizon(medium_term).
economic_horizon(long_term).

%% -----
%% 2. Economic functions supported by appraisals
%% -----

economic_function(credit_markets).
economic_function(investment_decisions).
economic_function(taxation_systems).
economic_function(financial_reporting).
economic_function(insurance_underwriting).
economic_function(estate_administration).
economic_function(corporate_transactions).

```

```

economic_function(public_policy_planning).

%% -----
%% 3. Stakeholders whose decisions depend on appraisals
%% -----

economic_stakeholder(borrowers).
economic_stakeholder(lenders).
economic_stakeholder(investors).
economic_stakeholder(governments).
economic_stakeholder(businesses).
economic_stakeholder(insurers).
economic_stakeholder(taxpayers).
economic_stakeholder(homeowners).

%% Capital flows more efficiently when appraisals are reliable.
economic_benefits(borrowers,    credit_markets).
economic_benefits(lenders,      credit_markets).
economic_benefits(investors,    investment_decisions).
economic_benefits(governments,  taxation_systems).
economic_benefits(governments,  public_policy_planning).
economic_benefits(businesses,   corporate_transactions).
economic_benefits(businesses,   financial_reporting).
economic_benefits(insurers,     insurance_underwriting).
economic_benefits(taxpayers,    taxation_systems).
economic_benefits(homeowners,   credit_markets).

%% -----
%% 4. Which functions are most salient at each horizon
%% -----

economic_supports(credit_markets,    short_term).
economic_supports(investment_decisions, short_term).
economic_supports(corporate_transactions, short_term).

economic_supports(credit_markets,    medium_term).
economic_supports(investment_decisions, medium_term).
economic_supports(taxation_systems,   medium_term).
economic_supports(insurance_underwriting, medium_term).

economic_supports(credit_markets,    long_term).
economic_supports(taxation_systems,   long_term).
economic_supports(public_policy_planning, long_term).
economic_supports(financial_reporting, long_term).
economic_supports(estate_administration, long_term).

%% -----
%% 5. Effects of appraisal quality, by horizon and function
%% -----

%% ---- SHORT TERM: high quality ----
economic_effect(high, short_term, credit_markets,
↳ reduced_transactional_uncertainty).
economic_effect(high, short_term, credit_markets,
↳ improved_lending_decisions).
economic_effect(high, short_term, corporate_transactions, fair_negotiations).
economic_effect(high, short_term, investment_decisions,
↳ prevention_of_immediate_losses).

```

```

%% ---- SHORT TERM: low quality ----
economic_effect(low, short_term, credit_markets,
↳ transactional_uncertainty).
economic_effect(low, short_term, credit_markets, poor_lending_decisions).
economic_effect(low, short_term, investment_decisions,
↳ losses_from_overvaluation).
economic_effect(low, short_term, investment_decisions,
↳ losses_from_undervaluation).

%% ---- MEDIUM TERM: high quality ----
economic_effect(high, medium_term, credit_markets, stable_mortgage_risk).
economic_effect(high, medium_term, investment_decisions,
↳ efficient_investment_allocation).
economic_effect(high, medium_term, taxation_systems, tax_equity).
economic_effect(high, medium_term, insurance_underwriting,
↳ adequate_insurance_coverage).
economic_effect(high, medium_term, investment_decisions,
↳ rational_market_pricing).

%% ---- MEDIUM TERM: low quality ----
economic_effect(low, medium_term, investment_decisions, inflated_asset_bubbles).
economic_effect(low, medium_term, credit_markets,
↳ impaired_lending_portfolios).
economic_effect(low, medium_term, public_policy_planning,
↳ distorted_development_incentives).
economic_effect(low, medium_term, credit_markets, increased_default_risk).
economic_effect(low, medium_term, corporate_transactions,
↳ increased_litigation_risk).

%% ---- LONG TERM: high quality ----
economic_effect(high, long_term, investment_decisions, sound_capital_formation).
economic_effect(high, long_term, public_policy_planning,
↳ sound_infrastructure_investment).
economic_effect(high, long_term, estate_administration,
↳ orderly_intergenerational_wealth_transfer).
economic_effect(high, long_term, taxation_systems, stable_public_revenues).
economic_effect(high, long_term, public_policy_planning,
↳ supported_housing_affordability).
economic_effect(high, long_term, credit_markets,
↳ financial_system_resilience).
economic_effect(high, long_term, investment_decisions,
↳ long_term_economic_growth).

%% ---- LONG TERM: low quality ----
economic_effect(low, long_term, investment_decisions,
↳ chronic_capital_misallocation).
economic_effect(low, long_term, credit_markets, banking_instability).
economic_effect(low, long_term, investment_decisions,
↳ reduced_investor_confidence).
economic_effect(low, long_term, public_policy_planning, inefficient_land_use).
economic_effect(low, long_term, public_policy_planning,
↳ inefficient_productive_asset_use).

%% -----
%% 6. Derived / query predicates
%% -----

```

```

economic_positive_effect(Horizon, Function, Effect) :-
    economic_effect(high, Horizon, Function, Effect).

economic_negative_effect(Horizon, Function, Effect) :-
    economic_effect(low, Horizon, Function, Effect).

economic_horizon_summary(Horizon, Effects) :-
    findall(Quality-Effect,
            economic_effect(Quality, Horizon, _Function, Effect),
            Effects).

economic_consequences_of_quality(Quality, List) :-
    findall(consequence(Horizon, Function, Effect),
            economic_effect(Quality, Horizon, Function, Effect),
            List).

economic_functions_at_horizon(Horizon, Functions) :-
    setof(F, economic_supports(F, Horizon), Functions).

economic_stakeholders_affected_by(Function, Stakeholders) :-
    setof(S, economic_benefits(S, Function), Stakeholders).

%% Overall claim: reliable appraisals improve allocative efficiency.
economic_reliable_appraisals_improve(capital_flow_efficiency).
economic_reliable_appraisals_improve(market_stability).
economic_reliable_appraisals_improve(long_term_economic_growth).

%% Top-level qualitative claim: appraisal quality matters at every horizon.
economic_appraisal_quality_matters_at(Horizon) :-
    economic_horizon(Horizon),
    \+ \+ economic_effect(high, Horizon, _, _),
    \+ \+ economic_effect(low, Horizon, _, _).

```

3 A importância das estimativas de valor de mercado como produto final das avaliações

As avaliações de valor de mercado desempenham um papel importante na análise do risco da garantia porque fornecem uma estimativa profissionalmente elaborada do preço provável que um imóvel ou ativo poderia alcançar em condições normais de mercado, servindo assim como referência fundamental para credores, investidores e analistas de garantias que avaliam a exposição de crédito e os desfechos potenciais de execução hipotecária. (Agarwal, Ben-David e Yao 2015) Embora o valor de execução hipotecária ou de liquidação possa diferir substancialmente do valor de mercado em razão de condições de comercialização adversas, prazos de venda acelerados, custos judiciais, deterioração física ou percepção desfavorável do mercado, os analistas de garantias recorrem com frequência às avaliações de valor de mercado como ponto de partida para estimar o potencial de recuperação em cenários de estresse. (Campbell, Giglio e Pathak 2011; Pennington-Cross 2006; Clauretie e Daneshvary 2009)

Ao analisar a relação entre valor de mercado, saldos devedores, razões empréstimo-valor, tendências de mercado, estado de conservação do imóvel, liquidez e condições locais de oferta e demanda, os analistas de garantias avaliam a probabilidade e a severidade das perdas em caso de inadimplência do tomador. (Qi e Yang 2009)

No curto prazo, avaliações precisas de valor de mercado ajudam os credores a tomar decisões prudentes de concessão e a estabelecer margens adequadas de garantia. No médio prazo, elas apoiam a gestão do risco de carteira, a securitização de créditos, a constituição de provisões e a

conformidade regulatória, ao ajudar as instituições a avaliar a suficiência variável das garantias ao longo dos ciclos econômicos. No longo prazo, práticas de avaliação consistentemente confiáveis contribuem para a estabilidade do sistema financeiro ao reduzir o risco de alavancagem excessiva, de crédito mal garantido e de precificação sistemicamente equivocada dos ativos dados em garantia, fatores capazes de ampliar as perdas em execuções hipotecárias e as perturbações econômicas mais amplas em períodos de estresse de mercado.(Eriksen et al. 2019)

```
%% appraisal_collateral.pl
%%
%% A Prolog knowledge base encoding the role of market value (MV)
%% appraisals in collateral risk analysis. Companion to
%% appraisal_economics.pl and appraisal_social.pl.
%%
%% Naming convention:
%%   - Predicates that have direct counterparts in the companion KBs
%%     (horizon, effect, summary, consequences, etc.) are prefixed
%%     `collateral_` so this module can be loaded alongside the others
%%     without name collisions, and so bridge rules can mention each
%%     vocabulary unambiguously.
%%   - Predicates that are unique to the collateral-analysis domain
%%     (value_basis, divergence_factor, risk_input, etc.) are NOT
%%     prefixed. They have no analogues in the other KBs and no
%%     collision risk, so the prefix would only add noise.

:- module(appraisal_collateral,
    [ collateral_horizon/1
      , collateral_analyst/1
      , collateral_effect/4                                % Quality, Horizon, Domain,
        ↪ Effect
      , collateral_positive_effect/3
      , collateral_negative_effect/3
      , collateral_horizon_summary/2
      , collateral_consequences_of_quality/2
      , value_basis/1
      , uses/2                                              % Analyst, ValueBasis
      , divergence_factor/1
      , stress_scenario/1
      , risk_input/1
      , risk_output/1
      , inputs_for/2
      , mv_is_benchmark_for/1
    ]).

:- disjoint collateral_effect/4.
:- disjoint uses/2.

%% -----
%% 1. Time horizons (namespaced -- parallel to economic/social_)
%% -----

collateral_horizon(short_term).
collateral_horizon(medium_term).
collateral_horizon(long_term).

%% -----
%% 2. Who uses MV appraisals in collateral risk analysis
%% -----
```

```

collateral_analyst(lenders).
collateral_analyst(investors).
collateral_analyst(collateral_analysts).
collateral_analyst(portfolio_risk_managers).
collateral_analyst(regulators).
collateral_analyst(securitization_arrangers).

%% -----
%% 3. Value bases distinguished in the passage
%% -----

value_basis(market_value).          % "normal market conditions"
value_basis(foreclosure_value).      % distressed
value_basis(liquidation_value).      % distressed, accelerated

%% Which value bases each analyst type works with.
uses(lenders, market_value).
uses(investors, market_value).
uses(collateral_analysts, market_value).
uses(collateral_analysts, foreclosure_value).
uses(collateral_analysts, liquidation_value).
uses(portfolio_risk_managers, market_value).
uses(securitization_arrangers, market_value).
uses(regulators, market_value).

%% Top-level claim of the opening sentence.
mv_is_benchmark_for(credit_exposure).
mv_is_benchmark_for(foreclosure_outcomes).
mv_is_benchmark_for(recovery_potential).

%% -----
%% 4. Why foreclosure / liquidation value diverges from MV
%% -----

divergence_factor(distressed_marketing_conditions).
divergence_factor(accelerated_sale_period).
divergence_factor(legal_costs).
divergence_factor(physical_deterioration).
divergence_factor(adverse_market_sentiment).

%% Stressed scenarios in which divergence factors apply.
stress_scenario(borrower_default).
stress_scenario(foreclosure).
stress_scenario(forced_liquidation).
stress_scenario(market_downturn).

%% -----
%% 5. Inputs the collateral analyst combines with MV, and the
%%     outputs they produce
%% -----

risk_input(market_value).
risk_input(loan_balance).
risk_input(loan_to_value_ratio).
risk_input(market_trends).
risk_input(property_condition).
risk_input(liquidity).
risk_input(local_supply_demand).

```

```

risk_output(likelihood_of_loss).
risk_output(severity_of_loss).
risk_output(recovery_estimate).
risk_output(loss_given_default).

%% inputs_for(+Output, -Inputs)
%% All risk inputs are used to produce each risk output.
inputs_for(Output, Inputs) :-
    risk_output(Output),
    setof(I, risk_input(I), Inputs).

%% -----
%% 6. Effects of appraisal quality by horizon
%% -----

%% ---- SHORT TERM: high quality ----
collateral_effect(high, short_term, underwriting, prudent_underwriting_decisions).
collateral_effect(high, short_term, underwriting, appropriate_collateral_margins).

%% ---- SHORT TERM: low quality ----
collateral_effect(low, short_term, underwriting,
    ↪ imprudent_underwriting_decisions).
collateral_effect(low, short_term, underwriting, inadequate_collateral_margins).
collateral_effect(low, short_term, underwriting, mispriced_credit_exposure).

%% ---- MEDIUM TERM: high quality ----
collateral_effect(high, medium_term, portfolio,
    ↪ effective_portfolio_risk_management).
collateral_effect(high, medium_term, portfolio, sound_loan_securitization).
collateral_effect(high, medium_term, portfolio, appropriate_reserve_allocation).
collateral_effect(high, medium_term, regulatory, regulatory_compliance).
collateral_effect(high, medium_term, portfolio,
    ↪ cycle_aware_collateral_evaluation).

%% ---- MEDIUM TERM: low quality ----
collateral_effect(low, medium_term, portfolio,
    ↪ weakened_portfolio_risk_management).
collateral_effect(low, medium_term, portfolio, mispriced_securitizations).
collateral_effect(low, medium_term, portfolio, inadequate_loss_reserves).
collateral_effect(low, medium_term, regulatory, regulatory_noncompliance_risk).
collateral_effect(low, medium_term, portfolio, misjudged_collateral_adequacy).

%% ---- LONG TERM: high quality ----
collateral_effect(high, long_term, system, financial_system_stability).
collateral_effect(high, long_term, system, restraint_on_excessive_leverage).
collateral_effect(high, long_term, system, well_secured_lending).
collateral_effect(high, long_term, system, accurate_systemic_collateral_pricing).
collateral_effect(high, long_term, system, dampened_foreclosure_amplification).

%% ---- LONG TERM: low quality ----
collateral_effect(low, long_term, system, financial_system_instability).
collateral_effect(low, long_term, system, excessive_leverage).
collateral_effect(low, long_term, system, poorly_secured_lending).
collateral_effect(low, long_term, system, systemic_collateral_mispricing).
collateral_effect(low, long_term, system, amplified_foreclosure_losses).
collateral_effect(low, long_term, system,
    ↪ amplified_economic_disruption_under_stress).

```

```

%% -----
%% 7. Derived / query predicates (namespaced)
%% -----

collateral_positive_effect(Horizon, Domain, Effect) :-
    collateral_effect(high, Horizon, Domain, Effect).

collateral_negative_effect(Horizon, Domain, Effect) :-
    collateral_effect(low, Horizon, Domain, Effect).

collateral_horizon_summary(Horizon, Effects) :-
    findall(Quality-Effect,
            collateral_effect(Quality, Horizon, _Domain, Effect),
            Effects).

collateral_consequences_of_quality(Quality, List) :-
    findall(consequence(Horizon, Domain, Effect),
            collateral_effect(Quality, Horizon, Domain, Effect),
            List).

```

4 A orientação temporal da atividade de avaliação

Os avaliadores geralmente elaboram pareceres de valor de mercado referidos a um ponto no tempo, em uma data-base específica —retrospectiva (passada), atual ou, com limitações definidas, prospectiva (futura)—, em vez de previsões probabilísticas ao longo de períodos prolongados. Essa orientação temporal e metodológica representa uma das distinções centrais entre a prática regulada da avaliação e o trabalho de analistas quantitativos, analistas financeiros, analistas de investimentos e economistas.(Quan e Quigley 1991)

Na maioria dos trabalhos de avaliação, em particular naqueles que envolvem valor de mercado sob normas como as Uniform Standards of Professional Appraisal Practice (USPAP, normas uniformes de prática profissional de avaliação) ou os International Valuation Standards (IVS, normas internacionais de avaliação), o avaliador produz uma estimativa pontual conciliada (ou, nos casos permitidos, um intervalo estreito) ancorada em evidências de mercado observáveis, nas condições vigentes e em metodologias reconhecidas em uma única data-base. As USPAP permitem explicitamente conclusões de valor expressas como uma quantia específica, um intervalo de valores ou uma relação com um parâmetro de referência; ainda assim, o parecer conciliado permanece sendo uma indicação pontual crível vinculada a essa data.(The Appraisal Foundation 2024; International Valuation Standards Council 2025)

Embora os avaliadores necessariamente incorporem as expectativas razoáveis dos participantes do mercado quanto a tendências, oferta e demanda, depreciação, taxas de capitalização ou de desconto e absorção, fazem-no para sustentar um valor respaldado pelo mercado na data-base, e não para prever movimentos especulativos. Os pareceres de valor prospectivo (por exemplo, o valor estabilizado à conclusão de um empreendimento) exigem pressupostos extraordinários ou condições hipotéticas explícitos e permanecem restritos a uma data-base futura definida, com respaldo derivado do mercado.

Em contraste, os analistas quantitativos, financeiros e de investimentos frequentemente tratam a incerteza por meio de estimativas que abrangem um leque de resultados possíveis ao longo de períodos, empregando distribuições de probabilidade, simulações de Monte Carlo, análise de cenários, testes de sensibilidade ou métricas de valor em risco. Seu foco costuma centrar-se no desempenho futuro esperado, nos retornos ajustados ao risco ou no valor de investimento para uma parte específica em horizontes variados, e não em um único ponto de valor de mercado

em uma data. (Hull 2018)

Esse trabalho também pressupõe competências técnicas que uma licença de avaliador nem exige nem certifica. Ajustar um modelo de multivariate adaptive regression splines (MARS, splines de regressão adaptativa multivariada), por exemplo, requer o discernimento necessário para controlar a complexidade do modelo, de modo que um forte ajuste dentro da amostra (um R^2 elevado) não seja comprado ao preço do sobreajuste, mas se transfira aos dados reservados (um R^2 elevado em validação cruzada), o que por sua vez pressupõe uma compreensão operacional dos algoritmos subjacentes. (Friedman 1991; Steurer, Hill e Pfeifer 2021) Ao lado dessa facilidade quantitativa, o trabalho recompensa a capacidade conceitual de lidar com conceitos altamente estruturados, de discernir distinções finas de significado e de assimilar métodos desconhecidos. Trata-se de habilidades, e não de credenciais: um avaliador licenciado pode possuí-las e um analista sem licença pode não as ter, razão exata pela qual elas não determinam, por si sós, de que lado da linha regulatória recai uma dada tarefa.

Os avaliadores também podem tratar de elementos limitados voltados ao futuro, como projetar fluxos de receita, vacância, despesas e valores de reversão em análises de fluxo de caixa descontado, ou estimar valores prospectivos em contextos de incorporação ou de litígio. Esses componentes permanecem dentro dos limites da avaliação quando fundamentados em evidências atuais de mercado e em normas profissionais que enfatizam a fundamentação e a recusa da especulação sem respaldo. (Diaz e Wolverton 1998; Gallimore 1996; Crosby, Lizieri e McAllister 2010) A prática moderna da avaliação integra cada vez mais ferramentas estatísticas, análise de cenários e analítica de riscos —particularmente na avaliação em massa, nos modelos automatizados de avaliação ou no trabalho de carteira—, mas o produto central continua sendo, em regra, um parecer de valor de mercado referido a um ponto no tempo. (International Association of Assessing Officers 2017; Office of the Comptroller of the Currency and Board of Governors of the Federal Reserve System and Federal Deposit Insurance Corporation and National Credit Union Administration and Consumer Financial Protection Bureau and Federal Housing Finance Agency 2024; Alexandrov, Goodman e Neal 2023)

Pode-se caracterizar a prática da avaliação como primordialmente probatória e voltada a estimativas pontuais do passado ao presente (ou prospectivas limitadas), ao passo que a análise financeira e quantitativa é mais probabilística, distribucional e explicitamente orientada ao futuro ao longo de horizontes temporais. (Geltner 1991) Os avaliadores perguntam: “Qual é a indicação de valor de mercado mais bem fundamentada na data-base especificada, dadas as evidências atuais e os pressupostos definidos?” Os analistas financeiros e quantitativos indagam com mais frequência: “Que leque de resultados é provável ao longo de um período futuro, com que probabilidades, e como os recursos devem ser alocados?”

O limite não é absoluto e continua a evoluir com a complexidade do mercado. Ainda assim, persistem diferenças fundamentais de finalidade (valor de mercado *versus* valor de investimento), de exigências probatórias e de marcos regulatórios.

A linha operativa, contudo, não é traçada pela habilidade, e sim pela lei: a pergunta definidora é simplesmente se uma dada tarefa é daquelas para as quais a licença de avaliador é legalmente exigida. Analistas quantitativos, analistas financeiros, economistas, analistas de investimentos, gestores de carteira e profissionais afins podem legalmente estimar, prever, modelar ou analisar valores futuros de ativos —inclusive projeções distribucionais ou por intervalos— sem uma licença de avaliador de imóveis, desde que o trabalho não constitua uma avaliação regulada. (Murphy 2012) As distinções essenciais assentam-se em:

1. o tipo de valor (por exemplo, uma estimativa pontual de valor de mercado *versus* um valor de investimento ou uma previsão probabilística);
2. o foco temporal (uma data-base específica *versus* um período ou horizonte);
3. a forma do produto (um ponto ou intervalo estreito *versus* uma distribuição de probabilidade completa);
4. a finalidade pretendida e se o trabalho é apresentado como uma avaliação; e

5. o envolvimento de contextos de crédito regulado ou de contextos jurídicos que exijam observância de normas como as USPAP.

Esse arcabouço originou-se em grande medida de regulamentações bancárias como a FIRREA (1989) nos Estados Unidos, que buscou assegurar pareceres de valor de mercado críveis e independentes para transações relacionadas ao âmbito federal. (United States Congress 1989; K'Akumu e Larsen 2024) Muitas atividades sofisticadas de avaliação —preços-alvo em análises de ações, estudos de viabilidade, modelagem de risco de carteira, precificação de opções ou analítica preditiva— ficam fora da exigência de licença porque servem a finalidades de investimento, previsão ou assessoria, e não a funções reguladas de avaliação.

```

%% appraisal_scope_regulation.pl
%%
%% Knowledge base for the boundary of regulated appraisal practice.
%%
%% TWO TIERS, deliberately kept separate:
%%
%% TIER 1 -- THE DEFINING TEST (regulation cluster). A task is a
%% regulated appraisal IFF it is one for which an appraiser's
%% license is legally required. requires_license/1 is that test;
%% it -- and only it -- decides regulated_appraisal/1 vs.
%% unregulated_valuation/1.
%%
%% TIER 2 -- DESCRIPTIVE TENDENCIES (scope cluster). How the two
%% populations -- `appraisal' and `quant_financial' -- typically
%% differ in approach, temporal scope, output form, and skill set.
%% These are CORRELATES, not the boundary: the populations overlap
%% (a licensed appraiser may do quant work; a quant may be
%% unlicensed), and nothing here determines licensure. No claim is
%% made about intelligence -- only about training and methodology.
%%
%% `temporal scope' and `approach' are DISTINCT axes: an approach has a
%% typical (defeasible) temporal mode, but the two are not the same
%% viewpoint -- which is why an appraiser using an AVM (appraisal
%% approach, distributional methods) is possible.

:- module(appraisal_scope_regulation,
    [ % --- TIER 1: the defining test (regulation) ---
      regulation_framework/1
    , regulation_framework_origin/2
    , regulation_distinguishing_factor/1
    , regulation_triggers_licensure/1
    , regulation_definition_of_appraisal/1
    , regulation_unlicensed_activity/2
    , regulation_licensed_activity/2
    , regulation_legal_exposure_condition/1
    , requires_license/1
    , exempt_from_appraisal_license/1
    , regulated_appraisal/1
    , unregulated_valuation/1
    , regulatory_asymmetry/2

      % --- TIER 2: descriptive tendencies (scope) ---
    , valuation_approach/1
    , scope_profession/1
    , scope_orientation/2
    , temporal_scope/1
    , temporal_mode/1
    ]
).
```

```

    , scope_temporal_focus/2
    , scope_output_form/2
    , scope_skill/1
    , scope_skill_demanded_by/2
    , scope_inputs_considered/2
    , scope_future_oriented_context/2
    , scope_professional_constraint/1
    , scope_contrast/4
    , scope_question/2
    , scope_convergence_area/1
  ]).

:- discontiguous regulation_unlicensed_activity/2.
:- discontiguous regulation_licensed_activity/2.
:- discontiguous scope_contrast/4.

%% =====
%% TIER 1 -- THE DEFINING TEST (the boundary)
%% =====

regulation_framework(firrea_1989).
regulation_framework(uspap).
regulation_framework(ivs).
regulation_framework(state_licensing).
regulation_framework(state_certification).

regulation_framework_origin(firrea_1989,    savings_and_loan_crisis).
regulation_framework_origin(uspap,          firrea_1989).
regulation_framework_origin(state_licensing, firrea_1989).

regulation_distinguishing_factor(kind_of_value).
regulation_distinguishing_factor(purpose_of_valuation).
regulation_distinguishing_factor(temporal_focus).
regulation_distinguishing_factor(output_form).
regulation_distinguishing_factor(represented_as_appraisal).
regulation_distinguishing_factor(federally_related_transaction_involvement).

regulation_triggers_licensure(federally_related_mortgage_transaction).
regulation_triggers_licensure(work_represented_as_appraisal).
regulation_triggers_licensure(work_meeting_state_appraisal_definition).
regulation_triggers_licensure(regulated_lending_transaction).

regulation_definition_of_appraisal(formal_opinion_of_value).
regulation_definition_of_appraisal(concerning_identified_real_property).
regulation_definition_of_appraisal(developed_per_recognized_methods).
regulation_definition_of_appraisal(for_compensation).

regulation_unlicensed_activity(equity_analyst,          stock_price_targets).
regulation_unlicensed_activity(investment_bank,
  ↪ enterprise_value_estimates).
regulation_unlicensed_activity(hedge_fund,
  ↪ commercial_real_estate_appreciation_forecasts).
regulation_unlicensed_activity(economist,
  ↪ housing_market_trend_projections).
regulation_unlicensed_activity(insurance_analyst,
  ↪ replacement_cost_estimates).
regulation_unlicensed_activity(developer,              feasibility_studies).

```

```

regulation_unlicensed_activity(internal_corporate_analyst,
    ↪ asset_impairment_estimates).
regulation_unlicensed_activity(internal_corporate_analyst,
    ↪ portfolio_value_estimates).
regulation_unlicensed_activity(quantitative_analyst,
    ↪ predictive_valuation_models).
regulation_unlicensed_activity(avm_developer,
    ↪ algorithmic_residential_price_estimates).

regulation_licensed_activity(licensed_appraiser,
    ↪ market_value_for_federally_related_transaction).
regulation_licensed_activity(licensed_appraiser,
    ↪ formal_opinion_of_value_for_compensation).
regulation_licensed_activity(licensed_appraiser,
    ↪ appraisal_for_regulated_mortgage_lending).
regulation_licensed_activity(licensed_appraiser,
    ↪ appraisal_represented_as_such_under_state_law).

regulation_legal_exposure_condition(holding_self_out_as_licensed_appraiser).
regulation_legal_exposure_condition(preparing_appraisal_as_defined_by_state_law).
regulation_legal_exposure_condition(involved_in_transaction_requiring_licensed_appraiser).

%% --- The test itself: does the TASK require an appraiser's license? ---
requires_license(Task) :-
    regulation_triggers_licensure(Task), !.
requires_license(Task) :-
    regulation_licensed_activity(_, Task), !.

exempt_from_appraisal_license(Task) :-
    regulation_unlicensed_activity(_, Task),
    \+ requires_license(Task).

:- dynamic activity_satisfies_definition/1.

activity_satisfies_definition(Task) :-
    regulation_licensed_activity(_, Task).

regulated_appraisal(Task) :-
    activity_satisfies_definition(Task),
    requires_license(Task).

unregulated_valuation(Task) :-
    regulation_unlicensed_activity(_, Task),
    \+ regulated_appraisal(Task).

%% Meta-tension: the regulatory asymmetry the passage closes on.
regulatory_asymmetry(current_market_value_for_lending,
    future_oriented_valuation_for_investment).
regulatory_asymmetry(heavily_regulated_appraisal_practice,
    lightly_regulated_quantitative_practice).
regulatory_asymmetry(consumer_and_lending_protection_focus,
    investment_advisory_focus).

regulatory_asymmetry_origin(banking_regulation).
regulatory_asymmetry_origin(consumer_protection).
regulatory_asymmetry_origin(legal_evidentiary_concerns).
regulatory_asymmetry_origin_disclaimer(

```

```

    'Not based on any exclusive claim that appraisers alone possess valuation
    ↪ expertise.').

%% =====
%% TIER 2 -- DESCRIPTIVE TENDENCIES (NOT the boundary)
%%
%% Overlapping populations; correlates of who typically lands on each
%% side. None of these determine licensure -- Tier 1 does.
%% =====

%% --- Axis: valuation approach (descriptive populations) ---
valuation_approach(appraisal).
valuation_approach(quant_financial).

scope_profession(appraiser).
scope_profession(quantitative_analyst).
scope_profession(financial_analyst).
scope_profession(investment_analyst).
scope_profession(portfolio_manager).
scope_profession(economist).
scope_profession(investment_strategist).
scope_profession(financial_forecaster).

scope_orientation(appraisal, evidentiary).
scope_orientation(quant_financial, probabilistic).

%% --- Axis: temporal scope (its OWN viewpoint, distinct from approach) ---
temporal_scope(retrospective).
temporal_scope(current).
temporal_scope(prospective_point).
temporal_scope(future_range_horizon).

temporal_mode(point_in_time).
temporal_mode(distribution_over_horizon).

%% Typical (defeasible) pairing of approach to temporal mode -- a LINK
%% between two distinct axes, not an identity.
scope_temporal_focus(appraisal, point_in_time).
scope_temporal_focus(quant_financial, distribution_over_horizon).

%% --- Axis: output form ---
scope_output_form(appraisal, point_estimate).
scope_output_form(appraisal, narrow_range).
scope_output_form(quant_financial, probability_distribution).
scope_output_form(quant_financial, range_of_outcomes).

%% --- Axis: skills / methodological toolkit (training, NOT intelligence) ---
scope_skill(mars_without_overfitting).
scope_skill(fit_vs_generalization_balance). % balancing R2 against CV-R2
scope_skill(algorithmic_understanding).
scope_skill(handling_structured_concepts).
scope_skill(discerning_fine_distinctions_in_meaning).
scope_skill(assimilating_new_methods).

%% Skills typically demanded by advanced quantitative valuation; an
%% appraisal license neither requires nor certifies them.
scope_skill_demanded_by(quant_financial, mars_without_overfitting).
scope_skill_demanded_by(quant_financial, fit_vs_generalization_balance).

```

```

scope_skill_demanded_by(quant_financial, algorithmic_understanding).
scope_skill_demanded_by(quant_financial, handling_structured_concepts).
scope_skill_demanded_by(quant_financial, discerning_fine_distinctions_in_meaning).
scope_skill_demanded_by(quant_financial, assimilating_new_methods).

%% --- Inputs the appraiser weighs to support a value at the effective date ---
scope_inputs_considered(appraisal, current_market_evidence).
scope_inputs_considered(appraisal, observable_transactions).
scope_inputs_considered(appraisal, prevailing_market_conditions).
scope_inputs_considered(appraisal, recognized_methodologies).
scope_inputs_considered(appraisal, market_trends).
scope_inputs_considered(appraisal, anticipated_supply_demand).
scope_inputs_considered(appraisal, depreciation).
scope_inputs_considered(appraisal, capitalization_rates).
scope_inputs_considered(appraisal, absorption_patterns).
scope_inputs_considered(appraisal, expectations_already_priced_in).

%% --- Bounded future-oriented contexts that remain within appraisal ---
scope_future_oriented_context(income_property, future_income_streams).
scope_future_oriented_context(income_property, vacancy_rates).
scope_future_oriented_context(income_property, expenses).
scope_future_oriented_context(income_property, reversion_values).
scope_future_oriented_context(income_property, dcf_analysis).
scope_future_oriented_context(development, prospective_stabilized_value).
scope_future_oriented_context(conservation, future_damages).
scope_future_oriented_context(environmental, future_damages).
scope_future_oriented_context(insurance, future_damages).
scope_future_oriented_context(litigation, future_damages).
scope_future_oriented_context(eminent_domain, future_damages).
scope_future_oriented_context(general, remaining_economic_life).
scope_future_oriented_context(general, anticipated_market_reactions).
scope_future_oriented_context(market_analysis, price_trend_direction).

%% --- Professional constraints on any future-oriented component ---
scope_professional_constraint(supportability).
scope_professional_constraint(market_evidence).
scope_professional_constraint(extraordinary_assumptions_or_hypothetical_conditions |
→ _disclosed).
scope_professional_constraint(avoidance_of_unsupported_speculation).

%% --- Contrasts across the descriptive axes:
→ scope_contrast(A,B,Dimension,Distinction) ---
scope_contrast(appraisal, quant_financial, orientation,
evidentiary_vs_probabilistic).
scope_contrast(appraisal, quant_financial, temporal_mode,
point_in_time_vs_distribution_over_horizon).
scope_contrast(appraisal, quant_financial, output_form,
point_or_narrow_range_vs_probability_distribution).
scope_contrast(appraisal, quant_financial, purpose,
market_value_vs_investment_worth).
scope_contrast(appraisal, quant_financial, skill_set,
appraisal_licensing_toolkit_vs_advanced_statistical_toolkit).
scope_contrast(appraisal, quant_financial, typical_methods,
sales_comparison_cost_income_vs_simulation_option_pricing_var).

%% --- The orienting question each population asks (verbatim from the passage) ---
scope_question(appraisal,

```

```

    'What is the most supportable market value indication as of the specified
    → effective date, given current evidence and defined assumptions?').
scope_question(quant_financial,
    'What range of outcomes is likely over a future period, with what
    → probabilities, and how should resources be allocated?').

%% --- Where the two converge (appraisal borrows distributional methods) ---
scope_convergence_area(complex_commercial_valuation).
scope_convergence_area(mass_appraisal).
scope_convergence_area(automated_valuation_models).
scope_convergence_area(institutional_portfolio_analysis).

```

5 O valor de mercado $\times$ LTV como proxy do valor de execução hipotecária

As seções anteriores expuseram para que serve a avaliação, o que ela produz e onde se situam seus limites doutrinários e regulatórios; esta seção alcança a tensão para a qual elas vinham convergindo.

Uma das realidades práticas mais profundas subjacentes ao crédito hipotecário e à análise de garantias é que, em sua função econômica efetiva, uma avaliação de valor de mercado utilizada na subscrição hipotecária muitas vezes serve como proxy implícita da recuperabilidade esperada da garantia ao longo de algum período futuro de detenção, ainda que formalmente a avaliação seja enquadrada como um parecer de valor de mercado presente em uma data-base específica.(Foster e Van Order 1984; Kau et al. 1995; Deng, Quigley e Van Order 2000)

Uma razão empréstimo-valor de 80% não é arbitrária. Historicamente, ela evoluiu em parte como uma reserva de risco destinada a proteger os credores contra:

- quedas moderadas de mercado,
- despesas de execução hipotecária,
- descontos de liquidação,
- custos de manutenção,
- despesas judiciais,
- juros acumulados,
- e a incerteza durante o tempo necessário para alienar a garantia após a inadimplência.(Foote, Gerardi e Willen 2008; Qi e Yang 2009)

Essa reserva é também a razão pela qual o próprio montante do empréstimo —o produto do valor de mercado pela razão empréstimo-valor permitida— pode ser lido como a estimativa operacional que o credor faz do valor recuperável em execução hipotecária: limitar o empréstimo a uma fração do valor de mercado é, com efeito, uma aposta de que o produto da venda forçada não ficará abaixo do saldo devedor. Quão grande *deveria* ser o deságio, no entanto, é bem mais difícil de precisar do que sugere a cifra redonda de 80%. As estimativas do desconto por execução hipotecária variam amplamente entre mercados, métodos e períodos; um estudo de vendas forçadas frequentemente citado o situa em cerca de 27% abaixo dos preços comparáveis entre partes independentes(Campbell, Giglio e Pathak 2011), enquanto outras análises —que corrigem por estado de conservação do imóvel, localização e seleção— relatam magnitudes materialmente distintas.(Pennington-Cross 2006; Clauret e Daneshvary 2009)

O limiar convencional de 80%, em contrapartida, é essencialmente um único parâmetro de alcance nacional, aplicado a mercados locais enormemente heterogêneos e a fases sucessivas do ciclo de crédito. Ele não consegue acompanhar o desconto de nenhum mercado em particular e, por isso, funciona como uma aproximação muito grosseira e de tamanho único, e não como um mapeamento preciso do valor de mercado para a recuperação em execução hipotecária. Se

algo se pode dizer, é que se trata de uma reserva delgada: um deságio de 20% situa-se *abaixo* do desconto de 27% que aquele estudo relata, de modo que o empréstimo convencional máximo pode exceder o produto bruto da venda forçada mesmo antes de subtraídos os custos de alienação enumerados acima.

Quando o patrimônio próprio do tomador é demasiado escasso para constituir a reserva, substitui-se por um seguro hipotecário. Isso é mais bem compreendido não como um botão de política pública, mas como um arranjo de alocação de riscos: um segmento de tomadores dispõe de poucos ativos, mas tem renda estável, e o risco incremental de inadimplência e de perda decorrente de emprestar-lhes acima de 80% de LTV é projetado e assumido por uma seguradora —de forma privada, por meio do seguro hipotecário privado em um empréstimo convencional de LTV elevado, ou pública, por meio da Federal Housing Administration (administração federal de habitação dos Estados Unidos)—. O mercado em que esse arranjo é rotineiro, contudo, é uma criação da política habitacional federal: as government-sponsored enterprises (empresas patrocinadas pelo governo) que dominam o crédito convencional estão proibidas por seus estatutos de adquirir empréstimos acima de 80% de LTV sem tal reforço de crédito (United States Code 1970). Os termos em que o seguro se encerra são igualmente escolhas institucionais, e não cálculos de risco, e diferem acentuadamente entre programas.¹

De modo que, embora o avaliador esteja oficialmente estimando o “valor de mercado atual”, o credor interpreta economicamente esse valor por meio de uma lente de risco voltada ao futuro. Com efeito, o credor está perguntando algo mais próximo de:

“Se este tomador vier a inadimplir em algum momento do futuro previsível, é provável que a garantia permaneça suficiente para cobrir a dívida em aberto e as perdas relacionadas à execução hipotecária?”

Essa é, inerentemente, uma pergunta orientada ao futuro.

Isso repousa sobre uma separação conceitual deliberada entre:

- o parecer do avaliador sobre o “valor de mercado atual”, e
- o uso prático que o credor faz desse parecer como reserva prospectiva contra a deterioração futura da garantia.

Histórica e juridicamente, os reguladores procuraram impedir que os avaliadores se tornassem previsores especulativos porque:

- a previsão especulativa contribuiu para crises de crédito passadas,
- os tribunais queriam avaliações vinculadas a evidências observáveis,
- e os reguladores bancários buscavam práticas de avaliação padronizadas e defensáveis.

Assim, a doutrina da avaliação evoluiu em direção a:

- valor presente,
- evidências atuais de mercado,
- datas-base definidas,
- e pressupostos restringidos.

Enquanto isso, credores, investidores e analistas de garantias utilizam rotineiramente essas mesmas avaliações em arcabouços probabilísticos de risco futuro.

Na realidade, a subscrição hipotecária é um sistema integrado que envolve:

¹O seguro hipotecário privado convencional pago pelo tomador extingue-se automaticamente a 78% de LTV e seu cancelamento pode ser requerido a 80%, nos termos da Homeowners Protection Act de 1998. (United States Congress 1998) Os prêmios da Federal Housing Administration são mais rigorosos: para empréstimos originados em 3 de junho de 2013 ou depois, com razão empréstimo-valor original superior a 90% (por exemplo, 3,5% de entrada), o prêmio anual vigora por toda a vida do empréstimo, enquanto nos empréstimos com LTV igual ou inferior a 90% ele é cancelado após onze anos. (U.S. Department of Housing and Urban Development 2013) Por isso, muitos tomadores da FHA só se livram do seguro hipotecário refinanciando para um empréstimo convencional próximo de 80% de LTV.

1. a avaliação de valor de mercado atual,
2. a estimativa de risco futuro,
3. a probabilidade de inadimplência,
4. o momento da execução hipotecária,
5. a severidade da liquidação,
6. as expectativas macroeconômicas,
7. e a análise de estresse da carteira.

O avaliador contribui formalmente, sobretudo, para o item (1), enquanto o credor combina internamente esse resultado com os demais.

A distinção torna-se difusa na prática. Se um credor não concederia um empréstimo a menos que acreditasse que a garantia provavelmente preservaria valor suficiente no futuro, então a avaliação funciona indiretamente como parte de um modelo de proteção do valor futuro, ainda que as normas de avaliação formalmente evitem descrevê-la assim.

Essa tensão tornou-se ainda mais pronunciada com:

- os AVMs,
- a analítica institucional de carteiras,
- os modelos de securitização,
- os sistemas de avaliação por aprendizado de máquina,
- a modelagem de capital de Basileia,
- e os testes de estresse hipotecário em larga escala.

A analítica moderna de garantias trata cada vez mais os valores dos imóveis de forma probabilística ao longo do tempo, e não como estimativas estáticas de ponto único. (Basel Committee on Banking Supervision 2017; Steurer, Hill e Pfeifer 2021; Board of Governors of the Federal Reserve System 2021)

Poder-se-ia argumentar que a regulação tradicional da avaliação reflete um compromisso histórico:

- os avaliadores fornecem âncoras disciplinadas de valor presente, baseadas em evidências,
- enquanto as instituições financeiras assumem a responsabilidade pela interpretação do risco futuro.

Se essa separação permanece plenamente coerente nos sistemas financeiros modernos é uma questão legítima e cada vez mais debatida.

Daí decorre uma implicação prática para a própria atividade de avaliação. A maioria dos avaliadores não tem consciência desse papel de avaliação da garantia —o uso que o credor faz do valor de mercado como reserva prospectiva raramente é visível de dentro do trabalho de avaliação— e, por isso, não trata suas conclusões de valor de mercado com a cautela quanto às tendências futuras de valor que esse papel exige. Reconhecer como o parecer de valor presente é efetivamente consumido é um primeiro passo para fechar essa lacuna, e uma área que esta exposição pretende abrir a trabalhos posteriores.

```
%% appraisal_proxy.pl
%%
%% A Prolog knowledge base encoding the thesis that market value
%% appraisals, while formally framed as present-value opinions, function
%% in practice as implicit proxies for future collateral recoverability
%% within mortgage underwriting and lender risk management.

:- module(appraisal_proxy,
    [ % --- central thesis ---
      proxy_thesis/1

      % --- LTV-as-buffer mechanism ---
```

```

, proxy_ltv_buffer_against/1
, proxy_historical_ltv_threshold/1

% --- formal vs practical interpretation ---
, proxy_formal_framing/1
, proxy_practical_interpretation/1
, proxy_underlying_question/1

% --- the artificial separation ---
, proxy_separation_party/2      % Party, Role
, proxy_doctrine_practice_gap/2 % FormalSide, PracticalSide

% --- historical compromise ---
, proxy_historical_reason/1
, proxy_doctrine_emphasis/1
, proxy_routine_practical_use/1

% --- modern intensifiers ---
, proxy_modern_driver/1
, proxy_modern_treatment/1

% --- responsibility allocation ---
, proxy_appraiser_responsibility/1
, proxy_lender_responsibility/1

% --- meta-claim about coherence ---
, proxy_open_question/1

% --- practical implication for appraisal practice ---
, proxy_practical_implication/1
, proxy_appraiser_blindspot/1
, proxy_recommended_practice/1
, proxy_further_work/1

% --- underwriting system view ---
, underwriting_component/2      % Number, Component
, underwriting_responsibility/2 % Component, Party
]).

:- discontiguous proxy_thesis/1.
:- discontiguous underwriting_component/2.

%% =====
%% PART 1: THE CENTRAL THESIS
%% =====

proxy_thesis(
  'A market value appraisal used in mortgage underwriting often serves as an
  ⇨ implicit proxy for expected collateral recoverability over a future holding
  ⇨ period.').
proxy_thesis(
  'The appraisal is formally framed as a present-market-value opinion as of a
  ⇨ specific effective date.').
proxy_thesis(
  'The lender economically interprets the appraisal through a forward-looking
  ⇨ risk lens.').
proxy_thesis(

```

```

    'There is a deliberate conceptual separation between the appraiser's
    ↪ current-market-value opinion and the lender's use of it as a forecast
    ↪ buffer.').
proxy_thesis(
    'The appraisal indirectly functions as part of a future-value protection model,
    ↪ even though appraisal standards formally avoid describing it that way.').
proxy_thesis(
    'Whether the separation between appraisal doctrine and lender practice remains
    ↪ fully coherent in modern financial systems is an open and increasingly
    ↪ debated question.').

%% =====
%% PART 2: THE LTV-AS-BUFFER MECHANISM
%% =====

proxy_historical_ltv_threshold(eighty_percent).

proxy_ltv_buffer_against(moderate_market_declines).
proxy_ltv_buffer_against(foreclosure_expenses).
proxy_ltv_buffer_against(liquidation_discounts).
proxy_ltv_buffer_against(carrying_costs).
proxy_ltv_buffer_against(legal_expenses).
proxy_ltv_buffer_against(accrued_interest).
proxy_ltv_buffer_against(uncertainty_during_disposition).

%% =====
%% PART 3: FORMAL FRAMING VS. PRACTICAL INTERPRETATION
%% =====

proxy_formal_framing(present_market_value_opinion).
proxy_formal_framing(opinion_as_of_specific_effective_date).
proxy_formal_framing(grounded_in_current_market_evidence).

proxy_practical_interpretation(forward_looking_risk_lens).
proxy_practical_interpretation(forecast_buffer_against_future_impairment).
proxy_practical_interpretation(input_to_future_collateral_recoverability_estimate).

proxy_underlying_question(
    'If this borrower defaults sometime in the foreseeable future, is the
    ↪ collateral likely to remain sufficient to cover the outstanding debt and
    ↪ foreclosure-related losses?').

%% =====
%% PART 4: THE ARTIFICIAL SEPARATION
%% =====

proxy_separation_party(appraiser, opinion_of_current_market_value).
proxy_separation_party(lender, forecast_buffer_against_future_impairment).
proxy_separation_party(appraiser, evidence_based_present_value_anchor).
proxy_separation_party(lender, future_risk_interpretation).

proxy_doctrine_practice_gap(present_market_value_opinion,
                             implicit_future_value_proxy).
proxy_doctrine_practice_gap(opinion_as_of_effective_date,
                             forecast_over_foreseeable_holding_period).
proxy_doctrine_practice_gap(evidence_based_present_anchor,
                             future_risk_buffer).
proxy_doctrine_practice_gap(static_single_point_estimate,

```

```

        probabilistic_value_over_time).

%% =====
%% PART 5: HISTORICAL COMPROMISE
%% =====

proxy_historical_reason(speculative_forecasting_contributed_to_past_crises).
proxy_historical_reason(courts_required_observable_evidence).
proxy_historical_reason(regulators_sought_standardized_defensible_practices).

proxy_doctrine_emphasis(present_value).
proxy_doctrine_emphasis(current_market_evidence).
proxy_doctrine_emphasis(defined_effective_dates).
proxy_doctrine_emphasis(constrained_assumptions).

proxy_routine_practical_use(lenders_use_in_probabilistic_future_risk_frameworks).
proxy_routine_practical_use(investors_use_in_probabilistic_future_risk_frameworks).
proxy_routine_practical_use(collateral_analysts_use_in_probabilistic_future_risk_f
↪ rameworks).

%% =====
%% PART 6: MODERN INTENSIFIERS
%% =====

proxy_modern_driver(avms).
proxy_modern_driver(institutional_portfolio_analytics).
proxy_modern_driver(securitization_models).
proxy_modern_driver(machine_learning_valuation_systems).
proxy_modern_driver(basel_capital_modeling).
proxy_modern_driver(large_scale_mortgage_stress_testing).

proxy_modern_treatment(probabilistic_over_time).
proxy_modern_treatment(distributional_rather_than_point_estimate).
proxy_modern_treatment(dynamic_rather_than_static).

%% =====
%% PART 7: ALLOCATION OF RESPONSIBILITY UNDER THE COMPROMISE
%% =====

proxy_appraiser_responsibility(disciplined_present_value_anchor).
proxy_appraiser_responsibility(evidence_based_valuation).
proxy_appraiser_responsibility(current_market_value_opinion).

proxy_lender_responsibility(future_risk_interpretation).
proxy_lender_responsibility(default_probability_estimation).
proxy_lender_responsibility(foreclosure_timing_estimation).
proxy_lender_responsibility(liquidation_severity_estimation).
proxy_lender_responsibility(portfolio_stress_analysis).
proxy_lender_responsibility(macroeconomic_expectation_integration).

%% =====
%% PART 8: OPEN QUESTION (META-CLAIM)
%% =====

proxy_open_question(
    'Whether the separation between disciplined present-value appraisal and
    ↪ lender-side future-risk interpretation remains fully coherent in modern
    ↪ financial systems.').

```

```

proxy_open_question(
    'Whether modern probabilistic-over-time collateral analytics make the
    ↪ formal/practical separation untenable in some sectors.').

%% =====
%% PART 9: THE INTEGRATED MORTGAGE UNDERWRITING SYSTEM
%% =====

underwriting_component(1, current_market_valuation).
underwriting_component(2, future_risk_estimation).
underwriting_component(3, default_probability).
underwriting_component(4, foreclosure_timing).
underwriting_component(5, liquidation_severity).
underwriting_component(6, macroeconomic_expectations).
underwriting_component(7, portfolio_stress_analysis).

underwriting_responsibility(current_market_valuation, appraiser).
underwriting_responsibility(future_risk_estimation, lender).
underwriting_responsibility(default_probability, lender).
underwriting_responsibility(foreclosure_timing, lender).
underwriting_responsibility(liquidation_severity, lender).
underwriting_responsibility(macroeconomic_expectations, lender).
underwriting_responsibility(portfolio_stress_analysis, lender).

%% =====
%% PART 10: PRACTICAL IMPLICATION FOR APPRAISAL PRACTICE
%% =====

proxy_practical_implication(
    'Most appraisers are not aware that market value functions, via the
    ↪ loan-to-value ratio, as a proxy for future foreclosure value.').
proxy_practical_implication(
    'The lender''s forward-looking use of market value is rarely visible from
    ↪ within the appraisal assignment.').
proxy_practical_implication(
    'Appraisers therefore do not treat market value conclusions with the caution
    ↪ toward future value trends that the collateral-valuation role warrants.').

proxy_appraiser_blindspot(collateral_valuation_role).
proxy_appraiser_blindspot(forward_looking_use_of_market_value).
proxy_appraiser_blindspot(market_value_as_ltv_proxy_for_foreclosure_value).

proxy_recommended_practice(recognize_how_present_value_opinion_is_consumed).
proxy_recommended_practice(treat_market_value_with_caution_re_future_trends).

proxy_further_work(closing_the_doctrine_practice_gap).
proxy_further_work(equipping_appraisers_for_the_collateral_valuation_perspective).

```

6 Conclusão

As cinco passagens codificadas nas bases de conhecimento anteriores descrevem uma única profissão a partir de cinco pontos de vista complementares. Tomada isoladamente, cada passagem se lê como uma exposição autossuficiente: os interesses sociais e de ordem pública que carregam, as funções econômicas que as avaliações sustentam, seu papel operacional na análise do risco da garantia, o limite regulatório que as separa do trabalho quantitativo adjacente e a tensão entre seu enquadramento formal e seu uso prático. Tomadas em conjunto, e com a

estrutura explicitada em Prolog, elas descrevem algo mais interessante — uma profissão cuja doutrina oficial e cuja função econômica foram deliberadamente mantidas apartadas, e cuja coerência interna é cada vez mais posta à prova pelas ferramentas analíticas hoje aplicadas a seus produtos.

As três primeiras bases de conhecimento compartilham um esquema comum: `effect(Quality, Horizon, Domain, Effect)`. Cada uma enquadra a qualidade da avaliação como um insumo causal cujas consequências se propagam por horizontes de curto, médio e longo prazo, diferindo apenas quanto ao domínio pelo qual essas consequências são lidas. A lente econômica as lê pelos mercados de crédito, a alocação de investimentos, a tributação e a formação de capital; a lente social as lê pela equidade, a confiança pública e os desfechos comunitários; a lente da garantia as lê pela concessão de crédito, a gestão de carteiras e a estabilidade do sistema financeiro. O esquema compartilhado não é uma coincidência. Ele reflete uma afirmação substantiva que atravessa as três passagens: a qualidade do trabalho de avaliação tem consequências, essas consequências escalam com o tempo, e os mesmos modos subjacentes de falha se manifestam de maneira distinta conforme os interesses postos em primeiro plano. Os átomos compartilhados no lado dos argumentos dos predicados — em particular os vocabulários de qualidade e de horizonte — tornam os três vocabulários combináveis, de modo que uma única falha, como a superavaliação sistemática, possa ser rastreada simultaneamente pela precificação econômica equivocada, pela iniquidade social e pela mensuração incorreta da garantia.

A quarta base de conhecimento muda de registro. Ela não descreve consequências da qualidade da avaliação; descreve o que a avaliação *é*, metodológica e juridicamente. Seus predicados são taxonômicos, e não causais, e sua maquinaria inferencial responde a perguntas sobre jurisdição, e não sobre impacto. Essa mudança é necessária porque o enquadramento consequencial das três primeiras passagens pressupõe uma definição da atividade avaliada, e essa definição é, ela própria, contestada. Os predicados de escopo registram o que os avaliadores fazem — pareceres de valor presente baseados em evidências, com componentes voltados ao futuro estreitamente delimitados —, enquanto os predicados de regulação registram as condições sob as quais essa atividade recai na prática licenciada. As regras de encadeamento progressivo (`requires_license/1`, `exempt_from_appraisal_license/1`, `regulated_appraisal/1`) fornecem um mecanismo para perguntar se um dado trabalho de avaliação cai dentro ou fora do perímetro regulado, pergunta que as três primeiras bases de conhecimento supõem já respondida.

A quinta base de conhecimento avança uma tese. Onde as demais descrevem, classificam ou rastreiam consequências, a base de conhecimento da proxy argumenta que a cuidadosa separação doutrinária entre a avaliação a valor presente e a previsão do valor futuro — a separação que a base de conhecimento da regulação documenta — está em tensão com a função econômica efetiva das avaliações no crédito hipotecário. A convenção de 80% de razão empréstimo-valor, os sete riscos específicos contra os quais ela funciona como reserva, a pergunta subjacente do credor registrada textualmente (“Se este tomador vier a inadimplir em algum momento do futuro previsível, é provável que a garantia permaneça suficiente . . .”) e o sistema integrado de subscrição de sete componentes sustentam todos a mesma afirmação: a avaliação é formalmente um parecer de valor presente e economicamente uma reserva de valor futuro. O compromisso histórico que produziu esse arranjo — os avaliadores como âncoras disciplinadas de valor presente, os credores como o locus da interpretação do risco futuro — foi uma resposta defensável às condições sob as quais surgiu o licenciamento da avaliação, mas os intensificadores modernos enumerados na passagem (os AVMs, os sistemas de avaliação por aprendizado de máquina, a modelagem de capital de Basileia, a analítica de securitização, os testes de estresse em larga escala) tratam os valores dos imóveis de forma probabilística ao longo do tempo. Essas ferramentas trazem camadas doutrinariamente separadas de volta a um único fluxo computacional, e a separação conceitual entre camadas torna-se, por isso, mais difícil de sustentar.

Se essa reconciliação é alcançável, e em que termos, é a questão com que se encerra a passagem sobre a proxy. As bases de conhecimento não a respondem. Fornecem, contudo, uma

exposição explícita e consultável das considerações relevantes para respondê-la: as consequências da qualidade da avaliação ao longo de horizontes e domínios, o limite jurisdicional que determina quais consequências recaem dentro da prática regulada e as razões estruturais pelas quais esse limite está hoje sob pressão. Essa exposição é o alicerce sobre o qual podem ser construídos argumentos mais específicos — sobre metodologia, sobre a reforma das normas, sobre a relação entre a prática de avaliação e o sistema mais amplo da avaliação quantitativa.

Notas sobre a representação do conhecimento

Vale registrar explicitamente três padrões estruturais, uma vez que emergiram do processo de codificação e não de qualquer passagem isolada.

Primeiro, a família de bases de conhecimento não é plana. Três das cinco compartilham um esquema; uma fornece as qualificações jurisdicionais que determinam quando esse esquema se aplica; uma avança uma tese sobre a lacuna entre o enquadramento formal e a função econômica. Uma camada de ligação que combine todas as cinco não as trata como cinco colunas de uma mesma matriz. Trata as três primeiras como descritoras de consequências, a quarta como delimitadora de onde essas descrições são operativas e a quinta como identificadora de onde a própria delimitação está sob pressão. Essa estrutura assimétrica é uma virtude, e não um defeito: ela espelha a organização real da prática profissional, na qual afirmações descritivas, afirmações jurisdicionais e afirmações críticas ocupam papéis epistêmicos genuinamente distintos.

*Segundo, a disciplina de prefixos que distingue os módulos (*economic_*, *social_*, *collateral_*, *scope_/regulation_*, *proxy_/underwriting_*) só faz trabalho real onde há risco de colisão.* Os predicados com contrapartes diretas entre módulos justificam seu prefixo; os predicados exclusivos de um único domínio conceitual, não. O vocabulário de valores compartilhado no lado dos argumentos — os átomos de qualidade, horizonte e domínio — é precisamente o que permite aos módulos comunicarem-se, e prefixar esses valores frustraria isso. O padrão que emergiu é geral: prefixo nos predicados, átomos compartilhados nos argumentos, com prefixação seletiva em vez de uniforme onde o predicado é exclusivo de uma lente.

Terceiro, a tensão central identificada na base de conhecimento da proxy é estreitamente relacionada, mas distinta, da assimetria regulatória identificada na base de conhecimento de escopo/regulação. As duas são registradas por predicados paralelos de dois lugares — *regulatory_asymmetry/2* e *proxy_doctrine_practice_gap/2* — com nomes deliberadamente diferentes, porque descrevem eixos distintos do mesmo arranjo subjacente. A assimetria regulatória diz respeito a quem é regulado e com que intensidade; a lacuna entre doutrina e prática diz respeito ao que se afirma formalmente sobre o produto do avaliador *versus* o que dele se faz economicamente. As duas tensões reforçam-se mutuamente. A regulação empurra os avaliadores em direção à doutrina do valor presente; a função econômica puxa seu produto em direção ao uso voltado ao valor futuro. O quadro resultante é o de uma profissão cuja autodescrição oficial é cada vez mais difícil de conciliar com os fluxos analíticos que seus produtos alimentam.

Referências

- Agarwal, Sumit, Itzhak Ben-David e Vincent Yao (2015). “Collateral Valuation and Borrower Financial Constraints: Evidence from the Residential Real Estate Market”. Em: *Management Science* 61.9, pp. 2220–2240. DOI: [10.1287/mnsc.2014.2002](https://doi.org/10.1287/mnsc.2014.2002).
- Alexandrov, Alexei, Laurie Goodman e Michael Neal (jan. de 2023). *Reengineering the Appraisal Process: Better Leveraging Both Automated Valuation Models and Manual Appraisals*. Rel. técn. Urban Institute. URL: <https://www.urban.org/sites/default/files/2023-01/Reengineering%20the%20Appraisal%20Process.pdf>.
- Basel Committee on Banking Supervision (2017). *Basel III: Finalising Post-Crisis Reforms*. Standards d424. Bank for International Settlements. URL: <https://www.bis.org/bcbs/publ/d424.pdf>.

- Board of Governors of the Federal Reserve System (mar. de 2021). *Dodd-Frank Act Stress Test 2021: Supervisory Stress Test Methodology*. Rel. técn. Board of Governors of the Federal Reserve System. URL: <https://www.federalreserve.gov/publications/files/2021-april-supervisory-stress-test-methodology.pdf>.
- Campbell, John Y., Stefano Giglio e Parag Pathak (2011). “Forced Sales and House Prices”. Em: *American Economic Review* 101.5, pp. 2108–2131. DOI: [10.1257/aer.101.5.2108](https://doi.org/10.1257/aer.101.5.2108).
- Clauretie, Terrence M. e Nasser Daneshvary (2009). “Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time”. Em: *Real Estate Economics* 37.1, pp. 43–67. DOI: [10.1111/j.1540-6229.2009.00235.x](https://doi.org/10.1111/j.1540-6229.2009.00235.x).
- Crosby, Neil, Colin M. Lizieri e Patrick M. McAllister (2010). “Means, Motive and Opportunity? Disentangling Client Influence on Performance Measurement Appraisals”. Em: *Journal of Property Research* 27.2, pp. 181–201. DOI: [10.1080/09599916.2010.499014](https://doi.org/10.1080/09599916.2010.499014).
- Deng, Yongheng, John M. Quigley e Robert Van Order (2000). “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options”. Em: *Econometrica* 68.2, pp. 275–307. DOI: [10.1111/1468-0262.00110](https://doi.org/10.1111/1468-0262.00110).
- Diaz Julian, III e Marvin L. Wolverton (1998). “A Longitudinal Examination of the Appraisal Smoothing Hypothesis”. Em: *Real Estate Economics* 26.2, pp. 349–358. DOI: [10.1111/1540-6229.00749](https://doi.org/10.1111/1540-6229.00749).
- Ebrahim, M. Shahid e Sikandar Hussain (2010). “Financial Development and Asset Valuation: The Special Case of Real Estate”. Em: *Journal of Banking & Finance* 34.1, pp. 150–162. DOI: [10.1016/j.jbankfin.2009.07.011](https://doi.org/10.1016/j.jbankfin.2009.07.011).
- Eriksen, Michael D. et al. (2019). “The Influence of Contract Prices and Relationships on Appraisal Bias”. Em: *Journal of Urban Economics* 111, pp. 132–143. DOI: [10.1016/j.jue.2019.03.001](https://doi.org/10.1016/j.jue.2019.03.001).
- Foote, Christopher L., Kristopher Gerardi e Paul S. Willen (2008). “Negative Equity and Foreclosure: Theory and Evidence”. Em: *Journal of Urban Economics* 64.2, pp. 234–245. DOI: [10.1016/j.jue.2008.07.006](https://doi.org/10.1016/j.jue.2008.07.006).
- Foster, Chester e Robert Van Order (1984). “An Option-Based Model of Mortgage Default”. Em: *Housing Finance Review* 3.4, pp. 351–372.
- Friedman, Jerome H. (1991). “Multivariate Adaptive Regression Splines”. Em: *The Annals of Statistics* 19.1, pp. 1–67. DOI: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Gallimore, Paul (1996). “Confirmation Bias in the Valuation Process: A Test for Corroborating Evidence”. Em: *Journal of Property Research* 13.4, pp. 261–273. DOI: [10.1080/095999196368781](https://doi.org/10.1080/095999196368781).
- Geltner, David M. (1991). “Smoothing in Appraisal-Based Returns”. Em: *The Journal of Real Estate Finance and Economics* 4.3, pp. 327–345. DOI: [10.1007/BF00161933](https://doi.org/10.1007/BF00161933).
- Ghent, Andra C., Walter N. Torous e Rossen I. Valkanov (2019). “Commercial Real Estate as an Asset Class”. Em: *Annual Review of Financial Economics* 11, pp. 153–171. DOI: [10.1146/annurev-financial-110118-123121](https://doi.org/10.1146/annurev-financial-110118-123121).
- Griffin, John M. e Gonzalo Maturana (2016). “Who Facilitated Misreporting in Securitized Loans?” Em: *The Review of Financial Studies* 29.2. Evidence of appraisal inflation and its consequences, pp. 384–419. DOI: [10.1093/rfs/hhv130](https://doi.org/10.1093/rfs/hhv130).
- Hendershott, Patric H. e Edward J. Kane (jul. de 1992). *Office Market Values During the Past Decade: How Distorted Have Appraisals Been?* NBER Working Paper 4128. Available at SSRN: <https://ssrn.com/abstract=226786>. National Bureau of Economic Research. DOI: [10.3386/w4128](https://doi.org/10.3386/w4128). URL: <https://www.nber.org/papers/w4128>.
- (1995). “U.S. Office Market Values During the Past Decade: How Distorted Have Appraisals Been?” Em: *Real Estate Economics* 23.2, pp. 101–116. DOI: [10.1111/1540-6229.00660](https://doi.org/10.1111/1540-6229.00660).
- Hull, John C. (2018). *Options, Futures, and Other Derivatives*. 10^a ed. New York: Pearson. ISBN: 978-0-13-447208-9.

- International Association of Assessing Officers (2017). *Standard on Mass Appraisal of Real Property*. International Association of Assessing Officers, Kansas City, MO. URL: https://www.iaao.org/wp-content/uploads/Standard_on_Mass_Appraisal.pdf.
- International Valuation Standards Council (2025). *International Valuation Standards (IVS)*. International Valuation Standards Council, London. Published 31 January 2024; effective 31 January 2025. URL: <https://www.ivsc.org/>.
- K'Akumu, Owiti A. e James E. Larsen (2024). "A Primer on Regulations and the Practice of Residential Property Appraisal". Em: *Journal of Real Estate Practice and Education*. Focus on regulatory and public interest aspects. DOI: [10.1080/15214842.2024.2377869](https://doi.org/10.1080/15214842.2024.2377869).
- Kau, James B. et al. (1995). "The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment". Em: *The Journal of Real Estate Finance and Economics* 11.1, pp. 5–36. DOI: [10.1007/BF01097934](https://doi.org/10.1007/BF01097934).
- LaCour-Little, Michael e Stephen Malpezzi (2003). "Appraisal Quality and Residential Mortgage Default: Evidence from Alaska". Em: *The Journal of Real Estate Finance and Economics* 27.2, pp. 211–233. DOI: [10.1023/A:1024728420837](https://doi.org/10.1023/A:1024728420837).
- Lee, Ming-Hsiang, Chien-Wen Peng e Hui-Fen Liao (2020). "An Analysis of Objectivity in the Real Estate Appraisal Process". Em: *International Real Estate Review* 23.4, pp. 483–504.
- Mayer, Yanling G. e Frank E. Nothaft (2022). "Appraisal Overvaluation: Evidence of Price Adjustment Bias in Sales Comparisons". Em: *Real Estate Economics* 50.3, pp. 862–881. DOI: [10.1111/1540-6229.12351](https://doi.org/10.1111/1540-6229.12351).
- Murphy, Edward V. (2012). *Regulation of Real Estate Appraisers*. CRS Report RS22953. Congressional Research Service. URL: <https://sgp.fas.org/crs/misc/RS22953.pdf>.
- Office of the Comptroller of the Currency and Board of Governors of the Federal Reserve System and Federal Deposit Insurance Corporation and National Credit Union Administration and Consumer Financial Protection Bureau and Federal Housing Finance Agency (2024). *Quality Control Standards for Automated Valuation Models*. Final Rule, 89 Fed. Reg. 64538 (Aug. 7, 2024). URL: <https://www.federalregister.gov/documents/2024/08/07/2024-16197/quality-control-standards-for-automated-valuation-models>.
- Pennington-Cross, Anthony (2006). "The Value of Foreclosed Property". Em: *Journal of Real Estate Research* 28.2, pp. 193–214. DOI: [10.1080/10835547.2006.12091177](https://doi.org/10.1080/10835547.2006.12091177).
- Qi, Min e Xiaolong Yang (2009). "Loss Given Default of High Loan-to-Value Residential Mortgages". Em: *Journal of Banking & Finance* 33.5, pp. 788–799. DOI: [10.1016/j.jbankfin.2008.09.010](https://doi.org/10.1016/j.jbankfin.2008.09.010).
- Quan, Daniel C. e John M. Quigley (1991). "Price Formation and the Appraisal Function in Real Estate Markets". Em: *The Journal of Real Estate Finance and Economics* 4.2, pp. 127–146. DOI: [10.1007/BF00173120](https://doi.org/10.1007/BF00173120).
- Steurer, Miriam, Robert J. Hill e Norbert Pfeifer (2021). "Metrics for Evaluating the Performance of Machine Learning Based Automated Valuation Models". Em: *Journal of Property Research* 38.2, pp. 99–129. DOI: [10.1080/09599916.2020.1858937](https://doi.org/10.1080/09599916.2020.1858937).
- The Appraisal Foundation (2024). *Uniform Standards of Professional Appraisal Practice (USPAP), 2024 Edition*. The Appraisal Foundation, Appraisal Standards Board, Washington, DC. Effective January 1, 2024. URL: <https://www.appraisalfoundation.org/>.
- U.S. Department of Housing and Urban Development (2013). *Revision of Federal Housing Administration Policies Concerning Cancellation of the Annual Mortgage Insurance Premium*. Mortgagee Letter 2013-04, Federal Housing Administration. Effective June 3, 2013; the annual MIP continues for the life of loans with original LTV above 90%, and for 11 years at or below 90%. URL: <https://www.hud.gov/sites/documents/13-04ml.pdf>.
- United States Code (1970). *Federal National Mortgage Association and Federal Home Loan Mortgage Corporation Charter Acts*. 12 U.S.C. § 1717(b)(2) (Fannie Mae) and 12 U.S.C. § 1454(a)(2) (Freddie Mac). Prohibit GSE purchase of conventional mortgages exceeding 80%

LTV absent credit enhancement (private mortgage insurance, seller recourse/repurchase, or a retained seller participation).

United States Congress (1989). *Financial Institutions Reform, Recovery, and Enforcement Act of 1989*. Pub. L. No. 101-73, 103 Stat. 183; Title XI codified at 12 U.S.C. §§ 3331–3356. URL: <https://www.govinfo.gov/content/pkg/STATUTE-103/pdf/STATUTE-103-Pg183.pdf>.

— (1998). *Homeowners Protection Act of 1998*. Pub. L. No. 105-216, 112 Stat. 897; codified at 12 U.S.C. §§ 4901–4910. Automatic termination of borrower-paid private mortgage insurance at 78% LTV; cancellation may be requested at 80%.

Residual Constraint Approach (RCA): estrutura metodológica e protocolo

Tradução para o português (Brasil) da versão atualizada do artigo originalmente publicado no Zenodo em 5 de março de 2025. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

William Bert Craytor, SRA

Volume 1, Número 1 · Verão de 2026

Publicado em 1º de julho de 2026

DOI: [10.67330/kgebx46](https://doi.org/10.67330/kgebx46)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da [licença Creative Commons Atribuição 4.0](#).

Resumo. Uma limitação importante dos métodos convencionais de avaliação está na sua dependência de técnicas manuais ultrapassadas, como a análise de pares comparativos (*matched pairs*), que ignoram as contribuições de valor das características do imóvel como se elas não existissem: o *Sales Grid* (quadro de homogeneização) tradicional não possui coluna para contribuições de valor, apenas para ajustes. Na verdade, as contribuições de valor fornecem a base necessária para as restrições matemáticas que impedem a supervalorização e a subvalorização pelos avaliadores tradicionais.

O Residual Constraint Approach (RCA, pt.: abordagem por restrição residual) faz avançar o método comparativo direto de dados de mercado (*Sales Comparison Approach*, SCA) tradicional ao integrar um processo de avaliação multifásico que emprega Multivariate Adaptive Regression Splines (MARS) juntamente com restrições matemáticas rigorosas sobre as contribuições de valor das características — em particular as características não medidas: variáveis latentes como estética, projeto, estado de conservação e padrão construtivo, cujos valores costumam ficar sujeitos ao juízo subjetivo e ao viés do avaliador tradicional.

As restrições matemáticas do RCA apoiam-se na aplicação especializada da regressão MARS para estimar as contribuições de valor das características mensuráveis do imóvel ao preço de venda. Contudo, essa estimativa normalmente responde por apenas cerca de 80% do preço de venda efetivo, excluindo características avaliadas subjetivamente, como projeto, utilidade funcional, estado de conservação e padrão construtivo. Na San Francisco Bay Area, o resíduo — que representa esses componentes subjetivos — gira em torno de 20% do valor total do imóvel. A cifra de 80/20 reflete a experiência do autor na San Francisco Bay Area sob condições típicas de dados e não deve ser lida como uma constante invariante entre mercados. Em mercados nos quais o R^2 alcançável é menor, a resposta da estrutura metodológica consiste em identificar e medir os determinantes de valor não medidos — a distância ao oceano em comunidades litorâneas, a altimetria em bairros de encosta e variáveis semelhantes coletadas de fontes externas quando os dados do MLS não as fornecem — em vez de aceitar a limitação e reportar faixas de erro mais amplas. Embora os resíduos do MARS sejam frequentemente tratados como erros de estimação em outros contextos, na avaliação de imóveis eles funcionam como medida indireta do valor das variáveis latentes. Ainda que isso possa parecer pouco prático à primeira vista, uma demonstração estabelece que o resíduo pode ser decomposto de maneira significativa em componentes descritivos, desde que suas contribuições somem o resíduo — sem afetar o resultado final da avaliação.

Esta metodologia resulta de duas décadas de aplicação empírica por parte do autor, desenvolvida por meio de implementações iterativas do MARS em ambientes R e Python.

Palavras-chave: avaliação de imóveis, avaliação residencial, Residual Constraint Approach, MARS, Multivariate Adaptive Regression Splines, método comparativo direto de dados de mercado, análise de resíduos, variáveis latentes, avaliação automatizada

Sumário

1	Introdução	4
2	Antecedentes	6
3	Engenheiro de Avaliações	9
4	Premissas	10
5	Fluxo de Trabalho do RCA	13
5.1	Tipos de Análise MARS	13
5.2	Fluxo de Processamento	13
5.3	Fluxo de Dados	14
5.3.1	Requisitos de Dados para o MARS	15
5.3.2	Protocolos: Listagens de Vendas Fechadas	16
5.4	Estrutura de Pastas do Projeto	18
5.4.1	Protocolos: Múltiplas Execuções do MARS (earth)	19
5.5	A Pasta de Trabalho Excel Central	21
5.6	Backups	21
6	Notação da Pasta de Trabalho	22
6.1	Dimensões Descritivas	22
6.1.1	A Dimensão do Projeto (“a”)	23
6.1.2	Dimensão da Versão de Execução (“r”)	24
6.1.3	Etapas do Fluxo de Trabalho (“w”)	24
6.1.4	Planilhas da Pasta de Trabalho Excel (“s”)	25
6.1.5	Listagens de Imóveis (“p”) — linhas do MLSData	25
6.1.6	Variáveis do MLS (“v”) — colunas do MLSData	25
6.1.7	Exemplo Final de Notação	26
6.1.8	Justificativa	26
6.1.9	Matrizes?	27
6.1.10	Primeira Simplificação	27
6.1.11	Segunda Simplificação	27
7	O Sales Grid do RCA	29

8	Variáveis	33
8.1	Ontologia da Origem das Variáveis	33
8.2	Ontologia da Aplicação das Variáveis	34
8.3	Termos de Segundo e Terceiro Graus	36
8.3.1	Variáveis de Ajuste	36
8.3.2	Variáveis de Agregação	36
8.3.3	Variáveis e Notação Especiais	37
8.3.4	Decompondo os Resíduos	37
9	Demonstrações Fundamentais do RCA	39
10	Modelo MARS	41
10.1	Funções “ g ” de Variáveis e Interações do Modelo	41
10.2	Indexação das Funções g	41
10.3	Dimensões dos Gráficos	42
10.4	A Equação Final do Modelo	42
10.5	Número de Gráficos Necessários Para Representar um Modelo	43
11	Características dos Dados	46
11.1	Finalidade do Resíduo	46
11.2	Notas de Estado de Conservação e Padrão Construtivo Como Regressores Aproximados	47
11.3	Erros nos Dados	47
11.4	O Que É Exatidão?	49
11.5	O Imóvel Avaliando Como Anomalia	50
11.6	A Importância do CVR2 no Trato de Dados Inesperados	51
12	Resíduos	52
12.1	Ordenação e Pontuação	52
12.2	Curva CQA-Resíduo	55
12.3	As Características da Curva CQA	55
12.3.1	Análise dos Componentes da Curva CQA	55
12.3.2	Dinâmica de Mercado e Implicações Sociais	56
12.4	Resíduo do Avaliando	56
12.5	Conclusão de Valor do RCA	58
12.5.1	Considerações Gerais	58
12.5.2	Revisão, Auditoria, Reclamações	58
13	Conclusão	60

1 Introdução

Este artigo apresenta o fundamento lógico e matemático de uma abordagem moderna para estimar com exatidão o valor de mercado de imóveis residenciais e de outros tipos de bens imóveis. O método, desenvolvido ao longo de 20 anos de experiência na aplicação do software Multivariate Adaptive Regression Splines (MARS) a residências do norte da Califórnia — em particular em comunidades da San Francisco Bay Area —, representa um avanço significativo na precisão da avaliação. O MARS eleva a qualidade da avaliação ao capturar com habilidade as relações não lineares entre as características do imóvel e o valor, ao mesmo tempo em que identifica e prioriza de forma eficiente as variáveis de maior impacto. Essa capacidade permite que os avaliadores descubram a dinâmica do valor em novas regiões de mercado muito mais rapidamente do que os métodos tradicionais permitem. O MARS constitui a espinha dorsal do Residual Constraint Approach (RCA) (pt.: abordagem por restrição residual) aqui proposto. O método RCA faz jus ao seu nome graças ao seu arcabouço rigoroso:

- Impõe restrições estritas sobre as contribuições de valor das características do imóvel, tanto medidas quanto não medidas:
 1. Todas as contribuições de valor derivadas do MARS, inclusive o resíduo, devem somar o preço de venda líquido de cada elemento comparativo.
 2. O resíduo pode ser decomposto em subcomponentes com valores atribuídos, desde que seu total seja igual ao resíduo do elemento comparativo.
- O RCA minimiza o viés na estimativa das características subjetivas — tais como o apelo estético, o estado de conservação e o padrão construtivo — ao vincular seu valor total ao resíduo, o qual é quantificado com exatidão quando um engenheiro de avaliações competente constrói o modelo MARS.

Embora concebido principalmente para imóveis residenciais, nos quais as características subjetivas influenciam de maneira significativa o valor de mercado, o RCA tem potencial para aplicações mais amplas na avaliação de ativos. Avaliações exatas são críticas diante do valor estimado de \$45 trilhões do mercado residencial de imóveis dos EUA. A Fannie Mae observa que as avaliações típicas se desviam cerca de 5% dos preços de venda efetivos, embora essa cifra mascare o viés, já que os avaliadores frequentemente enfrentam pressão para se alinhar aos preços de venda. Além disso, como 60%–70% das avaliações dão suporte a refinanciamentos — nos quais não há preços de venda comparáveis —, a necessidade de precisão, tal como a oferecida pelo RCA, torna-se evidente. A análise do Federal Reserve de 2012 estimou que 15%–20% das avaliações anteriores à crise de 2008 se desviavam em mais de $\pm 10\%$ dos preços de venda, faixa provavelmente ainda pertinente hoje para as avaliações de refinanciamento, com variações conforme o tipo de imóvel e o mercado. Por que a exatidão importa? Em operações com baixa relação empréstimo-valor (loan-to-value, LTV) — por exemplo, um comprador com bom histórico de crédito dando \$500,000 de entrada em uma residência de \$1 milhão — os credores podem dispensar a avaliação, supondo risco mínimo de execução hipotecária. Contudo, os compradores que dispensam a avaliação correm o risco de pagar demais e de descobrir defeitos somente após a compra. Isso ressalta o valor de métodos robustos de avaliação como o RCA. Quantificar a exatidão das avaliações continua sendo um desafio. Os pesquisadores carecem de métricas diretas dos desvios em relação ao “valor de mercado” e recorrem, em seu lugar, a indicadores indiretos. Os métodos tradicionais de avaliação, enraizados no juízo subjetivo e na competência variável dos avaliadores, complicam a análise. Entre as estratégias empregadas estão:

1. Tzioumis (2016) estudou 4,782 avaliadores em 259,436 avaliações em 11 estados e constatou que um terço ficava dentro de 1% do preço de contrato, metade entre 1%–3% acima e uma

minoria em média 5% acima — o que sugere viés de alta. O estudo identificou dois tipos de avaliador: os “ao valor de mercado” (próximos ao preço de contrato) e os “acima do valor de mercado” (consistentemente mais altos).

2. Krivorotov e LaCour-Little (2020) constatou que 90% das avaliações de compra igualavam ou superavam os preços de contrato, o que evidencia o “viés de ancoragem” — a tendência a se alinhar a um valor desejado —, especialmente em contextos de refinanciamento.

Esses estudos revelam viés nas práticas de avaliação, amplificado pelas limitações das ferramentas tradicionais. Se os avaliadores fossem uniformemente competentes e imparciais, suas avaliações convergiriam; a supervalorização persistente por parte de certos grupos sinaliza problemas sistêmicos.

Os estudos sobre viés citados acima, e o padrão sistêmico que eles apontam, refletem um traço mais profundo da prática tradicional de comparação de vendas que o RCA foi concebido para enfrentar. A metodologia tradicional opera em dois canais com estatuto epistêmico bastante distinto. No canal das características mensuráveis — GLA (área bruta habitável), número de cômodos, área do terreno, idade, localização — espera-se que o avaliador derive os valores de ajuste de maneira objetiva, por meio da análise de pares comparativos ou de regressão linear local sobre dados de mercado. Avaliadores competentes realizam esse trabalho, e os ajustes resultantes são, no mínimo, aproximadamente defensáveis diante dos dados. Mesmo aqui, contudo, o quadro de homogeneização tradicional registra somente ajustes; ele sequer possui uma coluna para contribuições de valor. Tampouco os ajustes tradicionais são calculados como diferenças entre contribuições de valor — eles são estimados diretamente, a partir de pares comparativos ou de procedimentos semelhantes, sem que as contribuições subjacentes cheguem a ser calculadas. No RCA, ao contrário, cada ajuste é a diferença entre as contribuições de valor do imóvel avaliando e as do elemento comparativo para a característica considerada — a estrutura cujas consequências a Demonstração II torna precisas. No canal das características subjetivas — estado de conservação, padrão construtivo, projeto, utilidade funcional, estética — os procedimentos de fundamentação disponíveis vão do tênue ao precário. Os ajustes de características subjetivas podem ser fundamentados por análise de pares comparativos de imóveis fora do conjunto imediato de comparáveis, por apelo à prática reconhecida dos pares ou por referência à experiência documentada do avaliador no mercado. Todos os três são formalmente conformes às normas profissionais. Nenhum deles atinge o rigor estatístico que o canal mensurável alcança rotineiramente.

O resultado é uma assimetria procedimental entre o que as normas exigem e o que a metodologia consegue entregar. As normas exigem fundamentação para cada ajuste, mensurável ou subjetivo. A metodologia entrega fundamentação robusta em um canal e um espectro de procedimentos mais fracos no outro, indo de um par comparativo anedótico de tamanho amostral um até a citação daquilo que outros avaliadores costumam utilizar. O avaliador honesto e competente que faz um trabalho cuidadoso de pares comparativos para um ajuste de vista produz fundamentação sensivelmente melhor do que aquele que cita a prática dos pares sem embasamento, mas um revisor que compare as duas avaliações no formulário padrão não consegue distingui-las facilmente, porque o formulário registra o valor do ajuste, e não o rigor de sua derivação. A metodologia tolera a assimetria porque nenhum procedimento mais forte para características subjetivas esteve, de modo geral, à disposição dos avaliadores em atividade.

O RCA aprimora ambos os canais, mas de maneiras categoricamente distintas. No canal mensurável, a regressão MARS captura relações não lineares, interações entre variáveis e uma estrutura de funções de base que a análise de pares comparativos e a regressão linear ordinária não conseguem representar. Ajustado contra o conjunto completo de vendas qualificadas da região de mercado — tipicamente várias centenas — em vez de contra os três a nove elementos comparativos do quadro tradicional, as contribuições de valor derivadas do MARS são ao mesmo

tempo objetivas e substancialmente mais exatas do que a metodologia que substituem. No canal subjetivo, o RCA faz algo diferente em espécie de qualquer dos procedimentos tradicionais de fundamentação. Ele não tenta, de modo algum, estimar valores de ajuste para características subjetivas individuais. As contribuições de todas as características subjetivas — variáveis latentes — aparecem coletivamente como o resíduo: a diferença entre o preço de venda líquido efetivo de cada elemento comparativo e o preço previsto pelo MARS, calculável a partir dos dados sem intervenção do avaliador. A decomposição desse resíduo em componentes nomeados — Projeto, Utilidade Funcional, Vista, Estado de Conservação, Padrão Construtivo e assim por diante — é restringida a somar o total do resíduo, e, pela Demonstração II (Seção 9), qualquer decomposição que satisfaça essa restrição produz a mesma conclusão de valor. O juízo que permanece no fluxo de trabalho é o posicionamento do imóvel avaliando na distribuição ordenada dos resíduos dos elementos comparativos, feito contra uma classificação inspecionável e explicitamente documentada no laudo. Uma avaliação RCA fraca continua sendo possível — um VE inexperiente pode produzir um modelo MARS sobreajustado, posicionar o avaliando de forma inverossímil ou deixar passar anomalias nos dados —, mas ela é fraca de maneiras que um revisor pode identificar ao inspecionar o modelo e a classificação, e não de maneiras que uma revisão por formulário padrão não consiga trazer à tona.

Para além do refinanciamento, os avaliadores devem aos proprietários avaliações exatas, especialmente em compras de alto valor, nas quais os erros podem custar dezenas ou centenas de milhares. Ainda assim, muitos hesitam em gastar \$600–\$1,000 em avaliações, duvidando de seu valor agregado. Métodos avançados como o RCA, embora inicialmente intensivos em recursos, prometem maior precisão. Com refinamento e experiência, sua eficiência — e sua adoção — devem aumentar.

2 Antecedentes

A formação acadêmica do autor é em matemática, ciência da computação e estatística. Por profissão, é engenheiro de software, com experiência secundária em avaliação e contabilidade. Realiza avaliações, de forma intermitente, desde 2002 e aplica o MARS desde cerca de 2004 para determinar ajustes de imóveis e tendências de preço em imóveis residenciais.

Avaliar com MARS é “experiência profunda” em avaliação. O MARS gera modelos que refletem, com mais detalhe do que qualquer outro método estatístico, tanto as variáveis que contribuem para o valor de mercado quanto o modo como mudanças nesses valores impactam o preço de venda. Assim, ao avaliar com MARS, o avaliador investiga constantemente as relações causais entre valores de variáveis e preços de venda tal como geradas pelos modelos MARS.

No interesse da transparência, a metodologia RCA é mais bem aplicada a áreas residenciais com estas características:

1. Bairros complexos, com imóveis bastante diferentes entre si
 - Residências mais antigas
 - Qualidades variadas de reforma
 - Muitas arquiteturas e projetos distintos
 - Variação em atributos do imóvel como GLA, área do terreno, altimetria e vista.
2. Condições de mercado complexas
 - Condições de mercado em rápida mudança
 - Muitos tipos distintos de compradores

3. Boas fontes de dados: o protocolo RCA baseia-se em mineração de dados e necessita de dados abundantes e de qualidade. Embora possa ser usado em áreas com dados escassos, talvez não compense o tempo e o custo exigidos.
4. Deseja-se estimativas exatas do valor de mercado. O RCA deve, na maioria dos casos, conseguir fornecer exatidão de $\pm 1-2\%$. O laudo do engenheiro de avaliações deve comentar as condições de exatidão na conclusão final de valor. Se faltarem dados ou informações, ou se padrões consistentes do mercado não tiverem sido descobertos por meio da mineração de dados, a exatidão será prejudicada e o cliente deverá ser informado da situação. Especificamente, um R2 e um CVR2 elevados (R2 validado cruzadamente, calculado pelo particionamento repetido do conjunto de dados MLS em partições aleatórias de treino e teste), um modelo MARS significativo e um padrão forte nas características CQA (condition-quality-appeal: estado de conservação, padrão construtivo e apelo) do imóvel nos resíduos ordenados são os principais meios de aferir exatidão e confiabilidade na conclusão de valor.

O desenvolvimento destes protocolos RCA ao longo dos últimos 20 e poucos anos apoia-se na disponibilidade dos dados da MLSListings Inc. (Sunnyvale, CA) e de outros MLS, atualmente acessíveis por meio da MLSListings, que representam mais de 95% da experiência de análise de dados do autor. A expertise do autor decorre de extenso trabalho de avaliação em 15 condados primários do norte da Califórnia, com experiência adicional em 10 condados secundários.

A metodologia RCA é mais eficaz em mercados que atendem a dois critérios essenciais:

1. Ampla disponibilidade de dados de avaliação de alta qualidade
2. Áreas residenciais complexas e heterogêneas, caracterizadas por:
 - Estilos arquitetônicos diversos
 - Graus variados de reforma dos imóveis
 - Padrões de urbanização ao longo de várias décadas

Experiência nos Condados da Califórnia

As afirmações feitas neste artigo baseiam-se na experiência de avaliar imóveis nos seguintes condados do norte da Califórnia:

San Mateo	Santa Clara	Marin	Napa	Monterey	Santa Rosa
Alameda	San Francisco	Contra Costa	Sacramento	El Dorado	San Joaquin
Stanislaus	Placer	Solano	Mendocino	Lake	Colusa
Merced	Nevada	San Benito	Santa Cruz	Sutter	Yuba

As áreas acima abrangem essencialmente o Vale do Silício e os condados que o cercam, estendendo-se para fora em direção a regiões mais rurais, incluindo o Vale Central e o sopé da Sierra. Cabe mencionar que o autor sempre teve acesso a dados abundantes e *relativamente* exatos em comparação com aquilo de que muitos avaliadores dispõem em outras áreas menos desenvolvidas dos EUA.

Os dados de MLS utilizados são certificados RESO, em conformidade com o RCF (RESO Common Format). Apesar da certificação RCF, dados de importância geralmente secundária às vezes estão ausentes ou incorretos para alguns imóveis, seja por erros de lançamento do órgão fiscal de origem, seja por erros posteriores de digitação do corretor no MLS. Com frequência esses

erros seguem um padrão, como simplesmente não fazer um lançamento quando a característica não existe. Por exemplo, em vez de digitar 0 em “carport spaces” (vagas cobertas sem fechamento) quando não há carport, nada é digitado. Em áreas onde carports são raros, o analista pode substituir a ausência de dado por 0, a fim de obter um ajuste útil para carports. Embora o R/earth e outras implementações de MARS consigam lidar com dados faltantes, cabe ao Valuation Engineer (VE, engenheiro de avaliações) decidir como tratá-los.

Os dados da MLSListings também estão disponíveis a partir dos MLS de alguns outros estados, da Flórida ao Havaí. Isso pode ser útil para testar o quão bem os protocolos e o código funcionam em estados distintos.

3 Engenheiro de Avaliações

Definição 1: Engenheiro de Avaliações

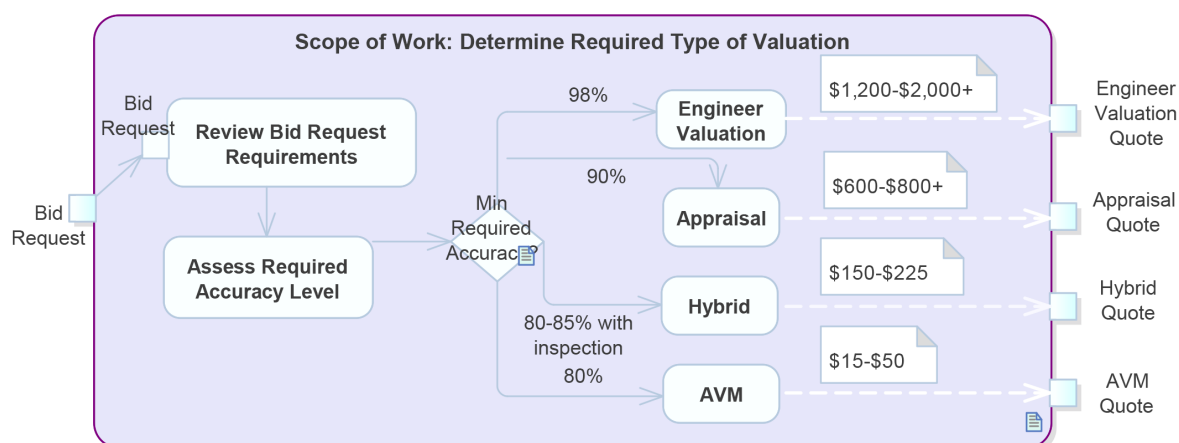
Para evitar ambiguidade, o termo “Valuation Engineer” (ou “VE”, em português engenheiro de avaliações) é usado para designar um indivíduo que reúne a expertise de um avaliador habilitado com conhecimentos, habilidades e experiência avançados no uso de mineração de dados, estatística, programação, protocolos detalhados e inteligência artificial para apurar o valor de mercado de imóveis complexos. Também é importante reconhecer que a avaliação é uma tarefa que exige alguma forma de habilitação, dado o potencial já estabelecido de fraude.

Os credores tipicamente oferecem várias opções de avaliação:

1. Automated Valuation Models (AVMs, modelos automatizados de avaliação) — a opção de menor custo
2. Avaliações híbridas — que utilizam inspetores não habilitados sob supervisão de avaliador habilitado
3. Avaliações padrão — realizadas por avaliadores habilitados

Ausentes da consideração típica para trabalhos de avaliação estão as avaliações em nível de engenharia. Simplesmente não há muitos avaliadores por aí com o conjunto de competências necessário para realizá-las. Deveria haver demanda por elas. Elas podem proporcionar exatidão significativamente maior e assegurar quase por completo uma objetividade razoavelmente boa com o emprego da metodologia RCA (MARS e restrições).

Para imóveis padronizados em loteamentos mais novos, as avaliações padrão em geral bastam. Contudo, avaliações de engenharia são recomendadas para residências mais antigas em áreas com histórico significativo de reformas ao longo de mais de 30 anos. A escolha entre esses métodos deve seguir um processo decisório estruturado, baseado nas características do imóvel e nas condições de mercado.



4 Premissas

Premissa 1: Área Geográfica e Tipo de Imóvel

O escopo deste artigo concentra-se primordialmente em residências unifamiliares e condomínios nas áreas metropolitanas e urbanas da Califórnia. Embora o método comparativo direto de dados de mercado se aplique a outros tipos de imóveis e regiões geográficas, os métodos da renda e do custo costumam desempenhar papel mais dominante na avaliação de terras agrícolas, imóveis multifamiliares e imóveis comerciais. Nesses casos, a abordagem RCA pode exigir modificação ou ter aplicabilidade limitada.

Premissa 2: Armazenamento de Dados

CSV e Excel são amplamente usados pelos sistemas MLS e assumidos como a principal forma de alimentar os programas RCA com dados. Ainda que feeds de corretoras possam ser usados para alimentar continuamente bancos de dados SQL. Backup e armazenamento intermediário frequentemente usam bancos de dados SQL, e dois que desempenham papéis importantes são os gratuitos e de código aberto SQLite e PostgreSQL. Arquivos de texto ou RTF também são usados para trechos de laudo, e arquivos PNG, JPG e TIFF são usados para gráficos. Note que o SQL Server da Microsoft também é usado por muitos, mas a edição gratuita “SQL Server Express” é muito limitada (por exemplo, só é possível usar 1 núcleo de CPU) e outras versões têm taxas de licença pesadas para uso comercial. Bancos de dados imobiliários na Califórnia podem ser muito grandes e as junções podem consumir muito tempo, de modo que um computador com 8 ou mais núcleos e 128G de RAM pode ser útil para junções grandes. Bancos de dados SQL também são importantes quando é preciso encontrar todos os imóveis em uma área delimitada arbitrariamente. Assim, se, por exemplo, você usasse análise de agrupamentos para definir bairros, tenderia a usar algum banco de dados SQL. Contudo, bancos de dados diferentes possuem funções internas distintas para o tratamento de informação geográfica e, portanto, não são compatíveis entre si. Para este conjunto de artigos sobre RCA, presume-se o PostgreSQL para tarefas SQL complexas e o SQLite para tarefas simples.

Premissa 3: “R/earth” É Usado Para o Processamento MARS.

Até 2017, o software MARS da Salford Systems era uma solução acessível, a \$270 por ano. Contudo, após a aquisição da Salford Systems pela Minitab, o MARS passou a ser vendido no pacote SPM a \$16,000 anuais — um patamar de preço proibitivo para a maioria das pequenas empresas. Felizmente, o pacote “earth” da linguagem de programação R (Milborrow 2024a), mantido por Stephen Milborrow, evoluiu a ponto de igualar ou superar as capacidades do MARS do SPM da Minitab. O pacote earth oferece vantagens adicionais: é gratuito, integra-se perfeitamente ao abrangente ecossistema estatístico do R e pode ser acessado a partir do Python usando Pandas, proporcionando excelente flexibilidade para diferentes fluxos de trabalho. As discussões relativas ao MARS pressupõem o uso específico do R/earth.

Premissa 4: Notação

Dados brutos de fontes externas, como arquivos CSV e planilhas Excel, são representados com caracteres latinos padrão. Isso inclui os dados de imóveis do Multiple Listing Service (MLS), as configurações do MARS (Multivariate Adaptive Regression Splines) usando o pacote R/earth e dados específicos do projeto. Essa notação estende-se às respectivas representações em data frame no R ou no Python.

Ao tratar de *características do imóvel* e de seus vários tipos e subconjuntos, a notação passa a usar letras gregas, para maior precisão e clareza analítica. Embora essa notação matemática possa parecer mais rigorosa, ela cumpre um propósito crucial: distinguir as características abstratas do imóvel de suas representações em dados brutos. Por convenção, os caracteres gregos denotam consistentemente características do imóvel, ao passo que os caracteres latinos indicam estruturas de dados de planilha. Por exemplo, o símbolo M representa a matriz de dados do MLS, organizada com os imóveis nas linhas e as respectivas características nas colunas.

Premissa 5: Linguagens de Programação

Nesta área, R e Python com Pandas são as linguagens dominantes, com C ou C++ sendo usados quando a velocidade é importante. O próprio R/earth é escrito em R e C.

Premissa 6: Indexação

A indexação pode variar entre linguagens de programação e matemática, o que pode gerar confusão:

- A matemática e o R iniciam a indexação em 1.
- C, C#, C++ e Python iniciam em 0.

Neste artigo, para manter tudo consistente e simples, toda a indexação começa em 0 (como em Python ou C). Veja como funciona:

- Para listas, matrizes ou data frames:
 - O índice 0 é reservado ao “imóvel avaliando” (o imóvel principal em estudo).
 - Exemplo: em uma tabela cujas linhas são imóveis, a linha 0 é o imóvel avaliando, e a coluna 0 contém identificadores únicos de todos os imóveis.
 - Uma dica prática: organize os imóveis com o avaliando primeiro (linha 0) e, em seguida, liste os demais da data de venda mais recente para a mais antiga, numerando-os sequencialmente (0, 1, 2 etc.).
- Quando não há imóvel avaliando:
 - Se uma estrutura de dados não inclui imóveis (por exemplo, apenas variáveis ou funções), o índice 0 é pulado e a indexação começa em 1.
 - Nota especial para usuários de R: como o R usa indexação baseada em 1, ao codificar em R a matemática deste artigo, some 1 a todos os índices para ajustar de base 0 para base 1.

5 Fluxo de Trabalho do RCA

Compreender o fluxo de trabalho do RCA é essencial para uma análise eficaz. O processo segue uma sequência estruturada de tarefas analíticas e pontos de decisão

5.1 Tipos de Análise MARS

O fluxo de trabalho da análise MARS depende de se é necessária apenas uma análise de mercado ou se é exigida a avaliação de um determinado imóvel. Note que a avaliação de um imóvel específico normalmente requer a análise de sua região de mercado.

- Análise de mercado: concentra-se em áreas amplas, como condados, cidades ou regiões delimitadas. Essa abordagem não envolve um imóvel avaliando específico e, portanto, não exige conclusão de valor. O fluxo de trabalho gira em torno do desenvolvimento do modelo, da validação e da geração de análises de apoio. O processo se encerra assim que o modelo ótimo é identificado e documentado.
- A análise de imóvel único abrange todos os elementos da análise de mercado, acrescidos da determinação do valor de um imóvel específico. Isso difere fundamentalmente dos métodos tradicionais de avaliação. Nas abordagens tradicionais, o valor é tipicamente obtido pelo ajuste de 3 a 9 dados de mercado e pelo cálculo de uma média ponderada. Em contraste, o RCA determina o valor primordialmente por meio de dois componentes:
 1. O valor previsto pelo modelo MARS
 2. O resíduo estimado, específico do imóvel (com o respectivo escore CQA)

O Sales Grid (quadro de homogeneização) no RCA serve primordialmente para explicar as diferenças de valor entre o avaliando e os elementos comparativos, ilustrando as contribuições das variáveis. Uma vantagem central do RCA é sua estabilidade: embora os clientes possam discutir a seleção dos comparáveis, a conclusão de valor permanece consistente independentemente de quais comparáveis sejam escolhidos para apresentação, pois todas as vendas (sejam 15 ou 600) são ajustadas com o mesmo modelo. Ainda que vendas mais antigas sejam automaticamente ajustadas às condições atuais de mercado, vendas recentes são em geral preferíveis. A eficácia do modelo em capturar os fatores de preço diminui com a distância temporal, na medida em que características de mercado não medidas se tornam mais significativas em períodos mais longos.

5.2 Fluxo de Processamento

O fluxo geral de processamento pode ser descrito por estas etapas:

1. Escopo do trabalho.
2. Configurar as pastas e os arquivos iniciais do projeto.
3. Baixar os dados do MLS. Consulte os Protocolos 1.1–1.3 quanto ao número de registros de imóveis necessários e à definição de limites para os intervalos de datas.
4. Completar as planilhas de configuração: variáveis de regressão, interações, agregações, cálculos.
5. Executar a regressão MARS (R/earth) para criar o modelo de preço.

6. Gerar as contribuições de valor e os resíduos dos imóveis (resíduo puro e resíduo/SF) a partir do modelo MARS.
7. Ordenar os imóveis por resíduo/SF.
8. Criar escores CQA para cada venda comparável.
9. Supondo que tenha sido produzida uma ordenação de qualidade dos imóveis que represente fielmente o apelo geral dos imóveis, encontrar a melhor posição do avaliando na ordenação e atribuir-lhe um escore CQA, estudando as tendências gerais do padrão na ordenação.
10. Atribuir o resíduo/SF correspondente ao imóvel avaliando como resíduo/SF · GLA (do avaliando).
11. **Somar a estimativa de preço do MARS e o resíduo atribuído para chegar à conclusão de valor do avaliando.**
12. Decompor o resíduo do avaliando em componentes descritivos e contribuições de valor.
13. Calcular as contribuições de valor e os ajustes dos comparáveis produzidos pelo modelo, além de todos os subtotais e dos preços de venda ajustados.
14. Os preços de venda ajustados serão iguais à conclusão final de valor do avaliando.
15. Selecionar os 6 a 12 imóveis mais semelhantes para o Sales Grid.
16. Decompor os montantes totais de resíduo dos comparáveis em contribuições de valor de componentes residuais, zerando o valor total do resíduo após concluída a decomposição.
17. Recalcular todos os subtotais e o preço de venda ajustado dos comparáveis do quadro. Conferir os valores.
18. Carregar o Sales Grid no laudo.

Observe que, com o RCA, a conclusão final de valor NÃO deriva diretamente dos valores dos comparáveis no Sales Grid. Com o método RCA, o Sales Grid torna-se apenas uma ferramenta para explicar por que imóveis que parecem semelhantes ao avaliando são vendidos por mais ou por menos do que a conclusão final de valor ou do que o valor apurado do imóvel avaliando. Isso precisa ser visto de outra maneira: a conclusão de valor do avaliando depende da estimativa de preço do modelo MARS e do resíduo estimado pelo engenheiro de avaliações. Mas essas duas quantidades dependem fortemente de todos os imóveis que entraram na regressão do R/earth, a qual criou a ordenação dos resíduos dos imóveis que determinou o escore CQA do imóvel avaliando e, assim, determinou seu resíduo por meio do posicionamento na ordenação feito pelo engenheiro de avaliações. Os ajustes no Sales Grid são, na verdade, determinados pelas restrições matemáticas impostas pelo RCA; embora a decomposição do resíduo fique de certo modo à mercê do VE, os totais das contribuições de valor da decomposição devem sempre somar o resíduo associado. Sabemos pela Demonstração 2 que, ainda que os valores efetivos dos componentes do resíduo possam variar conforme o juízo do VE, seus totais para um dado imóvel são completamente restringidos e, portanto, não alterarão o preço de venda ajustado do imóvel associado, independentemente de como o VE modifique a decomposição do resíduo.

5.3 Fluxo de Dados

Note que, neste artigo, apenas um fluxo de trabalho de alto nível é apresentado para o método RCA. Outros artigos aprofundarão as diversas etapas e subetapas da análise RCA.

5.3.1 Requisitos de Dados para o MARS

Será preciso especificar uma data inicial para as transações de venda e uma data final baseada na data-base da avaliação. Isso normalmente leva em conta o número estimado de variáveis de regressão e de interações que serão necessárias para cada grau de interação acima de 1.

5.3.2 Protocolos: Listagens de Vendas Fechadas

Protocolo 1.1: Determinação do Número de Listagens de Vendas Fechadas do MLS Necessárias

Este protocolo descreve como calcular o número mínimo recomendado de listagens de vendas fechadas a serem inseridas na regressão MARS. Você precisará fazer uma estimativa grosseira para começar, executar o MARS em rodadas de primeiro, segundo e possivelmente terceiro grau para gerar o modelo inicial e, em seguida, revisar o número de linhas conforme as regras a seguir para determinar se são necessárias mais linhas. Este protocolo pressupõe que as variáveis independentes sejam 100% independentes, o que raramente é o caso em imóveis; por exemplo, GLA e número de cômodos costumam ser altamente correlacionados, e é provável que a GLA também apresente correlação com a área do terreno. Isso nos leva ao cálculo dos “graus de liberdade (DOF)”. Cada variável verdadeiramente independente recebe um grau de liberdade. Mas, quando há dependências entre as chamadas variáveis independentes, o cálculo exato dos DOF exige uma análise que foge ao escopo deste artigo. Tenha apenas em mente que, em razão da provável colinearidade entre variáveis, as estimativas abaixo estão provavelmente superestimadas.

1. Para termos de primeira ordem (sem interações), a regra de ~ 15 linhas por variável funciona bem
2. Para cada interação de duas vias permitida, convém acrescentar aproximadamente 15-30 imóveis
3. Para interações de ordem superior (3 ou mais variáveis), cada uma deve ter mais de 30 imóveis.

Contudo, essa não é uma regra rígida. O número efetivamente necessário pode depender de:

- A relação sinal-ruído dos seus dados
- A complexidade das relações que você está tentando modelar
- A variabilidade da sua variável resposta
- A qualidade dos seus dados

Exemplo:

Assim, se você tiver:

- 10 variáveis principais
- 5 interações de duas vias permitidas
- 2 interações de três vias permitidas

Você poderia calcular seu número mínimo de vendas como:

$$\begin{aligned} N &= (10 \text{ variáveis} * 15 \text{ linhas/variável}) + \\ &\quad (5 \text{ interações-de-duas-vias} * 22 \text{ linhas/interação-de-duas-vias}) + \\ &\quad (2 \text{ interações-de-três-vias} * 30 \text{ linhas/interação-de-três-vias}) \\ &= 320 \text{ linhas} \end{aligned}$$

Alternativamente, você pode fazer o seguinte: depois de executar o R/earth sobre um conjunto de dados, é possível usar o número de funções de base do modelo, mais um, como estimativa dos graus de liberdade e, então, multiplicar esse número por 20-30 para obter o número de registros de entrada necessários.

Protocolo 1.2: Os Períodos de Venda Devem Abranger Mudança Mínima no Gosto dos Compradores

Ao coletar dados históricos de imóveis, não há limitação temporal rígida se as mudanças nas preferências de mercado puderem ser capturadas por variáveis apropriadas. Considere estes pontos-chave para lidar com a evolução arquitetônica no seu mercado: mudanças nos estilos arquitetônicos dominantes (como a transição do California Ranch para o mediterrâneo) podem ser modeladas usando:

- Identificadores de loteamento
- Códigos de área do MLS
- Intervalos de data de construção

Contudo, se essas mudanças de estilo não puderem ser adequadamente capturadas por variáveis específicas, você enfrenta uma restrição importante: enquanto a Data da Venda é tipicamente usada para ajustar as condições de mercado, usá-la simultaneamente para controlar as preferências arquitetônicas confundiria esses dois efeitos distintos. Nesses casos, a abordagem preferível é:

1. Limitar a coleta de dados ao período posterior à introdução do estilo arquitetônico mais recente
2. Restringir a seleção de comparáveis a loteamentos que compartilhem o mesmo estilo arquitetônico

Isso assegura que sua análise mantenha consistência nas características dos imóveis, ao mesmo tempo em que permite o ajuste adequado às condições de mercado ao longo do tempo. Alternativamente, ao menos neste caso, supondo que você consiga extrair dos dados o estilo arquitetônico, seria possível considerar a interação entre essas duas variáveis. Contudo, tal extração nem sempre é possível.

Protocolo 1.3: Se Estiver Numa Área Com Poucas Vendas, Faça o Melhor Que Puder

Ao trabalhar com modelagem MARS para dados de vendas, são necessárias aproximadamente 15 transações para uma análise significativa. Se você tiver menos vendas (por exemplo, 5 imóveis), pode duplicar cada registro duas vezes para atingir o limiar de 15 imóveis. Considerações importantes sobre essa abordagem:

1. Use apenas regressão de primeiro grau, para reduzir o sobreajuste
2. Ignore ou dispense a validação cruzada, uma vez que o conjunto de dados é demasiadamente limitado
3. Embora o MARS provavelmente supere os pares comparativos ou os métodos de regressão linear padrão nesse cenário, esteja ciente de que duplicar dados afeta a validade das medidas de incerteza e da significância estatística

A melhor solução de longo prazo é ampliar seu conjunto de dados com vendas reais adicionais. Contudo, se você tiver de prosseguir com dados limitados, a duplicação de dados oferece uma solução temporária viável, reconhecidas as suas limitações estatísticas.

5.4 Estrutura de Pastas do Projeto

Convém discutir a estrutura de pastas do projeto. É uma boa ideia manter pastas separadas para Análises de Mercado e Avaliações de Imóvel Único. Ambas as pastas têm estrutura de subpastas semelhante:

```

RCA_Projects/
├── MarketAnalysis/
│   ├── Burlingame_20240501/
│   ├── Pacifica_20240310/
│   ├── RedwoodCity_20240615/
│   └── ...
├── SingleFamily/
│   ├── Pacific_Rosewood_123_20240510/
│   │   ├── Code/
│   │   │   ├── R/
│   │   │   │   └── ...
│   │   │   └── Python/
│   │   │       └── ...
│   │   └── Data/
│   │       ├── MLS_Original.xlsx
│   │       ├── MLS/
│   │       │   └── VER202402510153248/
│   │       │       ├── MLS.xlsx
│   │       │       ├── MLS_Stage2.xlsx
│   │       │       ├── MLS_Stage3.xlsx
│   │       │       ├── MLS_Stage4.xlsx
│   │       │       ├── MLS_Stage5.xlsx
│   │       │       └── Documents/
│   │       │           ├── LinearModels.pdf
│   │       │           ├── Cum_Dist.pdf
│   │       │           ├── Model.txt
│   │       │           ├── PartContrib.pdf
│   │       │           ├── Pairs.pdf
│   │       │           ├── Residual1.png
│   │       │           ├── Residual2.png
│   │       │           ├── ResidualSF.png
│   │       │           ├── Rpart.pdf
│   │       │           ├── VariablesImp.pdf
│   │       │           ├── VariableVsSP.pdf
│   │       │           └── ...

```

O processamento da Etapa 1 criará a pasta da versão com suas subpastas. Ele copiará o `MLS_Original.xlsx`, ainda que essa planilha *devesse* nunca mudar e refletir exatamente o que foi baixado do MLS em determinada data e hora; contudo, existe a possibilidade, em um projeto mais longo, de que seja necessário um novo download do MLS. Em seguida, esta etapa fará uma cópia da pasta de trabalho `MLS.xlsx` expandida, com seus parâmetros de configuração, para a pasta da versão. Pode haver muitas pastas de versão, cada uma criada para uma execução completa do MARS.

Depois, as etapas seguintes gerarão acréscimos de colunas, ordenações e outras alterações na

pasta de trabalho e gravarão uma versão atualizada como `MLS_Stage(N)` na pasta de dados do projeto, quando a avaliação estiver concluída. Reiterando, a pasta raiz `MLS.xlsx` está em constante mudança, e assim a cópia da versão é uma mutação contínua da pasta de trabalho `MLS` original, com exclusões e inclusões de variáveis, além de outras alterações. O código R ou Python de processamento muito provavelmente será executado com entrada a partir da pasta de trabalho `Data MLS.xlsx`. Assim, sempre temos uma cópia de todos os dados por trás de cada versão da execução do RCA.

São criadas tantas versões quantas forem necessárias até que não seja mais possível obter melhorias, momento em que a melhor versão é selecionada como versão final. Pode-se alterar o nome da pasta acrescentando “_Final” ao seu fim, de modo que fique evidente ao percorrer o diretório. Algumas versões mais antigas, com modelos interessantes ou concorrentes, devem geralmente ser preservadas. Cada pasta de versão contém uma série de arquivos de texto, diagramas e gráficos que podem ser incorporados a um laudo. A pasta de trabalho final `MLS_Stage5.xlsx` contém tantas planilhas quantas forem necessárias ao laudo. Em particular, ela pode ter uma planilha para transferência a um Sales Grid da Fannie Mae. Também conterá necessariamente os cálculos ou a decomposição por trás das agregações. Esta é potencialmente a parte mais demorada da abordagem RCA — o esforço para encontrar o modelo perfeito em uma região de mercado difícil.

Tudo o que é necessário para recriar os quadros de homogeneização precisa ser copiado para a pasta da versão, inclusive o código R ou Python utilizado — caso o código sofra alterações frequentes.

5.4.1 Protocolos: Múltiplas Execuções do MARS (earth)

Protocolo 2.1: Execuções Sobreajustadas Para Estudar Cirurgicamente os Fatores de Preço da Região de Mercado

Durante as execuções iniciais do `earth`, o sobreajuste deve ser usado estrategicamente para descobrir fatores de preço subjacentes. Por exemplo, uma região de mercado representada por 800 vendas pode conter apenas 3 imóveis com oficina, e nesses casos isolados a oficina pode ter grande impacto no preço final de venda. Uma execução apenas com passo direto (forward pass) produzirá um modelo sobreajustado que mostra a contribuição de valor aproximada dessas oficinas. Contudo, como há apenas 3 vendas desse tipo em 800, os termos de oficina serão removidos no passo reverso e, certamente, sob validação cruzada. O valor que o passo apenas direto atribui às oficinas pode ser anotado e transferido para a análise de variáveis latentes, a ser usado para qualquer imóvel com oficina que apareça no Sales Grid. Esses modelos sobreajustados também fornecem estimativas detalhadas para muitas outras características com dados escassos, que seriam eliminadas pela poda do passo reverso ou por execuções validadas cruzadamente.

Salve cada uma dessas execuções exploratórias em sua própria pasta de versão: as contribuições de valor registradas são evidências que o VE vai querer citar mais tarde, ao decompor os resíduos. Compreenda, evidentemente, que o VE deve, em última instância, apoiar-se na poda e na validação cruzada para tornar o modelo final robusto diante de imóveis que a regressão nunca viu — como um imóvel avaliando que apareça posteriormente. Execuções sobreajustadas são instrumentos de estudo; jamais são o modelo final.

Protocolo 2.2: Múltiplas Execuções Validadas Cruzadamente Sob Diferentes Parâmetros de `earth()`

Regiões de mercado diferem em estrutura e tamanho, e o VE normalmente precisa experimentar uma série de execuções: para desenvolver a compreensão de quais variáveis mais impactam o preço; para julgar se o conjunto existente de variáveis medidas cumpre adequadamente a tarefa de estimar valor, ou se novos dados precisam ser acrescentados (distância à água, altimetria, distância a centros comerciais e assim por diante); e para descobrir novos tipos de interação, ao mesmo tempo em que se desautorizam outros por ineficientes ou enganosos. Ao final, o VE quer um bom R^2 nos dados de treino e o melhor CVR2 possível contra os dados de teste. Chegar lá pode exigir coletar novos dados, melhorar a exatidão dos dados existentes ou ajustar os parâmetros de `earth()` ao tamanho e ao caráter da região de mercado (por exemplo, aos graus de liberdade disponíveis — ver o Protocolo 1.1).

É precisamente por isso que a estrutura de projeto acima comporta muitas pastas de versão de execução: cada modelo candidato é preservado com seus parâmetros e seu R^2 /CVR2, de modo que a trajetória ao longo das execuções possa ser comparada. Uma versão cujo R^2 sobe enquanto o CVR2 estagna ou cai está ajustando ruído, e não sinal de mercado, e é o histórico de versões que torna isso visível.

Protocolo 2.3: Expansão em Prolog de Observações e Colunas de Características Descritivas

Embora esse recurso ainda esteja em fase um tanto experimental, o `earthUI` agora vem com um motor Prolog embutido (`vProlog`) e usa Definite Clause Grammars (DCGs, gramáticas de cláusulas definidas) para analisar as colunas de texto livre do MLS — Public Remarks (observações públicas) acima de tudo, mas também as observações do corretor e listas de características separadas por vírgula, como características dos cômodos ou benfeitorias existentes. A expansão converte expressões como “owned solar” (solar próprio), “3 car garage” (garagem para 3 carros) ou “40x50 workshop” (oficina de 40x50) em colunas de características comuns, que podem ser oferecidas à regressão, e as características com dados escassos assim recuperadas alimentam diretamente a análise de variáveis latentes (ver a subseção adiante sobre notas de estado de conservação e padrão construtivo como regressores aproximados; o recurso em si está documentado no artigo companheiro sobre o `earthUI` nesta edição).

A expansão de texto multiplica o conjunto de preditores candidatos, e isso torna necessárias múltiplas execuções `earth/RCA` para teste: as colunas expandidas devem ser testadas dentro e fora do modelo, consolidadas em contagens derivadas ou booleanos quando forem colineares, e conferidas contra o CVR2 para confirmar que carregam sinal de mercado, e não ruído. Trate cada experimento desses como sua própria pasta de versão de execução, exatamente como nos Protocolos 2.1 e 2.2. E, uma vez que a própria gramática de análise está em desenvolvimento, o VE deve revisar as colunas extraídas em confronto com o texto de origem antes de confiar nelas em um modelo.

5.5 A Pasta de Trabalho Excel Central

Como a maioria dos serviços de MLS baixa os dados de imóveis em planilhas CSV do Excel, o Excel é provavelmente a forma mais conveniente de armazenar dados e diversos parâmetros de configuração. Há uma pasta de trabalho Excel central para cada projeto, chamada `MLS.xlsx`, que possui uma planilha chamada “`MLSData`”, para armazenar as listagens de vendas fechadas de imóveis baixadas do MLS, além de várias planilhas de configuração e parâmetros:

```
Workbook: MLS.xlsx/  
├─ Sheet: Project Data  
├─ Sheet: MLSData  
├─ Sheet: Regression Variables  
├─ Sheet: Calculations  
├─ Sheet: Allowed Interactions  
├─ Sheet: Aggregations  
└─ Sheet: (MARS run parameters)
```

Os dados originais são armazenados em uma pasta de trabalho com nome do tipo `MLS_Original.xlsx`. Deve-se fazer uma cópia com nome diferente, digamos “`MLS.xlsx`”, que será carregada no programa RCA em R ou Python. Ambas as pastas de trabalho ficam armazenadas na pasta `Data`. Agora, a pasta de trabalho `Data/MLS.xlsx` provavelmente passará por muitas iterações de execuções, testando diferentes combinações de variáveis, parâmetros do MARS e alterações em outras configurações. Cada iteração cria uma nova pasta de entrada/saída nomeada com a data e a hora, no formato `VER_YYYYMMDDHHMMSS`. Cada pasta terá uma cópia da pasta de trabalho de entrada `MLS.xlsx` e conterá as pastas de trabalho Excel geradas para cada uma das quatro etapas de processamento (Etapas 2–5), bem como os documentos de texto, diagramas e gráficos de saída. Note que a Etapa 5, a etapa final, pode conter diversas outras planilhas para o laudo, dependendo do tipo de cliente e do escopo do trabalho.

5.6 Backups

Ao gravar planilhas Excel novas ou modificadas nas pastas de versão do projeto, também podem ser feitos backups em bancos de dados SQLite ou PostgreSQL na pasta do projeto. Os arquivos de banco de dados têm menor probabilidade de se corromper involuntariamente em períodos subsequentes, por uso em revisão ou consulta.

6 Notação da Pasta de Trabalho

6.1 Dimensões Descritivas

Há aproximadamente seis dimensões descritivas para uma pasta de trabalho “M”:

1. Avaliações “a”
2. Versões de Execução “r”
3. Etapas do Fluxo de Trabalho “w”
4. Planilhas “s”
5. Listagem de Imóveis “p”
6. Conjunto de Variáveis “v”

Assim, uma célula individual de uma determinada pasta de trabalho M pode ser descrita como:

$$\begin{matrix} r_* \\ w_* \end{matrix} \begin{matrix} a_* \\ W \\ s_* \end{matrix} \begin{matrix} p_* \\ v_* \end{matrix} \quad (1)$$

(Nota: “*” significa algum índice válido)

ou:

$$\begin{matrix} r_k \\ w_m \end{matrix} \begin{matrix} a_g \\ W \\ s_h \end{matrix} \begin{matrix} p_i \\ v_j \end{matrix} \quad \begin{matrix} i = 1, \dots, n_p \\ j = 1, \dots, n_v \\ k = 1, \dots, n_r \\ m = 1, \dots, n_m \\ g = 1, \dots, n_g \\ h = 1, \dots, n_h \end{matrix} \quad (2)$$

onde:

- n_p = n° de linhas (depende da planilha)
- n_v = n° de colunas (depende da planilha)
- n_r = n° de versões de execução
- n_m = n° de etapas do fluxo de trabalho
- n_g = n° de projetos de análise
- n_h = n° de planilhas da pasta de trabalho

Entenda que “W” faz referência à pasta de trabalho Excel inteira, com seis planilhas de dados centrais, cada uma com sua própria tabela, as quais, ao se executar o programa R ou Python associado, são carregadas em data frames internos separados. Tanto o R quanto o Python possuem funções internas capazes de tomar um nome e descobrir a que linha ou coluna do data frame ele pertence. Por exemplo, se no subscrito v_j tivermos $v = (\text{“GLA”}, \text{“LotSize”}, \text{“BathRms”}, \text{“BedRms”}, \dots)$ e $j=0$, então $v_j = \text{“GLA”}$. Tanto o R quanto o Python conseguirão internamente resolver “GLA” para a coluna em que está armazenada, que, por exemplo, pode ser a coluna 15 do data frame associado. Assim, neste artigo indexamos nossos vetores de nomes, os quais passam um nome ao programa em execução, que reindexa esse nome no índice real de coluna ou linha do data frame interno. É um pouco complicado. Mas, para ser preciso, isso é necessário e bastante conveniente. Trata-se de uma “separação de responsabilidades” entre o papel do analista e o papel do desenvolvedor; embora, na prática, o Valuation Engineer (VE, engenheiro de avaliações) possa desempenhar ambos os papéis. Note que um “data frame” do R é chamado de “dataframe” em Python.

Note também que a planilha de parâmetros de execução do MARS não é uma dimensão de W, já que W diz respeito à versão dos dados que entram no MARS, separadamente dos parâmetros que determinam como o MARS/earth trata esses dados.

Em prol da clareza, apresenta-se uma notação multidimensional complexa para a pasta de trabalho, que se relaciona mais diretamente aos data frames internos do R ou do Python nos quais as planilhas da pasta de dados são carregadas. É uma notação “pesada”. Mas, quando chegarmos às demonstrações e assim por diante, descartaremos boa parte dessa bagagem ao fazer suposições sobre aquilo com que estamos lidando. Ainda assim, essa notação ajuda a obter, desde o início, uma visão geral melhor do fluxo de trabalho e dos dados.

Começemos pela notação do conjunto de pastas de trabalho que iremos processar.

6.1.1 A Dimensão do Projeto (“a”)

Uma vez aprovado o Escopo do Trabalho (SOW), é preciso montar uma pasta de projeto para armazenar dados de entrada e de saída, incluindo parâmetros de configuração, gráficos, diagramas, planilhas e trechos de laudo. O VE deve identificar todos os projetos com um ID sequencial simples, como um inteiro, de modo que, se um laudo vier a sumir, isso fique evidente. Se projetos forem cancelados, deve haver um meio consistente de indicar que o projeto foi cancelado dentro da respectiva pasta de projeto, mas ela jamais deve ser excluída. Assumimos que os IDs de projeto começam em 0 e são incrementados de 1, embora haja outras possibilidades. Assumimos ainda que todos os IDs de projeto válidos são armazenados em um conjunto ordenado “a” criado com o programa de processamento.

Os dados do projeto de avaliação ficam na planilha de projeto da pasta de trabalho, mas também poderiam ser armazenados em algum banco de dados SQL central,

Eis os seguintes campos de projeto sugeridos:

1. ID local: um ID inteiro curto, gerado sequencialmente para a empresa. Faria uma excelente chave primária para armazenamento — e poderia ser gerado automaticamente pelo banco de dados a cada projeto sucessivo.
2. GUID global: se quiser, você pode criar um GUID (Globally Universal ID) usando

<https://www.uuidgenerator.net/>

3. Localização do imóvel (Análise de Mercado) ou endereço
4. Data-base da avaliação
5. Data do pedido
6. Data da vistoria
7. Data do laudo original
8. Data do primeiro laudo atualizado
9. Data do segundo laudo atualizado
10. Data do terceiro laudo atualizado
11. Nome do cliente
12. Endereço do cliente
13. Nome do credor
14. Endereço do credor
15. Nome do proprietário
16. Endereço do proprietário
17.

Exemplo de lista de IDs de projeto:

```
a = ("1",
      "2",
      "...",
      "141",
      "142")
```

6.1.2 Dimensão da Versão de Execução (“r”)

Em segundo lugar no fluxo de trabalho, executamos a regressão MARS sobre a entrada. Isso provavelmente envolverá de várias a muitas execuções, à medida que alteramos diversos parâmetros para aprimorar o modelo. Cada execução criará uma nova pasta de versão para a saída, com nome baseado na data e na hora da execução. Por exemplo,

```
r = ("ver20241210_083124",
      "ver20241209_151130",
      "ver20241208_160944")
```

```
latest_version : r1 = "ver20241210_083124"
```

Versões promissoras, com modelos e saídas competitivos, devem ser mantidas até que se decida por uma versão final, ou talvez por mais tempo se tiverem alguma utilidade. Talvez você queira modificar a pasta da versão para indicar que é a versão final, colocando “Final” ao fim do nome.

6.1.3 Etapas do Fluxo de Trabalho (“w”)

O fluxo de trabalho do RCA é separado em geralmente cinco etapas distintas, listadas a seguir. Cada etapa deve ser concluída com sucesso antes do início da etapa seguinte. O VE deve poder interromper o processamento ao fim de uma etapa para revisar os resultados antes de prosseguir. Isso é necessário porque as regiões de mercado são bastante complexas e frequentemente bem singulares. Podem surgir problemas incomuns que exijam uma nova abordagem. O VE precisa poder entrar no código R ou Python para fazer as alterações necessárias e até mesmo tentar novas abordagens. E, nesses casos, ele pode querer experimentar e tentar abordagens diferentes, exigindo múltiplas execuções da mesma etapa em que está trabalhando. Ainda assim, o VE também deve ter liberdade para executar todas as etapas com um único comando — com algumas ressalvas (por exemplo, a Agregação de 3º Grau já deve ter sido determinada em uma execução anterior da Etapa 2). Isso significa que, ao concluir cada etapa, ela deve armazenar todos os seus dados internos em arquivos ou bancos de dados, de modo que a etapa seguinte possa ser iniciada como se a etapa anterior ainda estivesse em memória, recarregando os dados da etapa anterior para a memória.

Eis as etapas:

```
w = ("stage1", -Configuração
      "stage2", -Processamento MARS
      "stage3", -Sales Grid final
      "stage4", -Trechos de laudo
      "stage5") -Revisão/Limpeza
```

```
stage2 : w2 = "stage2"
```

6.1.4 Planilhas da Pasta de Trabalho Excel (“s”)

Na Pasta de Trabalho Excel Central do RCA, temos as seguintes planilhas, cada uma mapeada em um data frame interno:

```
s = ("MLSData",
     "Regression Variables",
     "Calculations",
     "Interactions",
     "Aggregations 1st & 3rd Degree",
     "Aggregations 2nd Degree")
```

A razão de as agregações de 2º grau ficarem em planilha separada é que elas podem ser definidas mais facilmente com uma tabela simétrica que mostre todos os pares possíveis de interações, permitindo especificar uma variável de agregação à qual as contribuições de valor e os ajustes associados a cada par serão agregados (somados). Não poderíamos fazer isso com agregações de 3º grau, porque isso exigiria uma tabela simétrica tridimensional, o que não é possível no Excel. Além disso, uma tabela 20x20x20 teria 8,000 células a manejar. É mais fácil simplesmente listar quaisquer nomes de termos de terceiro grau na tabela junto com as agregações de 1º grau, por exemplo:

Variável	Agregar Para
AboveGradeFinArea	GLA
BelowGradeFinArea	BGLA
UnfinArea	UBA
LotSize	LotSize
...	...
GLA__LotSize	GLA
GLA__DateOfSale	GLA
BathRms__DateOfSale	GLA
Lat__Long	Location
Lat__Long__GLA	Location
...	...

6.1.5 Listagens de Imóveis (“p”) — linhas do MLSData

As listagens de imóveis formam as linhas da planilha MLSData. Elas não aparecem nas demais planilhas até a Etapa 4, na qual os trechos finais do laudo são criados com a seleção final de imóveis comparáveis. Eis uma lista exemplificativa de IDs de imóveis:

```
p = ("MLS823211", "MLS823456", "MLS823542", "MLS823721", "...")
```

6.1.6 Variáveis do MLS (“v”) — colunas do MLSData

Mais será dito adiante sobre o tratamento das variáveis. Basta afirmar, neste ponto, que a planilha MLSData tipicamente tem de 30 a 40 variáveis de vários tipos, mas normalmente apenas de 10 a 20 serão selecionadas como variáveis independentes para a regressão MARS. Essas variáveis também aparecem em outras planilhas, ainda que tipicamente como uma lista de todas

as variáveis possíveis que podem ser submetidas ao MARS. Um exemplo de lista de variáveis é:

$$v = (\text{"GLA"}, \text{"SalePrice"}, \text{"BedRms"}, \text{"BathRms"}, \text{"GarageSF"}, \text{"..."})$$

6.1.7 Exemplo Final de Notação

Vamos reunir a notação acima em um exemplo final. Usando as listas acima, seja

$$\begin{matrix} r_1 & & a_{141} \\ w_1 & \overline{W} & p_2 \\ & s_0 & v_0 \end{matrix} \quad (3)$$

Note que aqui estamos usando indexação em 0, de modo que se assume que o primeiro elemento das listas tem índice 0. O Python também usa indexação em 0, mas o R usa indexação em 1.

Assim, o acima significa que estamos falando do valor da célula para:

- A variável v_0 , ou coluna “GLA”
- Do imóvel p_2 , ou ID “MLS823542”
- Da planilha s_0 , ou “MLSDData”
- Para a etapa de fluxo de trabalho w_1 , ou “stage2”
- Da versão r_1 , ou
“ver20241209_151130”
- Do projeto a_{141} , ou “142”

A notação acima também pode então ser escrita como:

$$\begin{matrix} & & \text{"142"} \\ \text{"ver20241209_151130"} & \overline{W} & \text{"MLS823542"} \\ \text{"stage2"} & \text{"MLSDData"} & \text{"GLA"} \end{matrix} \quad (4)$$

6.1.8 Justificativa

“Por que não usamos simplesmente índices ou números na notação?” Temos dois requisitos para nossa notação:

1. Subconjuntos: por exemplo, o analista frequentemente seleciona subconjuntos de uma lista, como subconjuntos de variáveis. Pode haver de 35 a 40 variáveis na planilha MLSDData, mas apenas de 10 a 20 delas são usadas em uma dada execução. De uma execução para outra, o VE frequentemente muda as variáveis submetidas ao MARS/earth.
2. Iteração: por exemplo, $i = 1, 2, \dots, n_1$
Nossos algoritmos, protocolos e demonstrações normalmente precisam poder iterar por todos os itens de uma lista ordenada, do primeiro ao último — ou selecionar algum item em particular.

Assim, precisamos de uma letra para representar o subconjunto ou a lista de itens, e essa letra deve ter outro subscrito que possamos usar para iterar sobre os itens da lista, fornecendo-nos as cadeias de caracteres usadas para indexar as linhas ou colunas do data frame.

6.1.9 Matrizes?

E quanto às matrizes da álgebra linear? Matrizes só funcionam com números, inteiros ou reais. Matrizes como estrutura de dados são suportadas tanto em R quanto em Python. Contudo, muitas das variáveis com que trabalhamos nos dados de MLS são chamadas de “strings”, ou dados de caracteres. Mesmo que algumas variáveis sejam apenas números, muitas vezes não são números reais, e sim apenas nomes. Por exemplo, os códigos de área do MLS para uma cidade podem ser (“660”, “661”, “662”, “663”, ..., “672”). A regressão do earth permite que variáveis sejam especificadas como variáveis “factor” (fator), caso em que os números são tratados simplesmente como nomes, sem qualquer sequência. Nesses casos, o modelo produzido apenas fornecerá uma contribuição de valor para cada número, em vez de uma função que percorra a sequência. Outros exemplos incluem “C1”, ..., “C6” para Condition (estado de conservação) e “Q1”, ..., “Q6” para quality (padrão construtivo).

6.1.10 Primeira Simplificação

A notação acima para a pasta de trabalho Excel central do projeto é boa para amarrar as coisas, mas pesada demais para carregar ao explicar algoritmos ou demonstrações matemáticas. Quando estivermos no trabalho metódico, podemos assumir que ele se dá dentro de um dado projeto, versão, etapa e planilha. Apenas dizemos com o que estamos trabalhando, se isso já não estiver implícito. Tudo de que realmente precisamos do acima, uma vez indicado com o que estamos trabalhando, é:

$$M_{s_h}^{p_i} \quad \begin{matrix} i = 1, 2, \dots, n_p \\ j = 1, 2, \dots, n_v \\ h = 1, 2, \dots, n_s \end{matrix} \quad (5)$$

onde: n_p = n° de imóveis
 n_v = n° de variáveis de entrada
 n_s = n° de planilhas de entrada

6.1.11 Segunda Simplificação

É mais conveniente renomear as planilhas segundo os tipos específicos de data frames que elas suportam, em vez de usar índices. Para cada planilha, as linhas e as colunas têm funcionalidades diferentes. Assim, o conjunto W de data frames é decomposto nos data frames M, V, A, I e C, da seguinte forma:

$$\diamond \text{ Dados do MLS} \quad M_{v_j}^{p_i} = W_0^{p_i} \quad \begin{matrix} i = 0, 1, \dots, n_p \\ j = 0, 1, \dots, n_v \end{matrix} \quad (6)$$

onde: n_p = n° de imóveis
 n_v = n° de variáveis de entrada

Nota: o avaliando vem, por padrão, sempre em primeiro lugar nas listas de imóveis

Na planilha Excel “MLSData” e no data frame associado, a primeira linha está sempre reservada ao imóvel avaliando. Ela nunca é ordenada junto com os demais imóveis do MLS abaixo dela. Ela nunca entra na regressão MARS como dado independente; entra, isto sim, como dado-alvo, ou seja, como variável dependente. O mesmo vale para as listas de imóveis.

Em todas as listas de imóveis, o primeiro elemento, de índice 0, é sempre presumido como o ID do imóvel avaliando, salvo indicação em contrário.

Mesmo que a tarefa seja criar um modelo para uma região de mercado sem qualquer imóvel avaliando, a primeira linha é, ainda assim, carregada com alguns dados fictícios de imóvel. Isso torna muito mais conveniente escrever programas, especificar algoritmos ou fazer demonstrações. Assim, se estivermos usando c para representar a lista de imóveis, c_0 é o imóvel avaliando e c_i , $i = 1, \dots, n_p$, são as listagens comparáveis que serão entrada da regressão MARS/earth.

◇ **Especificações das Variáveis de Regressão**

$$V_{s_j}^{v_i} = W_1^{v_i}_{s_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_s \end{matrix} \quad (7)$$

onde: $n_v = n^\circ$ de variáveis disponíveis
 $n_s = n^\circ$ de especificações

Note que $n_s \leq n_v$, pois normalmente selecionamos um subconjunto das variáveis disponíveis para entrada na regressão.

◇ **Cálculos**

$$C_{c_j}^{v_i} = W_2^{v_i}_{c_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_c \end{matrix} \quad (8)$$

onde: $n_v = n^\circ$ de variáveis
 $n_c = n^\circ$ de especificações de cálculo

◇ **Interações**

$$I_{t_j}^{f_i} = W_3^{f_i}_{t_j} \quad \begin{matrix} i = 1, 2, \dots, n_f \\ j = 1, 2, \dots, n_t \end{matrix} \quad (9)$$

onde: $n_f = n^\circ$ de variáveis “de”
 $n_t = n^\circ$ de variáveis “para”

Note que o data frame I é um data frame simétrico, com as variáveis listadas em ordem alfabética nas linhas “from” (de) e nas colunas “to” (para). As células são marcadas com “1” para indicar que é permitida uma interação entre o par de variáveis associado. Caso contrário, “0” indica que nenhuma interação é permitida. Para uma interação de segundo grau, isso é relativamente simples, com apenas uma interação possível. Com uma regressão de terceiro grau, temos três interações de pares entre as três variáveis, e cada par de interações precisa ser permitido para que se gere o termo (ou variável) de interação de três variáveis.

◇ **Agregações**

$$A_{v_j}^{v_i} = W_4^{v_i}_{v_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_v \end{matrix} \quad (10)$$

onde: $n_v = n^\circ$ de variáveis

7 O Sales Grid do RCA

Tenha em mente que o protocolo RCA exige um programa em R ou Python capaz de acessar o R/earth e executar todas as etapas detalhadas. É intensivo em dados, e o trabalho não poderia, de modo algum, ser feito manualmente. Ao mesmo tempo, supondo que o programa esteja livre de defeitos, é preciso um programa de software para assegurar que o trabalho exigido seja feito com precisão e sem erro humano.

É também necessário que os VEs que realizam esse trabalho consigam modificar tais programas conforme necessário, sobretudo quando estão lidando com novas regiões de mercado, bairros, imóveis imprevistos ou outras exigências e problemas inéditos.

No processamento do RCA temos três tipos de Sales Grid:

1. O Sales Grid de Processamento do RCA, aqui designado como “MLSDData”, que é um data frame muito grande, com a respectiva planilha, e que pode conter muitos detalhes intermediários que acabam passando por um processo de agregação até se chegar ao Sales Grid do RCA mencionado adiante. Um exemplo seria grande demais para exibir aqui. Muito provavelmente os engenheiros de avaliações ajustarão essa planilha às suas preferências. Pretendo escrever um artigo separado sobre o desenvolvimento de um quadro desses em alguma data futura.
2. O Sales Grid Final do RCA, usado em laudos não destinados às GSE, que se parece com a [Table 2](#). Ele imita, por exemplo, o formulário da Fannie Mae, o suficiente para oferecer ao leitor alguma base de familiaridade. Contudo, observe que ele traz mais colunas por venda comparável do que o quadro da Fannie Mae: cada comparável tem uma coluna de “Contribuição de valor” ao lado de suas colunas de “Medida” e “Ajuste”, e o avaliando também tem uma — colunas de que o formulário da Fannie Mae carece. Um formato mais simples é também o da [Table 1](#).
3. O Sales Grid de Upload para as GSE, usado para carregar o quadro em softwares como o Alamode. Note, contudo, que nem todos os pacotes de software de formulários de avaliação oferecem a possibilidade de carregar dados de quadro de homogeneização a partir de planilha. Por necessidade, tais quadros não podem incluir coluna separada para contribuições de valor, das quais as GSE, como a Fannie Mae, não têm compreensão alguma. A estratégia é lhes dar o formulário exigido, mas respaldá-lo com um Sales Grid do RCA, bem como com outra documentação de apoio, tal como detalhes sobre eventuais agregações de variáveis.

Os Sales Grids de Processamento do RCA não se parecem com os Sales Grids da Fannie Mae, mas são, grosso modo, uma transposta do Sales Grid da Fannie Mae, com linhas e colunas trocadas: cada linha contém dados de algum imóvel, e cada coluna contém dados de alguma variável do imóvel. Pode haver facilmente centenas ou milhares de linhas de imóveis e 200 ou mais colunas. A primeira linha é reservada aos nomes das variáveis, e a segunda a um imóvel avaliando, mesmo que estejamos fazendo apenas uma análise de mercado. Quando a planilha é lida para dentro de um data frame, a primeira linha da planilha torna-se o atributo de nomes, e a segunda linha, do imóvel avaliando, torna-se a primeira linha do data frame.

As colunas do Sales Grid do RCA incluem colunas de identificação e, em seguida, variáveis primárias, externas e sintéticas, combinadas e alimentadas como variáveis independentes ao MARS. Dessas variáveis, o MARS devolve aquelas que julga ter impacto significativo sobre o Preço de Venda, também conhecido como variável dependente ou alvo. Na regressão de segundo grau, podemos obter pares dessas variáveis atuando como novas variáveis. Com a regressão de terceiro grau, podemos também obter triplas de variáveis atuando como as

chamadas variáveis independentes, embora sejam, evidentemente, altamente correlacionadas com as variáveis subjacentes.

Tabela 1: Planilha RCA de exemplo: disposição vertical

Avaliando			Comp 1			Comp 2		
	Medida	Contribuição de valor	Medida	Contribuição de valor	Ajuste	Medida	Contribuição de valor	Ajuste
Preço de venda			\$1,150,000			\$1,500,000		
Valor-base		\$500,000		\$500,000	\$0.00		\$500,000	\$0
		Características medidas	Características medidas			Características medidas		
Área habitável	1,530	\$463,500	1,350	\$382,500	\$81,000	1,700	\$540,000	-\$76,500
Área do terreno	6,000	\$200,000	6,500	\$220,000	-\$20,000	7,000	\$240,000	-\$40,000
Idade	53	\$63,600	57	\$68,400	-\$4,800	51	\$61,200	\$2,400
Banheiros	2	\$15,000	3	\$30,000	-\$15,000	3	\$30,000	-\$15,000
Dormitórios	3	\$6,000	4	\$9,000	-\$3,000	4	\$9,000	-\$3,000
Total medido		\$1,248,100.00		\$1,209,900.00	\$38,200		\$1,380,200.00	-\$132,100
		Características residuais	Características residuais			Características residuais		
Resíduo	CQA 5.8	\$30,000	CQA 2.5	-\$59,900	\$89,900	CQA 8.2	\$119,800	-\$89,800
Projeto		Decomposição do resíduo			Decomposição do resíduo			
Estado de conservação		7,000		-\$15,000	\$22,000		\$35,000	-\$28,000
Padrão construtivo		5,000		-\$8,000	\$13,000		\$20,000	-\$15,000
Utilidade funcional		8,000		-\$10,000	\$18,000		\$30,000	-\$22,000
Comodidades		5,000		-\$12,000	\$17,000		\$15,000	-\$10,000
Subtotal		5,000		-\$14,900	\$19,900		\$19,800	-\$14,800
		30,000		-\$59,900	\$89,900		119,800	-\$89,800
Total			Total			Total		
Contribuição de valor total		\$1,278,100.00		\$1,150,000.00			\$1,500,000.00	
Preço de venda ajustado					\$1,278,100.00			\$1,278,100.00

\$1,278,100.00

8 Variáveis

Do lado da entrada da execução do MARS, temos apenas variáveis simples, que são:

1. Características do imóvel carregadas da planilha do MLS.
2. Variáveis derivadas, criadas pelo VE para refinar as variáveis do MLS ou acrescentar transformações de variáveis existentes. Por exemplo, o VE pode preferir transformar a Data da Venda em “DaysOffMarket”, que equivale à data-base da avaliação menos a Data da Venda.
3. O VE pode importar dados de fontes de terceiros, como taxas de juros hipotecários. Ou, se dados de latitude, longitude ou altimetria estiverem faltando nos dados do MLS, ele pode importá-los de outras fontes. Essas variáveis podem ser chamadas de “variáveis externas”.

Do lado da saída do MARS, trabalhamos com um modelo de preço que é uma combinação de termos que:

1. Contêm apenas uma das variáveis de entrada e fornecem a equação da contribuição de valor dessa variável. São chamados de termos (ou variáveis de saída) de primeiro grau. O próprio termo é tratado como um valor de contribuição de valor e recebe um nome.
2. Contêm duas das variáveis de entrada e constituem a equação da contribuição de valor da interação entre essas duas variáveis. Esses termos podem ser chamados de termos (ou variáveis de saída) de segundo grau.
3. Contêm três das variáveis de entrada e constituem a equação da contribuição de valor da interação entre as três variáveis. Neste artigo, esses termos podem ser chamados de termos (ou variáveis de saída) de terceiro grau.

As variáveis de saída do MARS são, então, contribuições de valor, inclusive a contribuição de valor residual. Elas, por sua vez, são tratadas como variáveis nos cálculos do Sales Grid, tais como o cálculo de ajustes ou a criação de gráficos.

Podemos ter diferentes tipos de variáveis, que se enquadram em duas ontologias básicas: (1) Origem da Variável e (2) Aplicação da Variável, isto é, não residual versus residual

8.1 Ontologia da Origem das Variáveis

1. Variáveis de primeiro grau
 - 1.1 Variáveis de regressão do MLS
 - 1.2 Variáveis de regressão externas
 - 1.3 Variáveis de regressão derivadas

2. Variáveis fator (factor):

Para lidar com variáveis fator, é preciso uma função capaz de identificá-las. Uma vez que elas são especificadas na pasta de trabalho Excel de configuração do projeto, cria-se uma função que retorna verdadeiro se a variável passada a ela tiver sido classificada como do tipo fator.

3. Variáveis de interação: são criadas pelo MARS ao gerar o modelo, se o grau estiver definido como 2 ou mais. Contudo, 3 variáveis interagentes é o máximo que se poderia recomendar, em razão da dificuldade de interpretação, da probabilidade de sobreajuste e do impacto sobre os graus de liberdade, com o consequente aumento do número de registros de imóveis necessários para assegurar um modelo robusto.
 - 3.1 Variáveis de segundo grau: são comuns, e bons exemplos seriam GLA:BathRms, GLA:LotSize, DateOfSale:BathRms, Latitude _Longitude.
 - 3.2 Variáveis de terceiro grau: interações menos comuns, que sugerem cautela e possível sobreajuste.
4. Variáveis de decomposição do resíduo

8.2 Ontologia da Aplicação das Variáveis

1. Contribuições de valor geradas pelo MARS
 - 1.1 Variáveis de primeiro grau
 - 1.1.1. Variáveis de regressão do MLS
 - 1.1.2. Variáveis de regressão externas
 - 1.1.3. Variáveis de regressão derivadas
 - 1.2 Variáveis de interação
 - 1.2.1. Variáveis de segundo grau
 - 1.2.2. Variáveis de terceiro grau
 - 1.3 Variáveis de decomposição do resíduo
 - 1.4 Variáveis de agregação
2. Variáveis e valores gerados pelo engenheiro
 - 2.1 Variáveis de decomposição do resíduo
 - 2.1.1. Projeto/Estética
 - 2.1.2. Paisagismo
 - 2.1.3. Utilidade funcional
 - 2.1.4. Vista
 - 2.1.5. Estado de conservação
 - 2.1.6. Padrão construtivo
 - 2.1.7. ...

Esses grupos de variáveis receberão sua notação e serão descritos com mais detalhe na próxima seção

Protocolo 3.1: Variáveis Fator

Variáveis fator, ou simplesmente fatores, podem ocorrer em todos os termos, inclusive nas interações. Elas são importantes porque precisam de tratamento especial para a regressão MARS e para a construção de gráficos.

Ainda que variáveis com valores de caractere sejam necessariamente fatores, variáveis numéricas precisam ser sinalizadas como variáveis fator para que o MARS as reconheça como tal. Por exemplo, as áreas do MLS são muitas vezes identificadas por número. Em particular, as áreas de bairro na San Francisco Bay Area costumam ser identificadas por inteiros únicos. Se esses valores forem passados ao MARS como fatores, o MARS tentará descobrir se determinadas áreas têm um valor médio de contribuição às vendas de imóveis dentro de seus limites, sem levar em conta qualquer relação com outras áreas. Em contraste, com variáveis normais, ele suporá que há algum tipo de continuidade nas contribuições de valor conforme o valor da variável muda. Essa é uma distinção importante. Além disso, essas variáveis fator podem aparecer em termos do modelo junto com variáveis que não são fatores, o que dificulta a construção de gráficos. Especificamente, ao construir gráficos de termos de variável única com fatores, agrupamos todos esses fatores e simplesmente desenhamos um gráfico de barras mostrando suas contribuições, positivas ou negativas.

Se os fatores aparecerem em termos junto com outras variáveis, eles sempre aparecerão com um único valor possível, porque variáveis fator só podem ter um valor para um dado imóvel: se uma área tem o valor 662, por definição ela não pode ter nenhum outro valor. Se você achar isso confuso, para piorar, no modelo gerado pelo MARS não há um único nome de variável para o fator; o MARS toma o nome-raiz da variável fator que lhe foi dada como entrada e então cria novas colunas, com nomes de variável formados pela anexação dos valores possíveis. Assim, uma única coluna da planilha Excel, como `MlsArea`, terá múltiplas colunas de data frame criadas para cada valor possível, tais como `MlsArea660`, `MlsArea661` etc. Uma variável como `MlsArea661` terá valor 1 para imóveis da área 661; caso contrário, 0. Em seguida, o MARS fará a regressão sobre essas colunas como se representassem variáveis booleanas. Esses nomes modificados serão usados no modelo fornecido pelo MARS.

Você provavelmente vai querer converter esses nomes do modelo de volta ao nome original, para tornar os gráficos mais compreensíveis a clientes que gostam das coisas simples (ao menos tão simples quanto possível). Em vez de ver `MlsArea661=1`, veremos `MlsArea=661`. Podemos então construir gráficos dos termos simplesmente declarando no cabeçalho do gráfico algo como “Para MLS Area = 662”. Assim, termos de grau N com M variáveis fator são graficados como termos de grau (N-M), e a dimensão do fator desaparece no título.

Note que duas ou mais variáveis fator podem existir em um único termo, de modo que você provavelmente precisaria de um gráfico para cada par possível de valores. Se a variável x, com três valores distintos, e a variável y, com dois valores distintos, aparecerem no mesmo termo junto com a variável não fator z, então você precisaria de $2 \cdot 3 = 6$ gráficos. Um exemplo seria “`MlsArea`” + “`Architectural Style`” (estilo arquitetônico); por exemplo, casas vitorianas em San Francisco podem valer mais ou menos dependendo da área de San Francisco em que estejam. Felizmente, na maioria das áreas há em geral apenas de 1 a 3 estilos arquitetônicos importantes o bastante como contribuidores de valor para serem usados no modelo (embora possa haver 6 ou mais estilos arquitetônicos diferentes).

8.3 Termos de Segundo e Terceiro Grau

Com a regressão de segundo ou de terceiro grau, alguns dos termos do modelo de saída podem incluir duas ou, respectivamente, três variáveis. Por exemplo:

$$\text{BathRms_GLA} = -52 * \max(0, 1200 - \text{GLA}) * \max(0, \text{BathRms} - 1)$$

Note que os nomeamos ordenando alfabeticamente os nomes das variáveis e concatenando-os, talvez com um underscore. Fazemos o mesmo com os termos de terceiro grau, sobre as triplas de nomes de variáveis produzidas no modelo. Por exemplo, “BathRms__GLA__LotSz”.

Essas variáveis são acrescentadas às variáveis primárias de entrada na Etapa 3, quando o programa RCA calcula as contribuições de todas as variáveis para cada imóvel, inclusive as variáveis de interação.

8.3.1 Variáveis de Ajuste

O programa RCA também calculará os ajustes de todas as variáveis e imóveis na Etapa 3, depois de calculadas as contribuições de valor, subtraindo da contribuição do avaliando para uma característica a correspondente contribuição de valor do imóvel comparável para essa característica.

8.3.2 Variáveis de Agregação

Variáveis de agregação são definidas para substituir subconjuntos de variáveis, somando todas as suas contribuições de valor e ajustes em uma variável comum a ser usada no laudo. Por exemplo, podemos ter dividido tanto a área habitável acabada quanto a não acabada em variáveis de detalhe mais fino:

1. Área acabada legal acima do nível do solo
2. Área acabada legal abaixo do nível do solo
3. Área acabada irregular acima do nível do solo
4. Área acabada irregular abaixo do nível do solo
5. Área não acabada acima do nível do solo
6. Área não acabada abaixo do nível do solo

Sim, há áreas em que isso faz sentido.

A regressão MARS fará a regressão sobre todas as variáveis e gerará um modelo que fornecerá contribuições e ajustes para elas. Mas, para um laudo em formulário, essas diferentes contribuições provavelmente terão de ser combinadas ou agregadas em menos variáveis. Os detalhes da agregação devem ser fornecidos em anexo. Um exemplo mais frequente seria agregar os ajustes de interação a um dos ajustes das variáveis primárias. Por exemplo, poderíamos agregar o ajuste de GLA_Lotsize a GLA ou a LotSize, ou mesmo dividi-lo 50:50 entre ambos, conforme determinado pela planilha Aggregation.

8.3.3 Variáveis e Notação Especiais

π_i	Preço de venda líquido efetivo do imóvel $i = 1, 2, \dots, n_p$; para o avaliando, π_0 denota a conclusão de valor (ver a Demonstração I)
α_i	Preço de venda líquido ajustado do imóvel $i = 1, 2, \dots, n_p$
ρ_i	Preço de venda líquido estimado pelo MARS para o imóvel $i = 0, 1, \dots, n_p$
β	Valor-base do modelo
$\varphi_{i,j}$	Contribuição de valor da variável de regressão $j = 1, 2, \dots, n_r$ para o imóvel $i = 0, 1, \dots, n_p$
$\delta_{i,j}$	Ajuste da variável de regressão $j = 1, 2, \dots, n_r$ para o imóvel $i = 1, 2, \dots, n_p$
ϵ_i	Resíduo total do imóvel $i = 0, 1, \dots, n_p$
$\epsilon_{i,c}$	Contribuição de valor do componente residual c , $c = 1, 2, \dots, n_c$, para o imóvel $i = 0, 1, \dots, n_p$
$\xi_{i,c}$	Ajuste do componente residual c , $c = 1, 2, \dots, n_c$, para o imóvel $i = 0, 1, \dots, n_p$

onde

n_p = nº de imóveis comparáveis

n_r = nº de variáveis de regressão

n_c = nº de variáveis de decomposição do resíduo

8.3.4 Decompondo os Resíduos

Depois que a regressão MARS cria um modelo, ela gera os preços de venda estimados a partir das variáveis de entrada selecionadas pelo MARS. Subtrai esse valor do correspondente Preço de Venda Líquido para obter um valor residual de cada imóvel. Ora, o protocolo RCA estipula que, uma vez que o VE tenha decidido quais serão os comparáveis finais, normalmente de 6 a 12 listagens de imóveis, ele precisa percorrer cada listagem e decompor o resíduo calculado nas características (variáveis latentes) que julga mais provavelmente impactarem o valor, ignorando a possibilidade de erros aleatórios nos dados. Ele poderia, por exemplo, ter estas variáveis de

decomposição para um dado imóvel:

1. Projeto
2. Utilidade funcional
3. Venda acima/abaixo do mercado
4. Estado de conservação
5. Padrão construtivo

Imóveis diferentes provavelmente terão algumas variáveis de decomposição do resíduo distintas, que precisam ser reunidas em um único conjunto ordenado.

Estimando contribuições de características raras. Algumas características portadoras de valor são raras demais para o canal da regressão: uma oficina presente em 3 de 780 vendas não consegue receber um coeficiente defensável, e o passo direto normalmente ignorará por completo uma variável dessas. Trata-se de características residuais, e suas contribuições podem ser estimadas *no espaço residual*: ajusta-se o modelo sem a característica e, em seguida, regride-se os resíduos do modelo sobre a(s) coluna(s) excluída(s) por mínimos quadrados ordinários. Para um indicador binário, o coeficiente é uma contribuição de montante fixo (equivalentemente, o deslocamento médio do resíduo das vendas sinalizadas); para uma coluna numérica, é uma contribuição por unidade. Quando várias comodidades correlacionadas são estimadas *conjuntamente* (por exemplo, a área construída de edificações acessórias junto com um indicador de oficina), a regressão reparte o crédito entre elas, em vez de contar duas vezes os mesmos valores em duas linhas da decomposição. Estimativas apoiadas em menos de cerca de dez observações são apenas indicativas e devem ser corroboradas pelo custo depreciado ou por uma amostra ampliada (por exemplo, indicadores recuperados das observações públicas); a própria restrição residual limita qualquer valor que venha a ser finalmente atribuído. O earthUI (v0.11+) implementa esse método diretamente na aba *Residual Contribution* (contribuição residual), incluindo a estimativa conjunta, a evidência residual por venda e uma nota de documentação para o arquivo de trabalho.

Notação: usamos a notação da tabela acima para chegar a:

$$\epsilon_i = \epsilon_{i,1} + \epsilon_{i,2} + \dots + \epsilon_{i,n_c} \quad (11)$$

onde n_c é o número de componentes da decomposição do resíduo.

O VE precisa usar seu juízo para decompor o resíduo nos componentes dados. Poder-se-ia pensar que isso permitiria a entrada de viés na conclusão final de valor. Contudo, não é o caso, pois temos demonstração matemática de que, se os ajustes da decomposição somarem o resíduo, eles não impactarão a conclusão de valor. O propósito da decomposição não é estabelecer valor, mas explicar por que o resíduo é o que é em comparação com os resíduos das vendas de outros imóveis. O único ponto em que o viés do engenheiro pode impactar a conclusão final de valor é a estimativa do resíduo do avaliando, já que, evidentemente, não há preço de venda efetivo do avaliando que permita calcular um resíduo para ele.

Cada imóvel tem seu conjunto de contribuições de valor residuais, mas estas precisam ser fundidas em um único conjunto C sobre todos os imóveis:

$$\begin{aligned} \text{Fazendo: } A_i &= \{\{\epsilon_{i,c}\}_{c=1}^{n_c}\} \\ C &= \{x \mid x \in \bigcup_{i=0}^{n_p} A_i\} \end{aligned}$$

onde $[\epsilon_{i,c}]$ é o nome de variável para o valor $\epsilon_{i,c}$, A_i é o conjunto de tais variáveis e n_c é o número de valores em cada ϵ_i .

Assim, com os dados do avaliando e dos imóveis comparáveis em colunas separadas, então, abaixo de todas as contribuições de valor das variáveis do modelo de regressão de cada imóvel, deveríamos ver o conjunto C das n_c correspondentes contribuições de valor residuais.

9 Demonstrações Fundamentais do RCA

Seja:

$$\begin{aligned} n_p &= \text{n}^\circ \text{ de imóveis} \\ n_r &= \text{n}^\circ \text{ de variáveis de regressão} \\ n_c &= \text{n}^\circ \text{ de variáveis de componentes residuais} \end{aligned}$$

Demonstração I: Todos e cada um dos preços de venda ajustados dos comparáveis são iguais ao preço de venda estimado do avaliando.

Para qualquer comparável i , o preço de venda ajustado α_i é a soma do preço de venda líquido π_i mais os ajustes: as diferenças entre as contribuições de valor do avaliando e do comparável para o valor-base, para as variáveis de regressão $j = 1, 2, \dots, n_r$ e para o resíduo. Para o imóvel avaliando, que não tem venda efetiva, o preço de venda líquido π_0 é definido como a conclusão de valor $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$. Então:

$$\begin{aligned} \alpha_i &= \pi_i + (\beta - \beta) + (\varphi_{0,1} - \varphi_{i,1}) \\ &\quad + (\varphi_{0,2} - \varphi_{i,2}) + \dots + \\ &\quad (\varphi_{0,n_r} - \varphi_{i,n_r}) \\ &\quad + (\epsilon_0 - \epsilon_i) \\ &= \pi_i + (\beta + \varphi_{0,1} + \varphi_{0,2} + \dots \\ &\quad + \varphi_{0,n_r} + \epsilon_0) \\ &\quad - (\beta + \varphi_{i,1} + \varphi_{i,2} + \dots + \\ &\quad \varphi_{i,n_r} + \epsilon_i) \\ &= \pi_i + \pi_0 - \pi_i \\ &= \pi_0 \end{aligned} \tag{12}$$

Demonstração II: Alterar a decomposição do resíduo, sob a restrição de que todas as contribuições de valor dos componentes residuais somem o resíduo do comparável, não tem impacto algum sobre a conclusão final de valor do protocolo RCA.

Isto é quase trivial. Ainda assim, sem pensar na matemática, é fácil deixar passar, porque os avaliadores tradicionais focam os ajustes, e não as contribuições de valor. São as contribuições de valor que se cancelam aqui — um descuido de décadas. Para algum comparável k , $k > 0$, o ajuste residual total é igual à soma dos n_c componentes residuais da decomposição, que, por sua vez, é igual à soma das diferenças de contribuições de valor, o que é igual ao resíduo total do avaliando menos o resíduo total do comparável. E isso independe do número de componentes residuais criados ou do que eles representem. Desde que o valor dos componentes residuais seja igual ao resíduo total tanto para o avaliando quanto para o comparável escolhido, ξ_k não mudará.

$$\begin{aligned}
\xi_k &= \xi_{k,1} + \xi_{k,2} + \dots + \xi_{k,n_c} \\
&= (\epsilon_{0,1} - \epsilon_{k,1}) + (\epsilon_{0,2} - \epsilon_{k,2}) + \dots + \\
&\quad (\epsilon_{0,n_c} - \epsilon_{k,n_c}) \\
&= (\epsilon_{0,1} + \epsilon_{0,2} + \dots + \epsilon_{0,n_c}) - \\
&\quad (\epsilon_{k,1} + \epsilon_{k,2} + \dots + \epsilon_{k,n_c}) \\
&= \epsilon_0 - \epsilon_k
\end{aligned} \tag{13}$$

10 Modelo MARS

10.1 Funções “ g ” de Variáveis e Interações do Modelo

Um princípio central do método RCA é que o “raciocínio explicativo” é da máxima importância. Infelizmente, fornecer gráficos significativos para as várias equações de contribuição de valor, especialmente quando se usam interações e variáveis fator, pode ser desafiador. Considera-se, portanto, imperativo fornecer um protocolo claro e estrito para a criação desses gráficos. É o que se faz aqui, mas entenda que o protocolo se dirige ao desenvolvedor que escreve o código R ou Python para gerar os gráficos, e talvez interesse apenas superficialmente a quem não está envolvido com a codificação.

O modelo de preço gerado pelo MARS é a soma de uma constante-base mais termos adicionais, cada um envolvendo de 1 a 3 variáveis. Para analisá-lo, primeiro decompomos o modelo em termos individuais. Em seguida, múltiplos termos que compartilham o mesmo conjunto de variáveis, representando uma única interação, são combinados em expressões de interação únicas. Suponha que o grau máximo da regressão seja 3. Considere um modelo de regressão de terceiro grau composto por uma constante-base, n_1 expressões de primeiro grau, n_2 expressões de interação de segundo grau e n_3 expressões de interação de terceiro grau. Isso nos dá quatro grupos distintos de expressões, que são interações de variável única ou de múltiplas variáveis, cada qual graficada de modo diverso conforme dois fatores: (1) se incluem uma variável fator (uma variável categórica) e (2) o número de variáveis distintas que contêm. Especificamente:

- Expressões de primeiro grau (1 variável) são graficadas em 2D.
- Expressões de segundo grau (2 variáveis) são graficadas em 3D.
- Expressões de terceiro grau (3 variáveis) são graficadas como um *conjunto* de gráficos 3D.
- Se uma expressão contém uma variável fator, cada fator distinto reduz em um a dimensionalidade do gráfico (por exemplo, um gráfico 3D torna-se 2D se uma variável for fator).

A indexação de cada função g fornece a informação necessária para determinar como ela deve ser graficada, refletindo o número de variáveis distintas e a presença de variáveis fator.

10.2 Indexação das Funções g

Cada função g é indexada como $f_{j,k} g_k^j$, onde:

- $f_{j,k}$ é o primeiro índice (um sobrescrito antes de g):
 - j varia de 0 a 3 e identifica o grupo:
 - * $j = 0$: a constante-base (uma exceção à indexação usual).
 - * $j = 1$: expressões de primeiro grau.
 - * $j = 2$: expressões de segundo grau.
 - * $j = 3$: expressões de terceiro grau.
 - k é o índice de uma função específica dentro de seu grupo (por exemplo, $k = 1$ para a primeira função do grupo j).
 - O valor de $f_{j,k}$ indica o número de variáveis fator distintas na função:
 - * $f_{j,k} = 0$: sem variáveis fator.
 - * $f_{j,k} > 0$: o número de variáveis fator (por exemplo, 1, 2 ou 3).

- Índices posteriores em g :
 - O sobrescrito j após g (por exemplo, g^j) repete o número do grupo, por clareza.
 - O subscrito k após g (por exemplo, g_k) corresponde ao k de $f_{j,k}$, identificando a função dentro de seu grupo.

Embora os índices de f e g coincidam numericamente, seus papéis diferem: $f_{j,k}$ conta variáveis fator, o que é crítico para as decisões de construção de gráficos, ao passo que j e k em g definem seu grupo e sua posição.

10.3 Dimensões dos Gráficos

A dimensão d do gráfico de uma função $f_{j,k} g_k^j$ é calculada como:

$$d = j - f_{j,k} \quad (14)$$

onde:

- j é o número de variáveis distintas (ou o número do grupo, desconsiderado o caso da constante).
- $f_{j,k}$ é o número de variáveis fator.
- d é o número de variáveis contínuas (não fator) a serem graficadas, com máximo de 3 (já que o grau máximo é 3). Uma função com d variáveis contínuas é desenhada em $d + 1$ dimensões: $d = 1$ dá um gráfico 2D, $d = 2$ um gráfico 3D (ou um mapa de calor 2D) e $d = 3$ um conjunto de gráficos 3D.

Por exemplo:

- Se $j = 3$ (3 variáveis) e $f_{j,k} = 0$ (sem variáveis fator), então $d = 3 - 0 = 3$, exigindo um conjunto de gráficos 3D, cada um fixando uma das variáveis.
- Se $j = 3$ e $f_{j,k} = 1$ (uma variável fator), então $d = 3 - 1 = 2$, exigindo um gráfico 3D (ou mapa de calor 2D) para cada valor da variável fator que aparece no modelo.
- Se $j = 0$ (a constante), normalmente não é preciso gráfico, pois se trata de um único valor.

Quando $d = 3$ (por exemplo, uma expressão de terceiro grau sem variáveis fator), visualizar a função completa pode exigir múltiplos gráficos 3D, cada um fixando uma variável em valores diferentes para mostrar um subconjunto do comportamento da função.

10.4 A Equação Final do Modelo

O modelo MARS é decomposto em quatro grupos de expressões (constante, primeiro, segundo e terceiro graus). A indexação $f_{j,k} g_k^j$ nos informa:

- j : o grupo e o número de variáveis.
- $f_{j,k}$: o número de variáveis fator.
- $d = j - f_{j,k}$: a dimensão do gráfico.

Essa estrutura orienta como cada função g é visualizada, assegurando que o método gráfico corresponda à complexidade da expressão e aos tipos de variáveis.

A estimativa do preço de venda $\hat{P}$ fornecida pelo modelo MARS pode ser expressa como:

$$\begin{aligned} \hat{P} = & f_{0,1} g_1^0 + \sum_{q=1}^{n_1} f_{1,q} g_q^1(x_*) \\ & + \sum_{r=1}^{n_2} f_{2,r} g_r^2(x_*, y_*) \\ & + \sum_{s=1}^{n_3} f_{3,s} g_s^3(x_*, y_*, z_*) \end{aligned} \quad (15)$$

$n_1 = \text{n}^\circ \text{ de termos de primeiro grau}$
 $n_2 = \text{n}^\circ \text{ de termos de segundo grau}$
 $n_3 = \text{n}^\circ \text{ de termos de terceiro grau}$

10.5 Número de Gráficos Necessários Para Representar um Modelo

Quantos gráficos são necessários para visualizar todas as funções de um modelo? A resposta depende das variáveis, de suas interações e de serem elas variáveis fator ou não.

O VE pode limitar variáveis e interações no arquivo de configuração Excel destinado ao R/earth, influenciando quais interações são permitidas. Uma abordagem prática é deixar o R/earth avaliar todas as interações possíveis e selecionar aquelas com contribuições significativas para a estimativa do preço de venda, verificando ao mesmo tempo problemas como sobreajuste ou colinearidade. Algumas interações podem ser desautorizadas, afetando interações de grau superior (por exemplo, se um par é excluído em uma interação de terceiro grau, a combinação inteira fica excluída).

Para variáveis fator (por exemplo, MLS Area com 12 valores interagindo com GLA), a construção de gráficos torna-se mais complexa. Isso poderia exigir 12 gráficos — um por área. Se outro fator, como Architectural Style (6 valores), interagir com MLS Area e GLA, poderiam ser necessários até $12 \cdot 6 = 60$ gráficos. Contudo, o R/earth raramente considera todas as combinações significativas; tipicamente, apenas algumas poucas interações-chave (por exemplo, áreas e estilos específicos) merecem gráfico, conforme identificado pelo modelo.

Eis um guia simplificado do número de gráficos necessários:

1. **Funções de variável única (não fator):** 1 gráfico.
2. **Funções de variável única (fator):** 1 gráfico de barras combinando todos os valores do fator.
3. **Função-base:** nenhum gráfico necessário.
4. **Interações de 2º grau (sem fatores):** 1 gráfico (por exemplo, mapa de calor 2D com cores para o preço de venda).
5. **Interações de 2º grau (1 fator):** 1 gráfico por valor do fator presente no modelo (por exemplo, 2 valores = 2 gráficos).
6. **Interações de 2º grau (2 fatores):** 1 gráfico com uma matriz das combinações dos fatores (por exemplo, 3 áreas $\times$ 2 estilos = 6 células).
7. **Interações de 3º grau (sem fatores):** múltiplos gráficos (por exemplo, 4 gráficos para número de banheiros = 1, 2, 3, 4).
8. **Interações de 3º grau (1 fator):** 1 gráfico por valor do fator presente no modelo.

9. **Interações de 3º grau (2 fatores):** 1 gráfico por combinação de valores dos fatores presente no modelo.
10. **Interações de 3º grau (3 fatores):** 1 gráfico por valor do fator com o menor número de valores distintos, usando uma matriz de cores para os outros dois.

Cenários mais complexos podem exigir técnicas gráficas avançadas.

Há outras possibilidades, com gráficos provavelmente mais sofisticados.

Um exemplo de modelo completo gerado pelo MARS está na [Figure 1](#).

11 Características dos Dados

A seção seguinte trata do MARS. O MARS depende inteiramente dos dados que recebe, e a maior parte desses dados virá de algum MLS (Multiple Listing Service) ou de órgãos fiscais. Contudo, eles podem passar por outros provedores de dados que os mantêm e aprimoram para depois vendê-los com algum lucro. Em algumas áreas, são razoavelmente confiáveis; em outras, nem tanto. Em situações críticas, o vistoriador do imóvel precisa fazer suas próprias observações e medições, cuja exatidão pode ser questionada.

O engenheiro de avaliações precisa compreender o que os dados representam e o quão bem cumprem o papel de representar aquilo que deveriam representar. Isso suscita toda uma lista de questões:

- Como se classifica uma variável como sendo uma medição?
- Se uma variável não é uma medição, o que mais ela pode ser?
- Como a regressão pode ser útil se os dados que recebe não são exatos?

A maior parte dos dados que obtemos dos serviços de listagem do MLS e dos dados dos órgãos fiscais para regressão são dados numéricos que implicam medições ou contagens. Alguns são lógicos, verdadeiro ou falso.

11.1 Finalidade do Resíduo

O resíduo serve como medida indireta do valor agregado por variáveis não incluídas no modelo de regressão. Na linguagem da literatura hedônica, essas características não medidas são *variáveis latentes*: variáveis que influenciam o preço de venda mas não são medidas diretamente, na maioria dos casos porque não há maneira prática de medi-las objetivamente. Seus valores precisam, em vez disso, ser inferidos e, na prática, o valor de uma variável latente é muitas vezes estimado por juízo subjetivo — no RCA, um juízo ancorado na ordenação dos resíduos. Ao ordenar os imóveis com base em seu resíduo ou resíduo por pé quadrado, obtemos uma ordenação objetiva, dos imóveis de menor apelo aos de maior apelo. Esse método de ordenação é eficaz, desde que um analista habilidoso, com profundo conhecimento do mercado local, construa o modelo MARS (Multivariate Adaptive Regression Splines). Seguem alguns requisitos adicionais:

- Qualidade dos dados e do modelo: um modelo de regressão bem construído e dados de alta qualidade são pré-requisitos para resíduos confiáveis. Na San Francisco Bay Area, minha meta é desenvolver um modelo com R^2 de aproximadamente 75-80% e R^2 de validação cruzada (CV R^2) de 55-70%.
- Evitar o sobreajuste, limitar o R^2 : o engenheiro de avaliações deve buscar um R^2 máximo em torno de 80% na maioria dos mercados maduros de áreas de preço mais alto, como a SF Bay Area, porque boa parte do valor competitivo de um imóvel residencial reside em seu apelo subjetivo, que não é quantificável e, portanto, não pode ser incluído diretamente em modelos de regressão. Um R^2 significativamente acima de 80% pode indicar um modelo sobreajustado, exigindo exame adicional. Ainda assim, uma palavra de cautela: existem regiões de mercado com valores de R^2 acima de 90% sem sobreajuste, e existem áreas difíceis de avaliar, nas quais o melhor que se consegue talvez seja 50-60% ou menos.
- R^2 elevado em áreas específicas: contudo, há regiões, como loteamentos ou outras áreas residenciais uniformes, nas quais valores de R^2 mais altos são surpreendentemente comuns, em razão da homogeneidade dos imóveis.

11.2 Notas de Estado de Conservação e Padrão Construtivo Como Regressores Aproximados

As escalas de estado de conservação e padrão construtivo da Fannie Mae (C1–C6 e Q1–Q6) merecem comentário especial. Essas escalas são grosseiras demais para serem tratadas como verdadeiras características medidas e, por serem atribuídas pelo avaliador, são na prática notas subjetivas — variáveis latentes por direito próprio. É útil distinguir explicitamente dois níveis: a nota grosseira do avaliador usada como regressor, *condition-coarse-grain* (condition-1, estado de conservação de granulação grossa) e *quality-coarse-grain* (quality-1, padrão construtivo de granulação grossa), versus as variáveis latentes de granulação fina recuperadas do resíduo, *condition-fine-grain* (condition-2) e *quality-fine-grain* (quality-2). O VE pode aceitar as notas de granulação grossa do avaliador como aproximações e regredir sobre elas, obtendo contribuições de valor aproximadas que são úteis como atribuições iniciais de valor para as variáveis latentes de granulação fina. Onde isso é feito, o protocolo consiste em subtrair da estimativa do modelo as contribuições de valor dessas duas variáveis de granulação grossa estimadas pela regressão — retirando-as dos atributos medidos, sob a seção de regressão do quadro de homogeneização — e devolvê-las ao resíduo. Dali, elas são subsequentemente removidas do resíduo e atribuídas às correspondentes variáveis latentes de granulação fina, *res_condition1* e *res_quality1*. Informações adicionais extraídas das observações públicas do MLS (*public_remarks*) e de listas de texto relacionadas são então usadas para fazer microrrefinamentos nesses valores de variáveis latentes. O resultado final é que as contribuições de valor de granulação fina substituem as de granulação grossa: as notas grosseiras servem apenas de andaime para o modelo aproximado, e os valores reportados ao final para estado de conservação e padrão construtivo são os de granulação fina, derivados do resíduo.

11.3 Erros nos Dados

Com base em comentários de redes sociais, os dados do Multiple Listing Service (MLS) da Califórnia são em geral de qualidade superior aos de muitas outras regiões dos Estados Unidos. Tipicamente, deve-se esperar que os dados de MLS sejam de boa qualidade na maioria das áreas metropolitanas maduras, porém piores em áreas rurais.

Embora haja erros ocasionais nos dados por diversos motivos, esses erros devem em geral cancelar-se mutuamente e ter impacto mínimo sobre as médias usadas nos modelos de preço, caso exista dado suficiente.

Contudo, vieses sistemáticos podem ocorrer em áreas específicas. Por exemplo, considere um bairro em que muitas casas foram inicialmente construídas com segundo pavimento completo, mas alguns compradores tiveram a opção de suprimir cômodos do segundo pavimento para criar pés-direitos duplos em um ou mais cômodos do térreo, reduzindo significativamente a área bruta habitável (GLA). As alterações não foram posteriormente atualizadas nos registros do órgão fiscal, por uma razão ou outra. Em consequência, o MLS pode mostrar 2,400 pés quadrados quando o valor medido é de apenas 2,000 pés quadrados. Nesses cenários, os dados do MLS devem ser corrigidos antes de se executar a regressão MARS (Multivariate Adaptive Regression Splines). É importante notar que obter essa informação para os comparáveis pode exigir algum esforço investigativo. Discrepâncias devem ser registradas para referência futura.

Diante do exposto, é razoável supor que, em áreas de MLS de alta qualidade, os erros que transbordam para o resíduo sejam mínimos em comparação com o impacto de características subjetivas e não medidas, como projeto, estado de conservação e padrão construtivo. Essa é uma consideração importante ao estimar a exatidão do método RCA. Por exemplo, se o valor de R^2 de uma dada região de mercado é 80%, então aproximadamente 20% do preço é atribuído ao resíduo CQA. Um erro de $\pm 10\%$ no posicionamento do avaliando dentro da ordenação de

resíduos propaga-se, portanto, para um erro de $\pm 10\% \cdot 20\% = \pm 2\%$ na conclusão final de valor. É assim que se melhora a exatidão! Em muitos casos, estimativas de valor exatas bem abaixo de $\pm 1\%$ são possíveis.

Portanto, é razoável tratar os resíduos como indicadores indiretos do valor coletivo das variáveis não modeladas na regressão MARS, desde que o engenheiro de avaliações tenha experiência local de mercado suficiente para identificar onde e por que existem vieses sistêmicos de medição. Essa abordagem assegura que a análise reflita com exatidão os valores efetivos dos imóveis, reconhecendo as limitações e as nuances dos dados imobiliários.

Exemplo Prático: Medindo Determinantes Não Medidos em Sea Ranch

Quando o R^2 alcançável em uma região de mercado é menor do que a aparente homogeneidade do mercado levaria a esperar, a resposta da estrutura metodológica é identificar e medir a variável faltante, em vez de aceitar a limitação. A comunidade de Sea Ranch, no condado de Sonoma, ilustra isso diretamente.

Uma instância concreta da resposta “medir, não aceitar”. A comunidade de Sea Ranch, no condado de Sonoma, é um empreendimento litorâneo em que a distância ao oceano é um determinante primário de valor: uma casa a algumas centenas de pés da falésia alcança preço substancialmente diferente do de outra a um quarto de milha para o interior, mantidas constantes as demais características. Os dados do MLS da comunidade não registram a distância ao oceano como variável. Um modelo MARS inicial, ajustado sobre as variáveis disponíveis no MLS, produziu poder explicativo inferior ao que a aparente homogeneidade do mercado deveria sustentar, com uma estrutura de resíduos que não ordenava suficientemente os imóveis por nenhuma característica visível nos dados — sinal de que um determinante importante faltava ao modelo, e não de que o mercado fosse genuinamente ruidoso.

O remédio foi a medição direta. O autor obteve um mapa da comunidade mostrando a disposição dos lotes em relação à linha da costa e mediu a distância ao oceano de várias centenas de casas. As distâncias medidas foram acrescentadas como variável ao conjunto de dados do MLS e a regressão MARS foi refeita. A nova variável foi selecionada pelo modelo como significativa, o poder explicativo aumentou substancialmente e a ordenação dos resíduos passou a ordenar os imóveis por padrões consistentes com as características subjetivas remanescentes — vista, estado de conservação, padrão construtivo — em vez de pelo sinal latente da distância ao oceano, que antes a contaminava. A altimetria tem sido tratada de modo semelhante em bairros de encosta em que a variável do MLS está ausente ou não é confiável.

O princípio metodológico é geral. Quando o R^2 alcançável fica abaixo do que a aparente homogeneidade do mercado deveria sustentar, a primeira hipótese é a de que falta ao modelo uma variável importante, e a resposta apropriada é identificar qual é e coletar as medições diretamente. As fontes incluem mapas comunitários, dados de SIG, registros de órgãos fiscais, registros da prefeitura, imagens de satélite e vistoria de campo. O custo desse trabalho de medição é real, mas limitado; uma vez coletada a variável para uma dada região de mercado, ela pode ser aproveitada em trabalhos futuros naquela área. O RCA absorve novas variáveis de forma limpa, porque a estrutura metodológica é indiferente a quais variáveis entram na regressão MARS — a estrutura dos resíduos diz ao VE se a nova variável captura sinal real, e as contribuições de valor se ajustam em conformidade.

A resposta alternativa — aceitar um R^2 baixo e reportar faixas de erro mais amplas — é defensável quando os determinantes faltantes são genuinamente imensuráveis (preferências específicas do comprador, sentimento transitório de mercado, circunstâncias idiossincráticas da transação), mas não quando eles são apenas não coletados. O primeiro movimento do VE diante de um modelo fraco deve ser perguntar se o sinal faltante é deste último tipo.

11.4 O Que É Exatidão?

A exatidão perfeita de uma avaliação exige:

1. Conformidade com o valor de mercado: a transação hipotética de venda do imóvel avaliando adere 100% à definição dada de valor de mercado
2. Desenvolvimento de modelo exato e robusto: desenvolve-se um modelo MARS de alta qualidade, com R2 em torno de 75-80% e CVR2 de aproximadamente 55-70%, conforme a localização.
3. Boa análise de erros nos dados: todos os prováveis erros de dados das variáveis medidas são compreendidos e contidos, com pouco transbordamento para os resíduos.
4. Consideração de vieses sistêmicos: qualquer viés sistêmico nos dados deve ser contabilizado e ajustado.
5. Análise da ordenação dos resíduos: assegurar que os imóveis, na ordenação por resíduo, sigam um padrão razoável de características, do menor ao maior posto. Anomalias como vendas de inventário, vendas em short sale e leilões devem ser investigadas, explicadas e possivelmente removidas.
6. Posicionamento do avaliando: com uma compreensão completa da ordenação dos resíduos, o imóvel avaliando deve ser posicionado com exatidão. Devem ser identificados um imóvel de apelo significativamente maior e outro de apelo significativamente menor, posicionando-se o avaliando entre eles e pontuando-o. Depois, procure estreitar ao máximo a faixa da ordenação entre os imóveis superior e inferior.
7. Raciocínio explicativo: nunca sacrifique o raciocínio explicativo em nome da exatidão. O avaliador precisa ser capaz de explicar as diferenças de valor entre o avaliando e os imóveis comparáveis, ao menos a alguém de inteligência razoável. Por exemplo, um bom modelo MARS com R2 e CVR2 ótimos pode ser usado para explicar as contribuições de valor das características medidas, ao passo que a decomposição do resíduo pelo engenheiro de avaliações pode ser usada para explicar as contribuições de valor das demais características.
8. Sobreajuste: evite modelos sobreajustados, não usando características do imóvel capazes de identificar, isoladamente ou em combinação, um ou vários imóveis. Em particular, evite usar variáveis fator cujos valores possam especificar um número pequeno de imóveis. Uma variável fator que identifique bairros em conjuntos de dados de tamanho considerável é apropriada, mas uma que identifique imóveis com piscina em uma área em que existam pouquíssimas piscinas precisa ser usada com cautela. Uma característica que identifique imóveis únicos, como coordenadas GIS exatas, números de matrícula ou endereços, é absolutamente proibida.

Se o R2 da regressão é 80%, esperamos que cerca de 20% do valor do avaliando se deva ao resíduo. Ora, isso é apenas aproximado. Sabemos, pela curva CQA-Resíduo, que a maior parte da distribuição dos resíduos está nos extremos inferior e superior do CQA. Com base na curva CQA-Resíduo e na inclinação extrema nos valores baixos e altos de CQA, poderíamos esperar uma estimativa menos exata do resíduo nessas faixas. Contudo, isso é contrabalançado pela relativa escassez de imóveis nessas faixas e pela maior diferença de apelo que se observa entre imóveis adjacentes. Assim, embora possamos falar do impacto de uma faixa de $\pm 5\%$ na região central, nos extremos estaríamos mais preocupados com uma faixa um pouco mais estreita, de $\pm 2 - \pm 3\%$

Calcular a exatidão do RCA é, em geral, algo difícil, em razão da não linearidade da curva CQA-Resíduo. Mas é possível caso a caso, desde que se disponha de uma ordenação de resíduos razoável. A amplitude dos escores CQA é de 0,0 a 10,0. Com um bom modelo MARS, sua ordenação de resíduos deve mostrar um padrão claro do menor ao maior apelo, e você não deve ter dificuldade em pontuar dentro de $\pm 5\%$ da amplitude total, ou seja, ± 0.5 no escore CQA. Se houver 100 imóveis na regressão, isso equivaleria a classificar o imóvel avaliando com exatidão dentro de um grupo de 10 imóveis. Assim, com referência aos dados de vendas de Burlingame da Tabela 3, calculemos a exatidão (ou o erro) aproximada para um CQA estimado pelo VE de 7,5, de 5,1 e de 1,9, sendo sua conclusão de valor para o avaliando de \$3,000,000. Note que, como a Tabela 3 contém apenas uma amostra dos várias centenas de imóveis utilizados, será necessário interpolar o resíduo quando cabível (supondo linearidade). Para ser claro, estamos supondo que o VE esteja muito seguro, em cada caso, de que sua classificação do avaliando se encontra dentro de uma faixa de ± 5 imóveis na ordenação de resíduos de apenas 100 imóveis.

CQA= 7.5 ± 0.5					
CQA	Resíduo	Dif. abs.	Erro esp.	Preço hipot. de venda	Erro pct.
8.00	\$257,112				
7.50	\$197,113	\$59,999			
7.00	\$154,615	\$42,498	$\pm \$51,249$	\$3,000,000	$\pm 1.71\%$

CQA=5.1 ± 0.5					
CQA	Resíduo	Dif. abs.	Erro esp.	Preço hipot. de venda	Erro pct.
5.60	\$50,911				
5.10	\$19,559	\$31,352			
4.60	-\$28,207	\$47,766	$\pm \$39,559$	\$3,000,000	$\pm 1.32\%$

CQA= 1.9 ± 0.5					
CQA	Resíduo	Dif. abs.	Erro esp.	Preço hipot. de venda	Erro pct.
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	$\pm \$49,264$	\$3,000,000	$\pm 1.64\%$

Por fim, o único modo objetivo de verificar a exatidão é estabelecer um protocolo estrito (que este artigo apenas esboça em linhas gerais) e verificar sua observância. Isso é necessário porque os preços de venda efetivos podem desviar-se do valor de mercado em razão de imperfeições do mercado. Comparar avaliações e vendas subsequentes pode fornecer alguma indicação de exatidão. Contudo, as complexidades e a instabilidade do mercado imobiliário exigem a expertise de avaliadores ou de engenheiros de avaliações altamente qualificados para fornecer estimativas de valor baseadas em normas e protocolos estabelecidos. Estes devem ser rigorosos o bastante para assegurar que engenheiros de avaliações competentes cheguem independentemente a conclusões de valor quase idênticas. Essa deve ser uma meta alcançável na maioria dos casos em que haja bons dados disponíveis.

11.5 O Imóvel Avaliando Como Anomalia

Uma vez que estimar o valor de mercado de um imóvel exige a análise de transações de venda recentes, é justo perguntar se as transações recentes representam bem as diversas características de um imóvel avaliando mais antigo, carente de modernização e reparos. Isso é particularmente verdadeiro na San Francisco Bay Area, onde as transações de venda recentes tipicamente são de

casas que passaram por reformas para melhorar seu preço de venda. Assim, o padrão construtivo e o estado de conservação médios dos dados de mercado são em geral superiores aos de casas mais antigas que não estão no mercado há bastante tempo. Essa diferença de valor deve ser captada pela estimativa do resíduo do imóvel avaliando, ou seja, por seu escore CQA. Trata-se de uma consideração importante quanto à qualidade dos dados e à exatidão na avaliação do imóvel avaliando.

11.6 A Importância do CVR2 no Trato de Dados Inesperados

A regressão MARS baseia-se em múltiplas iterações do R/earth aplicadas a um subconjunto de imóveis selecionado aleatoriamente, que tipicamente compreende de 80% a 90% do conjunto de dados do MLS, denominado conjunto de treino. Os 10%-20% restantes de imóveis servem como conjunto de teste. Nessa fase, o modelo MARS é criado a partir do conjunto de treino e depois usado para prever os preços de venda dos imóveis do conjunto de teste, os quais são então comparados aos preços de venda efetivos. Essa metodologia tem sido reconhecida pelos profissionais de mineração de dados como a estratégia mais eficaz para desenvolver um modelo robusto, capaz de prever os preços de venda de imóveis ainda não observados.

Contudo, a principal utilidade da métrica CVR2 não está em refletir a exatidão factual do modelo de regressão quanto às vendas efetivas da região de mercado, mas sim em otimizar as capacidades preditivas do modelo quanto a vendas hipotéticas de imóveis em uma dada data-base. Esses imóveis não estão incluídos no conjunto de treino e podem apresentar combinações novas ou inesperadas de características, ausentes tanto nos dados do conjunto de treino quanto nos do conjunto de teste.

Nesse contexto, como podemos estimar o preço de venda de um imóvel que não passa por reforma há 15 anos, supondo que tenha sido anunciado, por exemplo, um mês antes da data-base da avaliação? Um modelo que demonstre CVR2 forte tem maior probabilidade de fornecer a estimativa mais exata do que um modelo que não tenha passado por essa validação cruzada repetitiva. É também vital que o conjunto de dados inclua um número suficiente de imóveis comparáveis com graus variados de reforma, reparo e modernização.

Em última análise, se dispusermos de um modelo robusto que preveja com exatidão as contribuições de valor de diversas características dos imóveis em nosso conjunto de dados de comparáveis, poderemos posicionar efetivamente o novo imóvel avaliando dentro da ordenação desses imóveis por resíduo produzida pelo modelo. Esse posicionamento nos permite atribuir ao imóvel um escore CQA e um resíduo apropriados.

12 Resíduos

Como indica o nome “Residual Constraint Approach”, os resíduos são centrais para esta metodologia. Aqui, resíduo refere-se à diferença entre o preço de venda líquido de um imóvel e o preço de venda previsto pela análise de regressão.

Um resíduo positivo sugere que o imóvel foi vendido por mais do que seu preço estimado, com base em atributos mensuráveis típicos como a área bruta habitável (GLA) e a área do terreno, o que implica que ele talvez tenha apelo acima da média. Um resíduo negativo indica que o imóvel foi vendido por menos, sugerindo potencialmente menor apelo ou outros atributos negativos não captados.

Dado que casas maiores provavelmente têm resíduos maiores, podemos usar o resíduo por pé quadrado com base na GLA para uma comparação mais normalizada. Uma discussão mais aprofundada dessa questão fica reservada a outro artigo.

12.1 Ordenação e Pontuação

Depois de criado o modelo MARS (Multivariate Adaptive Regression Splines) para as variáveis medidas, calculam-se os resíduos de cada imóvel. Esses resíduos representam a diferença entre os preços de venda líquidos observados e os preços de venda previstos pelo modelo. Deve-se criar uma coluna adicional para o Resíduo/SF, isto é, o resíduo dividido pela área habitável da residência, entendendo-se que, em geral, o resíduo de determinados imóveis é mais significativo conforme o tamanho da área habitável. Outras estatísticas possíveis podem funcionar melhor em alguns bairros, mas são assunto para outra ocasião.

Premissa 7: “Ordenação por Resíduo”

Neste artigo, as referências a “ordenação por resíduo” devem ser entendidas como ordenação por resíduo, por resíduo/SF, ou ordenação por alguma outra função dos resíduos, conforme o engenheiro de avaliações julgue mais apropriado à região de mercado. A experiência do autor na SF Bay Area indica que o resíduo/SF é em geral o mais confiável, especialmente quando se trata de um imóvel avaliando com GLA (Gross Living Area) incomumente pequena ou grande.

Em seguida, os imóveis são ordenados por seus resíduos, do mais negativo ao mais positivo. Gera-se então um escore de estado de conservação, padrão construtivo e apelo (Condition-Quality-Appeal, CQA) para cada imóvel, variando de 0,0 a 10,0. Esse escore é calculado da seguinte forma:

O escore CQA corresponde ao posto percentílico do resíduo de um imóvel entre todos os imóveis. Especificamente, o escore é calculado determinando-se o percentual de imóveis com resíduos inferiores ao do imóvel em questão e dividindo-se esse percentual por 10. Por exemplo, se o resíduo de um imóvel é superior aos resíduos de 50% dos imóveis, seu escore seria 5,0.

Em razão do arredondamento, um imóvel pode receber escore 10,0, representando o imóvel de maior resíduo. Contudo, um escore perfeito de 10,0 é teoricamente impossível porque, logicamente, nenhum imóvel pode ter resíduo superior ao de 100% dos imóveis, ele próprio incluído. Por outro lado, o escore mínimo possível de 0,0 é viável, pois ao menos um imóvel não tem nenhum outro imóvel com resíduo inferior. Note que mais de um imóvel pode ter escore 0,0 por arredondamento, se muitos imóveis estiverem sendo comparados.

Portanto, o escore CQA fornece uma medida intuitiva de quão bem os atributos de um imóvel se alinham às expectativas do modelo, com escores mais altos indicando melhor alinhamento

com as previsões do modelo.

É essencial reconhecer que o escore CQA só pode ser considerado indicador indireto do apelo de um imóvel comparável quando as condições da venda são compatíveis com seu valor de mercado. Cabe ao engenheiro de avaliações apurar se o valor residual incorpora elementos acima ou abaixo do mercado, fornecendo explicação abrangente e investigação minuciosa da transação de venda. Embora vendas significativamente acima ou abaixo do mercado não sejam frequentes, elas podem ocorrer em circunstâncias como vendas forçadas, inventário, incompetência, conluio etc.

A Tabela 3 apresenta um extrato semirrealista de dados do MLS, junto com os resíduos e os escores CQA, tanto totais quanto por pé quadrado. O número total de imóveis analisados neste estudo foi de aproximadamente 600. É digno de nota que várias variáveis submetidas à regressão não são exibidas. Em outras palavras, os resíduos apresentados são posteriores à consideração, pela regressão MARS, do impacto de outras variáveis, como a data da venda.

Tabela 3: Dados de mercado ordenados por resíduo

SalePrice	Estimated SP	Residual	CQA	Residual/SF	CQA SF	GLA	SaleDate
\$3,450,000	\$2,318,865	\$1,131,135	10	\$420.50	9.98	2,690	2021-04
\$4,360,757	\$3,399,922	\$960,835	9.9	\$290.28	9.82	3,310	2021-09
\$4,000,000	\$3,134,444	\$865,556	9.8	\$234.57	9.60	3,690	2020-11
\$5,850,000	\$5,150,966	\$699,034	9.7	\$109.91	8.35	6,360	2021-12
\$3,600,000	\$3,023,366	\$576,634	9.5	\$269.46	9.79	2,140	2022-06
\$4,425,000	\$3,987,873	\$437,127	9.2	\$127.44	8.68	3,430	2021-08
\$2,530,000	\$2,164,564	\$365,436	8.9	\$180.02	9.28	2,030	2019-12
\$4,250,000	\$3,884,719	\$365,281	8.8	\$87.18	7.91	4,190	2021-08
\$3,350,000	\$3,034,613	\$315,387	8.5	\$130.87	8.75	2,410	2021-09
\$3,800,000	\$3,522,028	\$277,972	8.2	\$77.43	7.72	3,590	2020-09
\$4,200,000	\$3,963,748	\$236,252	7.8	\$56.12	6.89	4,210	2021-01
\$2,900,000	\$2,702,887	\$197,113	7.5	\$89.19	7.92	2,210	2021-07
\$4,250,000	\$4,103,885	\$146,115	6.9	\$39.38	6.29	3,710	2022-08
\$5,000,000	\$4,864,958	\$135,042	6.8	\$28.25	5.85	4,780	2021-07
\$4,320,000	\$4,199,672	\$120,328	6.5	\$33.90	6.04	3,550	2022-07
\$2,600,000	\$2,479,904	\$120,096	6.5	\$60.96	7.00	1,970	2021-10
\$3,400,000	\$3,301,146	\$98,854	6.2	\$30.32	5.90	3,260	2021-06
\$2,000,000	\$1,901,987	\$98,013	6.2	\$73.14	7.54	1,340	2019-11
\$3,000,000	\$2,919,718	\$80,282	5.9	\$26.50	5.79	3,030	2020-05
\$6,750,000	\$6,699,089	\$50,911	5.6	\$7.75	5.20	6,570	2020-08
\$5,652,000	\$5,601,570	\$50,430	5.5	\$12.83	5.37	3,930	2022-05
\$2,600,000	\$2,579,082	\$20,918	5.2	\$10.51	5.33	1,990	2022-03
\$1,925,000	\$1,905,441	\$19,559	5.1	\$11.51	5.36	1,700	2021-01
\$3,200,000	\$3,209,500	-\$9,500	4.8	-\$3.21	4.88	2,960	2020-11
\$4,150,000	\$4,187,560	-\$37,560	4.5	-\$10.49	4.68	3,580	2021-04
\$2,700,000	\$2,776,510	-\$76,510	3.9	-\$23.61	4.14	3,240	2021-07
\$3,450,000	\$3,567,668	-\$117,668	3.4	-\$31.05	3.90	3,790	2020-09
\$1,975,000	\$2,109,255	-\$134,255	3.3	-\$55.48	3.24	2,420	2020-02
\$1,950,000	\$2,085,431	-\$135,431	3.2	-\$68.40	2.86	1,980	2021-02
\$3,695,000	\$3,859,154	-\$164,154	3.1	-\$47.04	3.56	3,490	2021-05
\$2,288,000	\$2,455,747	-\$167,747	2.9	-\$56.86	3.19	2,950	2020-06
\$2,635,000	\$2,802,781	-\$167,781	2.9	-\$56.87	3.18	2,950	2020-09
\$2,050,000	\$2,217,794	-\$167,794	2.9	-\$82.66	2.46	2,030	2021-07
\$1,550,000	\$1,731,804	-\$181,804	2.7	-\$130.79	1.32	1,390	2021-05
\$2,800,000	\$3,028,411	-\$228,411	2.4	-\$81.58	2.51	2,800	2021-02
\$3,700,000	\$3,976,381	-\$276,381	1.9	-\$76.99	2.62	3,590	2021-08
\$2,517,500	\$2,834,974	-\$317,474	1.5	-\$120.71	1.48	2,630	2019-11
\$2,550,000	\$2,905,330	-\$355,330	1.1	-\$115.74	1.64	3,070	2019-12
\$1,815,000	\$2,263,944	-\$448,944	0.7	-\$273.75	0.15	1,640	2021-11
\$3,711,000	\$4,683,454	-\$972,454	0	-\$226.68	0.37	4,290	2021-12
\$2,225,000	\$3,332,960	-\$1,107,960	0	-\$370.56	0.03	2,990	2021-08
\$3,550,000	\$4,746,489	-\$1,196,489	0	-\$274.42	0.14	4,360	2022-06

12.2 Curva CQA-Resíduo

Um gráfico do CQA versus o Resíduo/SF de um bairro resultará invariavelmente na curva característica CQA-Resíduo mostrada na [Figure 2](#). Este gráfico em particular representa dados de Burlingame, CA, por volta de 2022. Curvas semelhantes são em geral observadas ao se plotar escores CQA contra valores de resíduo; contudo, prefiro usar o Resíduo/SF por sua pertinência em contextualizar o CQA em relação à área bruta habitável (GLA) do imóvel avaliando. Multiplicando-se o Resíduo/SF pela GLA, pode-se converter o CQA de um imóvel em um valor residual correspondente.

O engenheiro de avaliações precisa avaliar cuidadosamente se o tamanho do imóvel, especificamente a área bruta habitável (GLA), influencia de modo significativo a magnitude dos resíduos. Essa aferição é crítica porque pode determinar o método mais exato de cálculo dos valores dos imóveis. Por exemplo, eletrodomésticos de cozinha de alto padrão e louças e metais de banheiro luxuosos podem elevar substancialmente o valor residual de uma casa. Contudo, em bairros nos quais o tamanho e o leiaute de cozinhas e banheiros são geralmente uniformes, essas melhorias podem não se correlacionar fortemente com a GLA. Seu impacto sobre o valor residual pode ter menos a ver com o espaço que ocupam e mais com a qualidade que representam.

Inversamente, características e materiais externos caros, que melhoram notavelmente o apelo visual do imóvel, podem ter relação mais direta com o tamanho do imóvel. Casas maiores comportam exteriores mais elaborados, como paisagismo extenso, materiais de cobertura superiores ou terraços amplos, que naturalmente ampliam o apelo geral do imóvel e, por consequência, seu valor residual.

Portanto, é essencial que os engenheiros de avaliações considerem essas dinâmicas ao analisar resíduos. Eles devem não apenas aferir o valor intrínseco agregado por acabamentos e características de alta qualidade, mas também compreender como essas melhorias interagem com as dimensões físicas do imóvel. Essa abordagem abrangente assegura que as avaliações reflitam com exatidão tanto os atributos qualitativos quanto os quantitativos de um imóvel.

Em consequência, recomenda-se plotar tanto a curva CQA-Resíduo quanto a curva CQA-Resíduo/SF e estudar as diferenças, que podem indicar padrões de uma região de mercado. O engenheiro de avaliações pode fazer seu programa calcular ambos os resíduos e tomar alguma média ponderada para calcular o resíduo a ser atribuído ao avaliando. Pode ou não fazer muita diferença.

12.3 As Características da Curva CQA

A curva de estado de conservação, padrão construtivo e apelo (CQA) oferece um modelo sofisticado para compreender como os valores dos imóveis flutuam em função do apelo, do estado de conservação e da dinâmica de mercado dentro de uma dada área. Esse modelo capta a interação de fatores econômicos e físicos que influenciam o mercado imobiliário, refletindo realidades matizadas da avaliação de imóveis nos diferentes segmentos de mercado.

12.3.1 Análise dos Componentes da Curva CQA

1. Decaimento exponencial no lado esquerdo da curva:

À medida que o escore CQA decresce de cerca de 2 para 0, os imóveis exibem características de rápida deterioração. Essa parte da curva é bem representada por uma função de decaimento exponencial, evidenciando com que rapidez um imóvel pode perder apelo e valor sem manutenção e reformas adequadas. A degradação física, como apodrecimento

da madeira, desbotamento de cores e enferrujamento de metais, acelera esse declínio, particularmente em casas que não recebem manutenção regular.

2. Crescimento linear na faixa intermediária:

O grosso do mercado tipicamente se enquadra nesta categoria, na qual o apelo cresce quase linearmente de escores em torno de 2 a 8. Esta seção reflete imóveis geralmente mantidos dentro do padrão, mas que não apresentam necessariamente qualidades ou localizações excepcionais. Representa a ampla classe média habitacional, abrangendo a maior parte do mercado de moradia, no qual as casas são precificadas competitivamente com base em suas características padrão e em seu estado geral.

3. Crescimento exponencial no lado direito da curva:

Aqui, a curva sobe exponencialmente dos escores 8 a 10, espelhando a extremidade de maior apelo do mercado. Os imóveis desse segmento com frequência pertencem a pessoas mais abastadas, que investem não apenas em manutenção superior, mas também em melhorias que ampliam significativamente o apelo e o valor do imóvel. O crescimento exponencial reflete a escassez de imóveis de tão alta qualidade e a forte demanda entre compradores de alta renda. E entenda: uma casa de preço mais baixo pode receber nota muito alta em termos de apelo, ao passo que uma casa grande e cara pode receber nota baixa. Essas dinâmicas são reveladas pelo uso especializado do MARS e geralmente passam despercebidas pela maioria dos avaliadores.

12.3.2 Dinâmica de Mercado e Implicações Sociais

A curva CQA narra implicitamente a história da riqueza relativa e da disparidade econômica em diferentes regiões. Em lugares como o condado de San Mateo, a variação dos valores dos imóveis entre áreas como Burlingame e bairros mais ricos, como Hillsborough, Woodside ou Atherton, ilustra divisões socioeconômicas significativas. Essa disparidade afeta tudo, da manutenção dos imóveis às avaliações de mercado, e em última instância influencia o perfil demográfico dos compradores e as tendências do mercado imobiliário.

Além disso, o caso de uma pessoa relativamente abastada que investe em uma casa de alta qualidade dentro de um bairro de faixa intermediária com o qual se sente à vontade e no qual, por diversas razões, prefere morar, resalta outro aspecto crítico da economia imobiliária — o conceito de imóveis “supervalorizados por sobreconstrução” (*overbuilt*). Tais imóveis, embora muito atraentes, podem não encontrar compradores locais capazes de pagar o preço que alcançariam em uma região de mercado de preços mais altos.

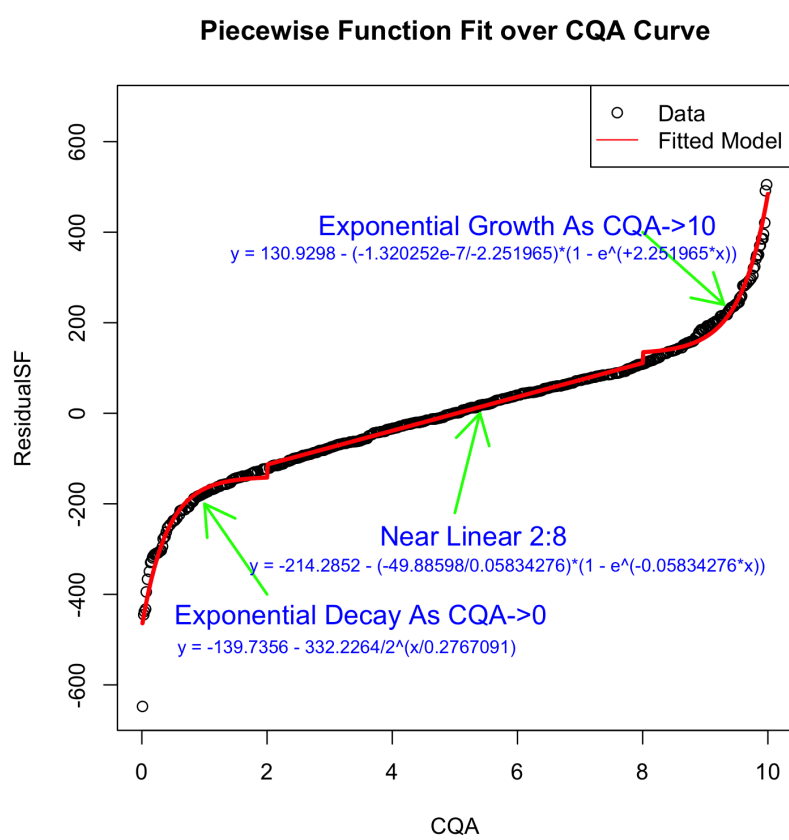
A curva CQA-Resíduo é característica de cada região de mercado e não deve ser negligenciada na sua análise.

12.4 Resíduo do Avaliando

Diferentemente dos comparáveis, o imóvel avaliando não possui resíduo determinado objetivamente; ele precisa ser estimado pelo engenheiro de avaliações (VE). Essa estimativa introduz um grau de subjetividade no protocolo do Residual Constraint Approach (RCA), marcando um ponto crítico em que o juízo pessoal desempenha papel decisivo.

Para estimar o escore de estado de conservação, padrão construtivo e apelo (CQA), o VE analisa meticulosamente os imóveis ordenados, buscando padrões discerníveis dentro da classificação. Essa análise frequentemente envolve exame detalhado das fotos do MLS associadas a cada imóvel. Um exame exaustivo é impraticável para um conjunto de dados grande, como o de 600 imóveis.

Figura 2: Função Característica CQA-Resíduo



Em vez disso, o VE aplica raciocínio prático para aproximar a provável posição do avaliando na classificação geral. Imóveis que exigem reparos extensos, ou “fixers” (imóveis para reforma), são geralmente posicionados perto do fim, ao passo que aqueles com melhores reformas e apelo tendem a se classificar perto do topo.

O VE então adota uma abordagem mais focada, “saltando” dentro da classificação para identificar comparáveis que correspondam de perto ao apelo do avaliando. À medida que a faixa se estreita, o VE presta atenção maior às semelhanças entre os imóveis, particularmente em aspectos como padrão construtivo, estado de conservação, estética do projeto (incluindo elementos como pintura, marcenaria, cantaria e esquadrias), utilidade funcional e outras características pertinentes

É crucial entender que os juízos do VE são fundamentalmente guiados pelas classificações estabelecidas pelo mercado. Isso assegura que os elementos subjetivos do processo fiquem ancorados nas realidades de mercado, buscando uma avaliação justa e exata do imóvel avaliando.

12.5 Conclusão de Valor do RCA

12.5.1 Considerações Gerais

Determinar um escore CQA é decisivo para estabelecer o valor residual estimado de um imóvel, pois esses elementos estão intrinsecamente ligados. Embora seja possível avaliar um imóvel sem CQA, essa métrica fornece contexto valioso ao ilustrar onde o imóvel se situa em comparação com os demais da região de mercado. Por exemplo, um escore CQA de 3,2 indica que o imóvel tem apelo maior do que 32% dos imóveis comparáveis das imediações. Se houver objeções ao escore atribuído, os contestantes precisam consultar as classificações e justificar quaisquer ajustes propostos com base em evidência fotográfica ou em outros dados substanciais. Se o sistema de ordenação for robusto e o engenheiro de avaliações (VE) tiver posicionado o imóvel com exatidão dentro desse arcabouço, é improvável que surjam discrepâncias significativas.

Na fase de elaboração do laudo, tipicamente são selecionados de 6 a 12 dados de mercado recentes e pertinentes para embasar a conclusão de valor. As diferenças entre esses comparáveis e o imóvel avaliando são analisadas, com cada variável sendo ajustada em conformidade. Esses ajustes são então aplicados aos preços de venda líquidos dos comparáveis para se obterem os preços de venda ajustados. Como demonstrado matematicamente, esses preços ajustados devem convergir para o preço de venda estimado do imóvel avaliando.

12.5.2 Revisão, Auditoria, Reclamações

Uma vez determinado o resíduo estimado do imóvel avaliando, os ajustes residuais podem ser calculados para todos os comparáveis. Esses ajustes servem como restrições para decomposições manuais adicionais pelo VE, adaptadas às variáveis CQA específicas consideradas pertinentes. É importante enfatizar que essas decomposições detalhadas são explicativas e não influenciam a avaliação final, como afirmado na Demonstração II.

Quanto a reclamações apresentadas por usuários insatisfeitos com a conclusão de valor da avaliação, um avaliador tradicional tipicamente precisa perder tempo explicando por que este ou aquele imóvel de preço mais alto ou mais baixo não tem o impacto sobre o valor que o usuário esperava. Com o RCA, todas as vendas comparáveis possíveis (normalmente de 100 a 600) são tipicamente avaliadas em precisamente o mesmo valor da conclusão de preço, de modo que é pouco custoso mostrar os ajustes de um imóvel qualquer. Todo o processo RCA é automatizado e objetivo, exceto pela pontuação do resíduo do avaliando. E assim outra vantagem do RCA é que a crítica a imperfeições no juízo subjetivo do VE fica restrita principalmente a essa única área — e pode ser gerida de forma eficaz e eficiente. Além disso, supondo um R2 em torno de

80%, o impacto de um erro de $\pm 10\%$ na estimativa do resíduo do avaliando corresponde a um erro de $\pm 2\%$ na estimativa final de valor.

Em muitos casos, o avaliando se encaixa tão perfeitamente na ordenação dos resíduos dos imóveis que se pode esperar exatidão dentro de $\pm 1\%$.

A soma do resíduo estimado com o preço de venda estimado pelo MARS para o avaliando produz a conclusão final de valor, conhecida como Conclusão de Valor do Residual Constraint Approach (RCA). Esse processo assegura uma abordagem abrangente e transparente da avaliação de imóveis, promovendo confiança na exatidão e na justeza da apuração.

13 Conclusão

O Residual Constraint Approach é uma estrutura metodológica com três contribuições substantivas: uma propriedade estrutural antiancoragem, derivada do ajuste do modelo sobre todo o mercado, e não sobre um pequeno conjunto de comparáveis (Demonstração I); um resultado de invariância que permite à decomposição do resíduo servir como camada explicativa sem perturbar a conclusão de valor (Demonstração II e a observação que a segue); e uma regularidade empírica — a curva característica CQA-Resíduo — que parece recorrer em diferentes regiões de mercado sob a forma característica exponencial-linear-exponencial. As limitações da estrutura são igualmente substantivas: a decomposição do resíduo é invariante, mas não identificadora; a etapa de posicionamento CQA apoia-se em juízo do VE que ainda não foi validado por estudo formal de concordância entre avaliadores; e as alegações de exatidão aguardam comparação com alternativas.

O que este artigo apresenta é uma estrutura metodológica e um protocolo. O que ele ainda não apresenta, mas o programa de pesquisa mais amplo agora em curso apresenta, é um pipeline executável e auditável que operacionaliza o protocolo em software com padrões defensáveis. Esse pipeline — o pacote earthUI no CRAN para a etapa das funções de base do MARS, o pacote glmnetUI para o modelo estrutural regularizado sobre essas bases e o pacote mgcvUI para a suavização espacial — é objeto de artigos companheiros nesta edição e constitui a ponte do RCA-como-método ao RCA-como-entregável. A lacuna entre doutrina e prática na regulação da atividade de avaliação (Craytor, 2026, nesta edição) fornece a razão pela qual essa ponte importa: os produtos da avaliação são economicamente consumidos como insumos prospectivos de risco, e uma estrutura que torne a opinião de valor presente auditável e reproduzível é pré-condição para qualquer engajamento honesto com esse uso prospectivo. O RCA é um movimento nesse projeto maior. Não é o projeto inteiro, e este artigo não pretende que seja.

Referências

- Ambrose, Brent W. et al. (2025). “Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation?” Em: *Review of Finance*. DOI: [10.1093/rof/rfaf046](https://doi.org/10.1093/rof/rfaf046).
- Craytor, William Bert (2026). “The Doctrine–Practice Gap in Real Estate Appraisal: A Structured Account of Functions, Boundaries, and Tensions”. Em: *Valuation Engineer Journal* 1.1.
- Friedman, Jerome H. (1991). “Multivariate Adaptive Regression Splines”. Em: *The Annals of Statistics* 19.1, pp. 1–141.
- Hastie, Trevor, Robert Tibshirani e Jerome H. Friedman (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second. Springer Series in Statistics. New York, NY: Springer. ISBN: 978-0-387-84857-0.
- James, Gareth et al. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics 103. New York: Springer. ISBN: 978-1-4614-7137-0.
- Krivorotov, George e Michael LaCour-Little (2020). “AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias”. Em: *SSRN Electronic Journal*. DOI: [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- Kuhn, Max e Kjell Johnson (2016). *Applied Predictive Modeling*. Corrected at 5th printing. New York: Springer. ISBN: 978-1-4614-6848-6.
- Milborrow, Stephen (2024a). *Earth: Multivariate Adaptive Regression Splines*. R Package. URL: <https://CRAN.R-project.org/package=earth>.
- (2024b). *Notes on the Earth Package*. Rel. técn. URL: <http://www.milbo.org/doc/earth-notes.pdf>.

- Milborrow, Stephen (2024c). *Variance Models in Earth*. Rel. técn. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.
- Perry, Andre M., Jonathan Rothwell e David Harshbarger (2018). *The Devaluation of Assets in Black Neighborhoods*. Rel. técn. Brookings Institution.
- Perry, Andre M., Jonathan Rothwell, J. Williamson et al. (2024). *How Racial Bias in Appraisals Affects the Devaluation of Homes in Majority-Black Neighborhoods*. Rel. técn. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Rel. técn. U.S. Department of Housing e Urban Development.
- Tzioumis, Konstantinos (2016). “Appraisers and Valuation Bias: An Empirical Analysis”. Em: *Real Estate Economics* 45.3, pp. 679–712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake e Mark Palim (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Rel. técn. Fannie Mae.
- Zhang, Wengang (2020). *MARS Applications in Geotechnical Engineering Systems: Multi-Dimension with Big Data*. Singapore: Springer. ISBN: 978-981-13-7421-0.

earthUI: uma interface interativa para o pacote earth (MARS)

Tradução para o português, elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

William Bert Craytor

Volume 1, Número 1 · Verão de 2026

Publicado em 1 de julho de 2026

DOI: [10.67330/7fj41b65](https://doi.org/10.67330/7fj41b65)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da [licença Creative Commons Atribuição 4.0](#).

Resumo. earthUI é uma interface gráfica do usuário para o pacote **earth** do R, que implementa Multivariate Adaptive Regression Splines (MARS). Oferece três modos de uso — modelagem preditiva geral, avaliação de imóveis e análise de região de mercado — e conduz o usuário pela importação de dados, pela configuração do modelo, pelo ajuste e pela interpretação dos diagnósticos e das saídas gráficas do earth. Este artigo documenta os requisitos de formato de dados do earthUI, o fluxo de trabalho de modelagem, as telas de saída e a referência completa de funcionalidades.

Palavras-chave: MARS, splines de regressão adaptativa multivariada, earth, R, interface gráfica do usuário, avaliação de imóveis

Nota do editor (julho de 2026): Este artigo descreve o earthUI 0.10.0. Pouco antes da publicação, o earthUI 0.11.0 absorveu os aplicativos complementares glmnetUI e mgcvUI (descritos nos dois artigos seguintes), consolidando os fluxos de trabalho de MARS, elastic net e GAM em um único aplicativo. A interface e o fluxo de trabalho aqui descritos não são afetados pela consolidação.

Sumário

1	Introdução	4
1.1	O que é o earthUI?	4
1.2	Um pacote, três rotinas	4
1.3	Três modos de finalidade	4
1.4	Recursos específicos para imóveis	5
1.5	Instalação do earthUI e do vProlog	5
1.6	Primeiros passos	6
2	Requisitos dos dados de entrada do MLS	7
2.1	Formato e estrutura do arquivo	7
2.2	Colunas necessárias no modo de avaliação	7
2.3	Convenções de nomenclatura de colunas especiais	8
2.4	Qualidade e completude dos dados	8
2.5	Posicionamento da linha do avaliando	9
3	Modo de finalidade geral	10
3.1	Visão geral	10
3.2	O fluxo de trabalho da barra lateral	10
3.3	Abas do painel principal	10

3.4	Persistência das configurações	11
3.5	Modo escuro	11
3.6	Localidade e configurações regionais	11
3.6.1	Países suportados	12
3.6.2	Sobreposição do papel	12
3.6.3	Salvando padrões	12
4	Modo de avaliação	14
4.1	Tratamento da linha do avaliando	14
4.2	Data de referência e idade da venda	14
4.3	Designações especiais de colunas	14
4.4	Contribuição residual (estimação de características raras)	15
4.5	Visão geral dos ajustes RCA	16
5	Modo de análise de região de mercado	17
5.1	Diferenças em relação ao modo de avaliação	17
5.2	Quando usar o modo de mercado	17
6	Seleção de variáveis	18
6.1	Variável(is)-resposta	18
6.2	A tabela de preditores	18
6.3	Deteção de tipos de dados e sobreposições	19
6.4	Marcadores Factor e Linear	19
6.5	Referência de tipos especiais de coluna	19
6.6	Múltiplas respostas	20
7	Seleção de parâmetros	21
7.1	Parâmetros básicos	21
7.2	Passagem direta (forward pass)	21
7.3	Interações permitidas	22
7.4	Poda	22
7.5	Validação cruzada	22
7.6	Filtragem por subconjunto	23
7.7	Valores recomendados de parâmetros	23
7.7.1	Parâmetros da passagem direta	23
7.7.2	Parâmetros de poda	25
7.7.3	Parâmetros de validação cruzada	25
7.7.4	Modelo de variância	26
7.7.5	Arquivo de saída do Earth	26
7.8	Padrões das configurações	26
8	Ajustando o modelo Earth	27
8.1	O botão Fit	27
8.2	Janela modal de ajuste e saída de rastreamento	27
8.3	Log de ajuste	27
9	Baixando os dados	28
9.1	Colunas de saída	28
9.2	Ordenação das colunas e formatação do Excel	28
9.3	Escores CQA	29
9.4	Ordenação e linha do avaliando	29
9.5	Exportação automática para o mgcvUI	29

9.6	Importação no glmnetUI	30
9.6.1	Principais benefícios	30
9.6.2	Importante: a coluna de pesos precisa coincidir	30
9.6.3	Fluxo de trabalho	30
10	Cálculos RCA e download	31
10.1	Abrindo a caixa de diálogo RCA	31
10.2	Interpolação do escore CQA	31
10.3	Colunas de saída	31
10.4	A aba RCA Adjustments	32
11	Grade de comparação de vendas	33
11.1	Visão geral e finalidade	33
11.2	Janela modal de seleção de comparáveis	33
11.3	Leiaute da grade	33
11.4	Preço de venda, concessões e Net SP	34
11.5	Linhas agrupadas (Location, Site, Age)	34
11.6	Linhas das variáveis do modelo	34
11.7	Linhas CQA Residual e de características residuais	35
11.8	Fórmula do preço de venda ajustado	35
11.9	Proteção da planilha e células editáveis	35
11.10	Trabalhando com a grade no Excel	35
12	Baixando relatórios	37
12.1	Generate Quarto Report	37
12.2	Convert Quarto Report	37
12.3	Conteúdo do relatório	37
13	Os Post Processing Modules: elastic net e GAM	39
13.1	Por que ajustes de apoio?	39
13.2	Inicialização e estado compartilhado	39
13.3	A rotina de elastic net	39
13.4	A rotina de GAM	39
13.5	Um fluxo de trabalho prático	40
14	Conjunto de dados de demonstração: Appraisal_1.csv	41
14.1	Descrição	41
14.2	Colunas	42
14.3	Início rápido sugerido	42
15	Experimental: extração de texto com Prolog e regras de derivação	44
15.1	O que ele faz	44
15.2	Requisitos: vProlog (Trealla), e não SWI-Prolog	45
15.3	Habilitando e executando	45
15.4	Regras de derivação	45
15.5	Arquivo de dados ativo versionado	46
15.6	Interface programática e arquivos incluídos	46
15.7	Experimento relacionado: prioris por região climática	46
15.7.1	Além da Califórnia: pacotes de jurisdição e módulos de região de mercado	46

1 Introdução

1.1 O que é o earthUI?

earthUI é uma interface gráfica do usuário para o pacote **earth** do R, que implementa Multivariate Adaptive Regression Splines (MARS) — splines de regressão adaptativa multivariada (Friedman 1991). O motor de modelagem é o pacote **earth** de Stephen Milborrow (Milborrow 2024), documentado em detalhe por Milborrow (s.d.[a]) e Milborrow (s.d.[b]). Ele roda como um aplicativo Shiny local — não há login, nem servidor, nem contas. Você o inicia a partir do R, importa um conjunto de dados (CSV ou Excel), configura seu modelo e o ajusta interativamente.

O aplicativo oferece um fluxo de trabalho completo: importação de dados, configuração de variáveis, ajuste do modelo, gráficos de diagnóstico, importância das variáveis, equações do modelo e relatórios para download em formato HTML, Word ou PDF.

1.2 Um pacote, três rotinas

A partir da versão atual, o pacote earthUI também *contém* seus dois Post Processing Modules (módulos de pós-processamento), que em versões anteriores eram os pacotes separados glmnetUI e mgcvUI. Instalar o earthUI instala as três rotinas; cada uma roda como seu próprio aplicativo Shiny local:

- **earthUI::launch()** — o aplicativo earth (MARS) descrito neste artigo, na porta 7878. **earth()** é o motor central de avaliação.
- **earthUI::launch_glmnet()** — o Post Processing Module de elastic net (porta 7879), que importa as funções de base do modelo earth e as reestima sob regularização lasso/elastic net (veja “glmnetUI Import” adiante).
- **earthUI::launch_mgcv()** — o Post Processing Module de GAM (porta 7880), que usa as posições dos nós do modelo earth como pontos de partida para os termos suavizados do mgcv (veja “mgcvUI Auto-Export” adiante).

As três rotinas compartilham uma mesma árvore de projetos, um mesmo banco de dados de configurações e uma mesma cópia da infraestrutura de apoio, de modo que um projeto aberto em um aplicativo fica imediatamente disponível para os demais. As rotinas glmnet e mgcv são evidência de apoio à conclusão de valor do modelo earth, e não fluxos de trabalho concorrentes: o padrão previsto é ajustar e refinar primeiro o modelo earth e, depois, executar os módulos sobre o mesmo projeto quando se deseja corroboração.

1.3 Três modos de finalidade

O earthUI organiza o trabalho em **projetos** — entidades nomeadas que mantêm juntos seus arquivos de entrada, saídas e configurações em um local fixo. Todo projeto carrega uma **Purpose** (finalidade), escolhida no assistente New Project quando o projeto é criado. Há três modos:

- **General** (geral) — Regressão Earth para qualquer tipo de população ou conjunto de dados. Este é o modo padrão. Fornece o fluxo de trabalho completo de modelagem MARS, sem acréscimos específicos de domínio.
- **For Appraisal** (para avaliação) — Regressão Earth adaptada à avaliação de imóveis. Acrescenta recursos específicos da avaliação de um único imóvel, incluindo o tratamento do imóvel avaliando, as designações especiais de colunas, o Residual Constraint Approach (RCA) (abordagem por restrição residual) e a geração do Sales Comparison Grid (grade de comparação de vendas).
- **Market Area Analysis** (análise de região de mercado) — Regressão Earth adaptada a estudos de regiões de mercado. Acrescenta recursos para analisar grupos de imóveis em um mercado definido, em vez de avaliar um único imóvel avaliando.

Nos três modos, o motor central de modelagem é idêntico — você está sempre ajustando um modelo Earth (MARS). A configuração de finalidade controla quais ferramentas e elementos de interface adicionais ficam disponíveis.

A finalidade é fixa por projeto. A finalidade de um projeto — junto com seu país, estado, município e cidade — fica travada quando o projeto é criado. Mudar para um projeto diferente redefine o earthUI para um estado padrão limpo (os dados importados, os resultados do modelo, as abas, a configuração de variáveis e os parâmetros do earth são todos limpos) e então restaura a configuração salva daquele projeto para sua finalidade.

1.4 Recursos específicos para imóveis

Quando **For Appraisal** ou **Market Area Analysis** está selecionado, o earthUI ativa vários recursos concebidos para a análise de imóveis:

- **Designações especiais de colunas** — Cada preditor pode ser marcado com um papel especial, como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Essas designações controlam como a coluna é tratada durante o ajuste, a saída e a geração do Sales Grid (grade de vendas). Veja o Capítulo 6 para a referência completa.
- **Arredondamento de latitude e longitude** — Colunas designadas como latitude ou longitude são automaticamente arredondadas para 3 casas decimais, a fim de evitar sobreajuste.
- **Coluna Sale Age** — Quando uma coluna é designada como `contract_date` e uma Effective Date (data de referência) é informada, o earthUI calcula uma coluna `sale_age` (dias entre a data da venda e a data de referência) e a substitui como preditor.
- **Cálculos RCA (somente avaliação)** — No modo de avaliação, depois de ajustar o modelo, o earthUI pode calcular a saída do Residual Constraint Approach (RCA), que produz ajustes por comparável, resumos de ajustes líquidos/brutos e um preço de venda ajustado para o imóvel avaliando.
- **Sales Comparison Grid (somente avaliação)** — Gera uma pasta de trabalho do Excel com várias planilhas e fórmulas para os ajustes dos comparáveis, detalhamentos residuais por característica e preços de venda ajustados. Veja o Capítulo 11.

***Dica:** O modo de finalidade General funciona com qualquer conjunto de dados — financeiro, científico, de engenharia ou imobiliário. Os modos de avaliação e de mercado apenas acrescentam recursos de conveniência para os profissionais do setor imobiliário.*

1.5 Instalação do earthUI e do vProlog

O earthUI está publicado no CRAN, de modo que a instalação padrão é uma linha no R:

```
install.packages("earthUI")
```

Este artigo descreve a versão atual. Se o CRAN ainda estiver distribuindo uma versão mais antiga quando você ler isto, a versão de desenvolvimento pode ser instalada a partir do GitHub (observe o `subdir` — o pacote fica na pasta `pkg/` do repositório):

```
remotes::install_github("wcraytor/earthUI", subdir = "pkg")
```

vProlog (opcional). O recurso experimental de processamento de observações (veja “Prolog Text Extraction” perto do fim deste artigo) usa o pacote vProlog — um motor Prolog embarcado e autocontido. Ele deliberadamente *não* é uma dependência obrigatória: o earthUI instala e roda sem ele, e os recursos de Prolog simplesmente permanecem desabilitados até que ele esteja

presente. O vProlog é distribuído pelo repositório r-universe do autor, que fornece binários pré-compilados para Windows e macOS (sem necessidade de cadeia de compilação):

```
install.packages("vProlog",  
                 repos = c("https://wcraytor.r-universe.dev",  
                           getOption("repos")))
```

Não é necessária nenhuma instalação de Prolog no sistema (SWI-Prolog ou outra) — o motor Trealla é compilado dentro do próprio pacote. Todo o restante deste artigo funciona sem o vProlog.

1.6 Primeiros passos

Para usar o earthUI:

1. **Instale o pacote** a partir do CRAN, conforme descrito acima.
2. **Inicie o aplicativo:** execute `earthUI::launch()` no R, ou, pela linha de comando, execute `Rscript -e 'earthUI::launch()'`. O aplicativo abre em seu navegador na porta 7878. Você também pode acessar o aplicativo diretamente navegando até `http://localhost:7878` em seu navegador. (Os Post Processing Modules iniciam da mesma forma: `launch_glmnet()` na porta 7879, `launch_mgcv()` na porta 7880.)
3. **Crie ou abra um projeto** na Seção 1 da barra lateral. O assistente New Project registra o nome do projeto, a Purpose (General, For Appraisal ou Market Area Analysis), a localização (país, estado, município, cidade) e, opcionalmente, um arquivo de dados inicial, que é copiado para a pasta `in/` do projeto.
4. **Importe seus dados** na Seção 2, que lista os arquivos CSV e Excel presentes na pasta `in/` do projeto ativo. Arquivos adicionados manualmente pelo Finder ou pelo Explorer aparecem depois de clicar em Refresh.
5. **Configure as variáveis** — escolha sua(s) variável(is)-resposta e os preditores, defina os tipos de dados e atribua quaisquer papéis especiais de coluna.
6. **Defina os parâmetros do Earth** — degree, nprune, filtros de subconjunto e outros argumentos de `earth()`.
7. **Ajuste o modelo** — clique em “Fit Earth Model” e examine os resultados no painel principal.
8. **Exporte** — baixe as previsões em Excel, gere pacotes de relatório Quarto ou (no modo de avaliação) calcule os ajustes RCA e as Sales Comparison Grids.

A configuração é salva no servidor, em um banco de dados de projeto, indexada por projeto e finalidade, por meio dos botões “Save current as default” das Seções 3 e 4 (veja o Capítulo 3, “Persistência das configurações”).

2 Requisitos dos dados de entrada do MLS

Para fluxos de trabalho de avaliação de imóveis e de análise de mercado, seus dados de entrada normalmente vêm de uma exportação de Multiple Listing Service (MLS). Este capítulo descreve a estrutura de arquivo esperada e as colunas que o earthUI pode utilizar.

2.1 Formato e estrutura do arquivo

O earthUI aceita arquivos **CSV** e **Excel** (.xlsx, .xls). Na importação, os nomes das colunas são automaticamente convertidos para **snake_case** — por exemplo, “Living SqFt” torna-se **living_sqft**, “Contract Date” torna-se **contract_date** e “Sale Price” torna-se **sale_price**. Essa normalização garante referências consistentes de colunas em todo o fluxo de trabalho. O separador de CSV e o marcador decimal usados durante a importação são determinados pelas configurações de localidade (veja o Capítulo 3, “Localidade e configurações regionais”).

Seu arquivo de dados deve ser uma tabela plana com uma linha por imóvel e uma coluna por atributo. A primeira linha do arquivo deve conter os cabeçalhos das colunas.

2.2 Colunas necessárias no modo de avaliação

Embora o earthUI funcione com qualquer conjunto de colunas, o fluxo de trabalho completo de avaliação (ajustes RCA + Sales Comparison Grid) se beneficia da presença das seguintes colunas em sua exportação do MLS:

Coluna	Tipo especial	Finalidade
Sale Price	(resposta)	Variável-resposta do modelo
Contract Date	contract_date	Usada para calcular sale_age (dias desde a data de referência)
Listing Date	listing_date	Usada com a data do contrato para calcular o DOM caso não exista coluna de DOM
Days on Market	dom	Dias no mercado; exibido no Sales Grid
Concessions	concessions	Concessões na venda; Net SP = Sale Price – Concessions no Sales Grid
Living Area (SF)	living_area	Habilita resíduos por pé quadrado (residual_sf , cqa_sf)
Lot Size	lot_size	Agrupada na linha "Site Size Dimensions" do Sales Grid
Site Dimensions	site_dimensions	Agrupada com o tamanho do lote no Sales Grid
Latitude	latitude	Arredondada para 3 casas decimais; usada no cálculo de proximidade e no grupo Location
Longitude	longitude	Arredondada para 3 casas decimais; usada no cálculo de proximidade e no grupo Location
Area ID	area	Agrupada na linha "Loc: Long Lat Area" do Sales Grid
Actual Age	actual_age	Agrupada na linha "Actual Age Effective Age" do Sales Grid
Effective Age	effective_age	Agrupada com a idade real no Sales Grid
Parcel Number	(nenhum)	Exibida na linha APN do Sales Grid
Listing ID	(nenhum)	Exibida na linha APN do Sales Grid
Address	display_only	Exibida na linha Address do Sales Grid; excluída do modelo

Os nomes de colunas da planilha podem estar em outro idioma — os nomes “especiais” estão em inglês para que o programa em R possa dar-lhes tratamento especial. No mais, os nomes de coluna informados aparecem nos modelos de regressão, nos gráficos e (no caso de avaliações) na Intermediate Sales Grid.

Nem todas as colunas são obrigatórias. O earthUI se adapta — se uma coluna estiver ausente, o recurso correspondente é simplesmente omitido. Por exemplo, se nenhuma coluna **concessions** for designada, a linha Net SP do Sales Grid mostra o Sale Price sem dedução de concessões. Contudo, para modelos de precificação imobiliária certas colunas são altamente recomendáveis para se obter um ajuste aceitável:

1. **Sale Age** — o número de dias entre a data de contrato da venda e a data de referência da avaliação ou análise. Se for usado um histórico de vendas de vários anos, especialmente para períodos superiores a 5 anos, **sale_age** costuma desempenhar papel central na estimativa do preço de venda. Na verdade, com frequência é tão importante que, sem ele, o earthUI não consegue fornecer modelo algum.
2. **Living Area** — também aparece com nomes como “Living Sqft”, “GLA” (gross living area) e assim por diante. Este é outro determinante de primeira ordem do preço de venda.
3. **Total Bath Count** — o número total de banheiros completos, de um quarto, de meio e de três quartos. Por exemplo, dois banheiros completos e um lavabo corresponderiam a um valor de 2,5.
4. **Garage Bays** ou **Garage Area** — o número de vagas de garagem ou a área da garagem.
5. **Lot Size** — a área de terreno do imóvel, tipicamente em pés quadrados ou acres.
6. **Longitude**, **Latitude** e, se disponível, **Area ID**. Os ajustes correspondentes serão combinados em um único ajuste de Location no Sales Grid.

2.3 Convenções de nomenclatura de colunas especiais

O earthUI identifica as colunas por sua **designação de tipo especial**, e não pelo nome da coluna. Você pode nomear suas colunas como quiser na exportação do MLS — o que importa é atribuir o tipo especial correto na tabela de Variable Configuration (Capítulo 6).

Por exemplo, seu MLS pode exportar a área construída como “GLA”, “Living SqFt”, “liv_area” ou “gross_living_area”. Depois da importação (quando o nome vira snake_case), basta designá-la como **living_area** no menu suspenso Special. O earthUI então a usará para cálculos de resíduo por pé quadrado e para o agrupamento no Sales Grid, independentemente de seu nome original.

2.4 Qualidade e completude dos dados

- **Valores ausentes (NA)**: linhas com valores NA em qualquer coluna preditora ou de resposta são automaticamente removidas antes do ajuste. O **na.action** é sempre definido internamente como **na.fail**, com a remoção de NA feita antes da chamada a **earth()**.
- **Colunas de data**: as datas devem estar em um formato que o R consiga interpretar (p. ex., 2025-06-15, 06/15/2025, June 15, 2025). O earthUI detecta automaticamente colunas de data quando pelo menos 80% dos valores são interpretados com sucesso. Anos com dois dígitos (p. ex., 06/15/25) são lidos como o ano correspondente mais próximo, de modo que exportações recentes do MLS são interpretadas corretamente; anos com quatro dígitos são sempre tomados literalmente.
- **Colunas numéricas**: preço de venda, área construída, tamanho do lote, concessões e campos semelhantes precisam ser numéricos. Se seu MLS exportar preços com símbolos monetários ou separadores de milhar (p. ex., “\$350,000” ou “350.000,00”), talvez seja necessário limpá-los antes da importação.
- **Colunas de fator**: variáveis categóricas como identificador de região, estilo ou estado de conservação devem conter um número gerenciável de valores distintos. O earthUI detecta

fatores automaticamente, mas você pode sobrepor a detecção.

Dica: Revise a coluna T/F/NA na tabela de preditores após a importação — ela mostra as contagens de true/false/NA de cada coluna. Colunas com mais de 30% de valores ausentes são sinalizadas em vermelho. Considere excluir colunas com muitos NA ou limpar os dados antes da importação.

2.5 Posicionamento da linha do avaliando

No modo **Appraisal**, a linha 1 deve ser o imóvel avaliando. Todas as linhas restantes são vendas comparáveis. O preço de venda do avaliando pode ser deixado em branco (NA) ou definido com qualquer valor — o earthUI o trata como NA durante o ajuste, de todo modo.

No modo **Market Area Analysis**, colocar um avaliando na linha 1 é opcional, e nenhuma linha é excluída automaticamente. Para manter uma linha de avaliando (ou qualquer outra linha) fora do ajuste, atribua a ela valor zero em uma coluna de pesos designada ou use um filtro de subconjunto.

No modo **General**, não há tratamento especial de linhas — todas as linhas são tratadas igualmente.

3 Modo de finalidade geral

3.1 Visão geral

O modo General Purpose é o padrão quando você inicia o earthUI. Ele oferece o fluxo de trabalho completo de modelagem MARS para qualquer conjunto de dados — não apenas imóveis. Você pode usar o earthUI para dados científicos, análise financeira, estudos de engenharia ou qualquer problema de regressão em que queira explorar relações não lineares e interações entre variáveis.

No modo General, a interface omite os recursos específicos de imóveis (colunas especiais, idade da venda, arredondamento de coordenadas, RCA). A barra lateral é simplificada para focar em seleção de variáveis, configuração de parâmetros, ajuste do modelo e exportação.

3.2 O fluxo de trabalho da barra lateral

A barra lateral é organizada em seções numeradas e recolhíveis que conduzem você da seleção do projeto até a exportação:

1. Project — Escolha um projeto existente ou crie um novo pelo assistente New Project, que registra a finalidade, o país e a localização administrativa (todos travados após a criação). O projeto ativo determina de onde os arquivos de entrada são lidos e para onde todas as saídas são gravadas — não existe um campo separado de pasta de saída.

2. Import Data — Escolha um arquivo CSV ou Excel na pasta `in/` do projeto ativo; um link Refresh reexamina a pasta em busca de arquivos adicionados manualmente. Para arquivos Excel com várias planilhas, aparece um seletor de planilha. Os nomes das colunas são automaticamente convertidos para `snake_case`.

3. Variable Configuration — Seletor da variável-resposta (admite múltiplas respostas), Effective Date e a tabela de preditores com caixas de seleção Include, Factor, Linear e Latent. Veja o Capítulo 6 para os detalhes completos.

4. Earth Call Parameters — Todos os argumentos da função `earth()`: degree, penalty, nk, método de poda, validação cruzada e outros. Veja o Capítulo 7 para a referência completa dos parâmetros.

Execute Prolog Processing — Uma seção opcional e sem numeração, entre as Seções 4 e 5, desabilitada a menos que o recurso experimental de Prolog esteja habilitado em Settings. Veja o Capítulo 15.

5. Fit Earth Model — Um campo de semente aleatória e o botão que executa o modelo de forma assíncrona. Veja o Capítulo 8 para os detalhes do ajuste.

6. Download Estimated Target Variable(s) & Residuals — (intitulada “Download Estimated Sale Prices & Residuals” nos modos de avaliação e de mercado) Exporta previsões, resíduos, escores CQA e contribuições por função g como um arquivo Excel. Veja o Capítulo 9.

7–8. RCA Adjustments e Sales Grid — Somente no modo de avaliação. Veja os Capítulos 10 e 11.

9. Generate Quarto Report — Grava um pacote de relatório Quarto autocontido (fonte .qmd mais recursos) na pasta de saída do projeto. Numerada como 7 fora do modo de avaliação. Veja o Capítulo 12.

10. Convert Quarto Report — Renderiza qualquer arquivo .qmd em HTML, Word ou PDF. Veja o Capítulo 12.

3.3 Abas do painel principal

Após o ajuste, o painel principal oferece dez abas:

Aba	Conteúdo
Data	Pré-visualização dos dados importados. No modo de avaliação, dividida nas tabelas Subject Property (imóvel avaliando) e Comparable Sales (vendas comparáveis).
Equation	A equação do modelo ajustado exibida em LaTeX/MathJax, mostrando cada função de base e seu coeficiente.
Summary	Métricas principais (R^2 , GRSq, GCV, RSS, CV R^2) e uma tabela de coeficientes.
Variable Importance	Gráfico de barras e tabela ordenada dos escores de importância dos preditores.
Contribution	Gráficos de contribuição das funções g. Selecione uma função g no menu suspenso; preditores univariados exibem gráficos de linha, e interações bivariadas ($\text{degree} \geq 2$) exibem gráficos de perspectiva 3D e de contorno.
Correlation	Mapa de calor das correlações entre preditores.
Diagnostics	Três gráficos: Resíduos vs. Ajustados, Q-Q normal e Observado vs. Previsto.
RCA Adjustments	(somente avaliação) Histogramas de Residual Adj %, Net Adj % e Gross Adj % entre os comparáveis, preenchidos após a conclusão do Passo 7 (Calculate RCA Adjustments).
Residual Contribution	Estima as contribuições de mercado de características raras ou excluídas, regredindo conjuntamente os resíduos do modelo sobre colunas que o modelo nunca utilizou (veja subsection 4.4).
ANOVA	Tabela de decomposição ANOVA mostrando a contribuição de cada função de base para a RSS.
Earth Output	Saída de texto bruta de 'earth::summary()', incluindo os detalhes da etapa de poda.

3.4 Persistência das configurações

O earthUI salva sua configuração no servidor, em um banco de dados de projeto (`projects.sqlite` na raiz `regProj`), indexada por projeto e finalidade. O salvamento é **acionado por botão**: o botão “Save current as default” da Seção 3 salva a resposta, a data de referência e a configuração dos preditores, enquanto o botão da Seção 4 salva os parâmetros da chamada ao earth e a matriz Allowed Interactions. Cada botão atualiza apenas a sua metade, de modo que nunca se sobrepõem, e **o ajuste não salva automaticamente** — você pode experimentar livremente sem sobrescrever seus padrões salvos. Os vários arquivos de entrada de um projeto (um pequeno extrato de teste, o conjunto de dados completo) compartilham uma configuração e, como as configurações vivem na árvore do projeto, elas viajam com ela quando você sincroniza, faz backup ou compartilha a pasta `regProj`.

3.5 Modo escuro

Clique no ícone de lua/sol no canto superior direito para alternar entre os temas claro e escuro. A preferência de tema é salva no armazenamento local e persiste entre sessões.

3.6 Localidade e configurações regionais

O earthUI oferece suporte a convenções internacionais de números, datas e formatação de CSV por meio de um sistema de localidades baseado em país. O menu suspenso **Country** no menu **Settings** (engrenagem) da barra de menus superior seleciona uma predefinição entre 31 países suportados. Cada predefinição configura:

- **Separador de CSV** — vírgula (,) para EUA/Reino Unido/Japão ou ponto e vírgula (;) para a maior parte da Europa, onde a vírgula é usada como marcador decimal.
- **Marcador decimal** — ponto (.) ou vírgula (,).
- **Separador de milhar** — vírgula (EUA/Reino Unido/Japão), ponto (Alemanha/Itália/Espanha), espaço (Finlândia/França/Polônia/Bálticos/Ucrânia/Rússia) ou apóstrofo (Suíça).
- **Formato de data** — MM/DD/AAAA (EUA), DD/MM/AAAA (maior parte da Europa) ou AAAA-MM-DD (Suécia/Lituânia/Japão/Canadá).
- **Tamanho do papel** — Letter (EUA/Canadá/México) ou A4 (todos os demais).

3.6.1 Países suportados

País	Sep. CSV	Decimal	Milhar	Data	Papel	
EUA, Canadá (csv)	,	.	,	MDY/YMD	Letter	
Reino Unido, Irlanda, AU, NZ	,	.	,	DMY	A4	
Alemanha, Áustria, Bélgica, Países Baixos, Itália, Espanha, Portugal, Dinamarca, Brasil	;	,	.	DMY	A4	
França, Polônia, Rep. Tcheca, Noruega, Letônia, Estônia	;	,	(espaço)	DMY	A4	
Finlândia, Suécia	;	,	(espaço)	DMY/YMD	A4	
Suíça	;	,	'	DMY	A4	
Ucrânia, Rússia	;	,	(espaço)	DMY	A4	
Turquia	;	,	.	DMY	A4	
Japão, Coreia do Sul	,	.	,	YMD	A4	
México	,	.	,	DMY	Letter	
Lituânia	;	,	(espaço)	YMD	A4	

3.6.2 Sobreposição do papel

Abaixo do seletor de país, um menu suspenso **Paper** (Letter ou A4) permite sobrepor o tamanho de página do relatório sem trocar de país. O separador de CSV, o marcador decimal e a ordem do formato de data seguem a predefinição do país selecionado; quando você muda de país, a sobreposição de papel volta ao padrão daquele país.

3.6.3 Salvando padrões

Clique em **Save** no menu Settings para armazenar globalmente suas preferências de localidade. Esses padrões se aplicam a todas as sessões futuras, independentemente do projeto que você abrir. As configurações por projeto (resposta, preditores, parâmetros) são salvas separadamente no banco de dados do projeto; os padrões de localidade persistem em todos os projetos por meio de um banco de dados SQLite no nível do usuário. Essa abordagem em dois níveis foi pensada para organizações como empresas de auditoria que trabalham com dados de vários países — defina o país mais frequente como padrão e ajuste quando necessário.

Dica: A formatação numérica nos eixos dos gráficos e nos rótulos de inclinação adapta-se automaticamente à sua localidade. A localidade alemã usa pontos para milhares (200.000), a finlandesa usa espaços (200 000) e a suíça usa apóstrofes (200'000). Nenhum símbolo monetário é exibido — o earthUI é agnóstico quanto à moeda.

4 Modo de avaliação

Quando você seleciona **For Appraisal** como Purpose, o earthUI se configura para a avaliação de um único imóvel. Todos os recursos descritos no Capítulo 3 continuam disponíveis; este capítulo trata apenas dos acréscimos específicos da avaliação.

4.1 Tratamento da linha do avaliando

No modo de avaliação, **a linha 1 do seu conjunto de dados é o imóvel avaliando** e todas as linhas restantes são vendas comparáveis. Seu arquivo de entrada deve estar organizado de acordo (veja o Capítulo 2). O preço de venda do avaliando pode ser deixado em branco ou definido com qualquer valor — o earthUI automaticamente o trata como NA durante o ajuste.

Após a importação, a aba Data se divide em duas seções: **Subject Property** (linha 1) e **Comparable Sales** (linhas 2 em diante). A linha 1 é sempre excluída do ajuste do modelo — a notificação “Skipping row 1 (subject). Fitting on N rows.” confirma isso. Após o ajuste, o modelo ainda gera previsões para a linha do avaliando, exibidas como **est_<target>** na saída.

4.2 Data de referência e idade da venda

Um campo **Effective Date** aparece na seção Variable Configuration em todos os modos de finalidade (com a data de hoje como padrão) — análises gerais e de mercado também têm uma data de referência, não apenas as avaliações — mas ele só é *usado* nos modos de avaliação e de mercado. Nesses modos, se você designar uma coluna como **contract_date** no menu suspenso Special, o earthUI calcula uma coluna **sale_age** — o número de dias inteiros entre a data de contrato de cada venda e a data de referência. Essa coluna substitui a coluna de data original como preditor.

Na primeira vez em que você clicar em Fit depois de designar uma data de contrato, o earthUI cria a coluna **sale_age** e o avisa para clicar em Fit novamente a fim de incluí-la.

4.3 Designações especiais de colunas

Nos modos de avaliação e de mercado, um menu suspenso **Special** aparece para cada preditor na tabela Variable Configuration. A lista completa de tipos especiais e seus efeitos:

Tipo especial	Efeito
no	Padrão — sem tratamento especial
contract_date	Dispara o cálculo automático de <code>sale_age</code> a partir da Effective Date
listing_date	Usada com a data do contrato para calcular o DOM quando não existe coluna <code>dom</code>
dom	Identifica a coluna de dias no mercado; exibida no Sales Grid
concessions	Identifica as concessões da venda; usada na fórmula Net SP (Sale Price – Concessions) do Sales Grid
latitude	Valores arredondados para 3 casas decimais; usados na proximidade de Haversine e no grupo Location do Sales Grid
longitude	Valores arredondados para 3 casas decimais; usados na proximidade de Haversine e no grupo Location do Sales Grid
area	Agrupada na linha "Loc: Long Lat Area" do Sales Grid
living_area	Habilita resíduos por pé quadrado (<code>residual_sf</code> , <code>cqa_sf</code>) na saída de download
lot_size	Agrupada na linha "Site Size Dimensions" do Sales Grid
site_dimensions	Agrupada com o tamanho do lote no Sales Grid
actual_age	Agrupada na linha "Actual Age Effective Age" do Sales Grid
effective_age	Agrupada com a idade real no Sales Grid
display_only	Coluna incluída nas exportações, mas excluída do ajuste do modelo (permite várias)
remarks	(Experimental) Coluna de texto livre de observações, convertida em colunas de características pela etapa opcional de Prolog (permite várias); veja o Capítulo 15
list	(Experimental) Coluna delimitada por vírgulas, dividida em colunas TRUE/FALSE one-hot pela etapa opcional de Prolog (permite várias); veja o Capítulo 15

Só é permitida uma coluna por tipo especial (exceto `display_only`, `remarks` e `list`). Atribuir um especial a uma segunda coluna automaticamente o remove da primeira. Um pequeno selo azul aparece ao lado do nome da variável, mostrando o tipo especial atribuído.

4.4 Contribuição residual (estimação de características raras)

Características raras demais para o canal da regressão — uma oficina presente em um punhado de vendas — pertencem à camada residual do RCA. A aba **Residual Contribution** estima sua contribuição de mercado regredindo os resíduos do modelo ajustado sobre colunas que o modelo nunca utilizou, por mínimos quadrados ordinários. Selecionar várias colunas as estima *conjuntamente*, de modo que amenidades correlacionadas (p. ex. `outbuilding_sqft` mais `pr_has_workshop`) dividem o crédito em vez de serem contadas em dobro. Indicadores binários produzem estimativas de valor global e listam o resíduo de cada venda sinalizada como evidência; colunas numéricas produzem contribuições por unidade; o erro padrão reportado é um EP ($t \geq 2$ como critério usual de significância). Termos apoiados em menos de dez observações recebem automaticamente uma RESSALVA na nota gerada para o dossiê, que documenta método, estimativas e advertências em um único parágrafo para o relatório. Somente os preditores que o modelo efetivamente *usou* são excluídos da lista de candidatos — um indicador raro oferecido ao ajuste, mas nunca selecionado pela passagem direta, permanece elegível, pois o modelo não absorveu nada de seu valor.

4.5 Visão geral dos ajustes RCA

O botão **Calculate RCA Adjustments & Download** (seção 7 da barra lateral, visível apenas no modo de avaliação após o ajuste) calcula ajustes derivados do mercado para cada comparável em relação ao avaliando. O fluxo de trabalho completo do RCA é descrito no Capítulo 10.

Depois de calcular os ajustes RCA, o botão **Generate Sales Grid & Download** (seção 8 da barra lateral) fica disponível. O fluxo de trabalho do Sales Grid é descrito no Capítulo 11.

Dica: O escore CQA que você atribui ao avaliando controla quanto da distribuição dos resíduos é atribuído ao avaliando. Um escore de 5,00 posiciona o avaliando na mediana dos comparáveis.

5 Modo de análise de região de mercado

Quando você seleciona **Market Area Analysis** como Purpose, o earthUI oferece os mesmos recursos específicos de imóveis do modo de avaliação (colunas especiais, idade da venda, arredondamento de coordenadas), mas orientado à análise de um grupo de imóveis, em vez da avaliação de um único avaliando.

5.1 Diferenças em relação ao modo de avaliação

- **Sem exclusão automática do avaliando** — diferentemente do modo de avaliação, o modo de mercado não exclui automaticamente a linha 1 do ajuste. Todas as linhas são incluídas; para manter uma linha de avaliando fora do ajuste, atribua a ela valor zero em uma coluna de pesos designada (como no conjunto de dados de demonstração) ou use um filtro de subconjunto.
- **Sem seções de RCA ou Sales Grid** — as etapas “Calculate RCA Adjustments & Download” e “Generate Sales Grid & Download” não estão disponíveis. O modo de mercado concentra-se no ajuste do modelo e na saída, e não em ajustes por comparável.
- **Numeração da barra lateral** — sem as etapas de RCA e Sales Grid, a seção Generate Quarto Report é numerada como 7 em vez de 9 (a seção Convert Quarto Report continua sendo a 10).

5.2 Quando usar o modo de mercado

O modo Market Area Analysis é apropriado quando você está:

- Construindo um modelo de regressão para um bairro ou região de mercado a fim de compreender os fatores determinantes de valor
- Analisando como variáveis como área construída, idade, tamanho do lote e localização afetam os preços de venda de um grupo de imóveis
- Preparando fundamentação para uma análise de condições de mercado ou uma delimitação de bairro
- Trabalhando com um conjunto de dados que inclui um imóvel avaliando na linha 1, mas querendo a opção de incluí-lo ou excluí-lo do ajuste

Dica: O modo de mercado também é útil para regressões imobiliárias em geral, quando você quer os recursos de colunas especiais (idade da venda, arredondamento de coordenadas), mas não precisa do fluxo de trabalho de ajustes RCA.

6 Seleção de variáveis

A Seção 3 da barra lateral — **Variable Configuration** — é onde você escolhe quais colunas participam do modelo e como elas são tratadas.

6.1 Variável(is)-resposta

O menu suspenso **Target (response) variable(s)**, no topo da Seção 3, lista todas as colunas do seu conjunto de dados. Selecione uma coluna para um modelo padrão de resposta única, ou várias colunas para um modelo de múltiplas respostas. Quando vários alvos são selecionados, o modelo ajusta todas as respostas simultaneamente usando `earth(cbind(y1, y2, ...) ~ .)`.

As colunas selecionadas como resposta são automaticamente excluídas da lista de preditores.

6.2 A tabela de preditores

Abaixo do seletor de resposta, uma tabela lista todas as colunas restantes com os seguintes campos:

Coluna	Descrição
Variable	Nome da coluna (nome completo exibido na dica de tela se truncado), precedido de um alternador $\times$ / $+$ (veja abaixo). Nos modos de avaliação/mercado, um selo azul mostra o tipo especial atribuído.
Type	Menu suspenso de tipo de dados: <code>numeric</code> , <code>integer</code> , <code>character</code> , <code>logical</code> , <code>factor</code> , <code>Date</code> , <code>POSIXct</code>
Include	Caixa de seleção — incluir esta coluna como preditor no modelo
Factor	Caixa de seleção — tratar esta coluna como variável categórica
Linear	Caixa de seleção — forçar entrada apenas linear (sem funções dobradiça)
Latent	Caixa de seleção — manter esta coluna inteiramente fora do modelo (excluída do <code>earth</code> e dos dados exportados para o <code>glmnetUI</code> / <code>mgcvUI</code>), reservando-a para trabalho com variáveis latentes
Special	Menu suspenso (somente avaliação/mercado) — veja a Referência de tipos especiais de coluna, abaixo
T/F/NA	Três contagens por coluna: para lógicas, TRUE/FALSE/NA; para numéricas, $\neq 0$ / $= 0$ /NA. Exibidas em vermelho quando mais de 30% dos valores estão ausentes

Uma linha de dica acima da tabela explica as abreviações: “Type = column data type, Inc = include as predictor, Factor = treat as categorical, Linear = linear-only (no hinges), Special = column role (e.g. `contract_date`).”

Desabilitação suave em vez de exclusão. O alternador $\times$ ao lado de cada nome de variável esmaece a linha e mantém a coluna fora do modelo sem removê-la dos dados; ele se transforma em um $+$ verde que reabilita a coluna. Colunas nunca são destruídas.

6.3 Detecção de tipos de dados e sobreposições

O earthUI detecta automaticamente os tipos de dados na importação. Colunas numéricas, inteiras, lógicas, de fator e de data são reconhecidas. Colunas de caracteres que parecem datas (pelo menos 80% dos valores interpretáveis segundo formatos de data comuns) são classificadas como **Date**.

Você pode sobrepor qualquer detecção alterando o menu suspenso **Type**. Ao mudar uma coluna para **character** ou **factor**, a caixa Factor é marcada automaticamente. Alterar tipos afeta como a coluna é passada à função **earth()**.

6.4 Marcadores Factor e Linear

- **Factor** — Quando marcado, a coluna é tratada como variável categórica. Isso é apropriado para colunas como **style**, **area_id** ou **grade**, que representam grupos discretos, e não medições contínuas. Colunas de fator entram no modelo como variáveis indicadoras (*dummy*).
- **Linear** — Quando marcado, a coluna é forçada a entrar no modelo linearmente — nenhuma função dobradiça (linear por partes) é criada para ela. Isso é útil quando você sabe que uma variável tem relação estritamente linear com a resposta.

6.5 Referência de tipos especiais de coluna

Nos modos de avaliação e de mercado, o menu suspenso **Special** oferece as seguintes opções. Cada tipo pode ser atribuído a no máximo uma coluna (exceto **display_only**, **remarks** e **list**, que permitem várias):

Tipos de data e hora:

- **contract_date** — Dispara o cálculo automático de **sale_age**. A coluna de data original é substituída por uma coluna inteira que mede os dias entre a data da venda e a Effective Date.
- **listing_date** — Usada como alternativa para calcular os dias no mercado (DOM = data do contrato – data do anúncio) quando nenhuma coluna **dom** explícita é designada.
- **dom** — Identifica a coluna de dias no mercado. Exibida na linha APN e na linha Date of Sale do Sales Grid.

Tipos monetários:

- **concessions** — Identifica concessões da venda (créditos do vendedor, incentivos ao comprador etc.). Usada no Sales Grid para calcular o preço de venda líquido: Net SP = Sale Price – Concessions.

Tipos de tamanho e localização:

- **latitude** — Os valores são automaticamente arredondados para 3 casas decimais, a fim de evitar sobreajuste. Usados nos cálculos de proximidade de Haversine (distância do avaliando a cada comparável) e agrupados na linha Location do Sales Grid.
- **longitude** — Mesmo tratamento de arredondamento da latitude. Usada para proximidade e para o grupo Location.
- **area** — Tipicamente um identificador de bairro ou região. Agrupada com latitude e longitude na linha “Loc: Long | Lat | Area” do Sales Grid.
- **living_area** — Habilita os cálculos de resíduo por pé quadrado (**residual_sf** e **cqa_sf**) na saída de download.
- **lot_size** — Agrupada na linha “Site Size | Dimensions” do Sales Grid.
- **site_dimensions** — Agrupada com o tamanho do lote no Sales Grid (p. ex., “75x120”).

Tipos de idade:

- **actual_age** — Agrupada na linha “Actual Age | Effective Age” do Sales Grid.
- **effective_age** — Agrupada com a idade real no Sales Grid.

Tipos de exibição:

- **display_only** — A coluna é incluída nas exportações para Excel, mas totalmente excluída do ajuste do modelo. Use isto para campos de endereço, números do MLS, identificadores de matrícula ou outros dados de referência que não devem ser preditores. Várias colunas podem receber essa designação.

Tipos de extração de texto (experimentais):

- **remarks** — Uma coluna de texto livre (p. ex. **public_remarks**) a ser convertida em colunas de características pela etapa opcional de Prolog. Não faz nada até que esse recurso seja habilitado em Settings. Várias colunas são permitidas. Veja o Capítulo 15.
- **list** — Uma coluna delimitada por vírgulas (p. ex. **existing_structures**) a ser dividida em colunas TRUE/FALSE <col>_<value> one-hot pela mesma etapa. Várias colunas são permitidas. Veja o Capítulo 15.

6.6 Múltiplas respostas

Selecionar mais de uma variável-resposta ajusta um modelo Earth de múltiplas respostas. Quando vários alvos são selecionados:

- O modelo prevê todas as respostas simultaneamente, compartilhando funções de base entre os alvos
- O botão **wp (response weights)** torna-se ativo, permitindo que você atribua um peso numérico a cada alvo
- Os modelos de variância (**varmod.method**) são desabilitados (sem suporte para múltiplas respostas)
- As abas de resultados exibem métricas, equações e gráficos de diagnóstico por resposta

Dica: Colunas designadas como **display_only** permanecem no seu conjunto de dados e aparecem na exportação para Excel, mas são inteiramente excluídas da tabela de preditores. Use isto para colunas de identificação, endereços ou outros campos de referência.

7 Seleção de parâmetros

A Seção 4 da barra lateral — **Earth Call Parameters** — dá acesso a todos os argumentos aceitos pela função `earth()`. Cada parâmetro tem um ícone azul de ajuda (?) com uma explicação em dica de tela. Os parâmetros são organizados em subseções recolhíveis.

7.1 Parâmetros básicos

Parâmetro	Padrão	Descrição
subset	(vazio)	Expressão de filtro de linhas. Veja “Filtragem por subconjunto”, abaixo.
weights	NULL	Seletor de coluna para pesos de caso (linha). Somente colunas numéricas são listadas.
wp	NULL	Pesos das respostas em modelos de múltiplos alvos. O botão abre uma caixa de diálogo com uma entrada numérica por alvo (padrão 1.0 cada).
keepxy	desligado	Reter x, y, subset e weights no objeto do modelo.
trace	0	Nível de rastreamento da saída de ajuste (0 a 5).
glm	nenhum	Família GLM opcional: gaussian , binomial ou poisson .
degree	1	Ordem máxima de interação. Definir $\text{degree} \geq 2$ habilita automaticamente a validação cruzada e revela a matriz Allowed Interactions.
penalty	2	Penalidade GCV por nó. Valores mais altos produzem modelos mais simples.

7.2 Passagem direta (forward pass)

Parâmetro	Padrão	Descrição
nk	automático	Máximo de termos antes da poda. Preenchido automaticamente com o valor recomendado para os dados carregados (veja “Valores recomendados de parâmetros”, abaixo).
thresh	0.001	Limiar da passagem direta. Valores menores permitem mais termos.

Parâmetro	Padrão	Descrição
minspan	automático	Vão mínimo entre nós. Preenchido automaticamente com o valor recomendado. Valores negativos definem o número máximo de nós por preditor.
endspan	automático	Vão de extremidade — distância mínima de um nó à borda dos dados. Preenchido automaticamente com o valor recomendado.
newvar.penalty	0	Penalidade por introduzir uma nova variável (incentiva o reaproveitamento de preditores já existentes).
fast.k	20	Número de termos-pai considerados no algoritmo fast MARS.
fast.beta	1	Controla o fator de envelhecimento do fast MARS.

7.3 Interações permitidas

Quando **degree** é definido como 2 ou mais, uma matriz de interações aparece abaixo dos parâmetros básicos. Trata-se de uma grade triangular superior de caixas de seleção, uma para cada par de preditores. Uma caixa marcada significa que os dois preditores podem interagir; desmarcá-la proíbe aquela interação específica.

Clicar no nome de um preditor (rótulo de linha ou coluna) alterna todas as interações daquela variável. As caixas **Allow All** e **Clear All** no topo fornecem controle em massa. A matriz usa cabeçalhos fixos, de modo que os rótulos de linha e coluna permanecem visíveis durante a rolagem.

Um alerta informativo lembra: “Interaction terms increase the risk of overfitting. Cross-validation has been enabled (10-fold).” (Os termos de interação aumentam o risco de sobreajuste. A validação cruzada foi habilitada, com 10 dobras.)

7.4 Poda

Parâmetro	Padrão	Descrição
pmethod	backward	Método de poda: backward , none , exhaustive , forward , seqrep ou cv .
nprune	NULL	Máximo de termos após a poda. Deixe vazio para seleção automática.

7.5 Validação cruzada

Parâmetro	Padrão	Descrição
nfold	10	Número de dobras da validação cruzada. Defina 0 para desabilitar a VC.
ncross	20	Número de repetições da validação cruzada.
stratify	ligado	Estratifica as amostras da VC para que cada dobra tenha distribuição de resposta semelhante.

Quando $\text{degree} \geq 2$, a validação cruzada é habilitada automaticamente (nfold definido como 10) para ajudar a proteger contra o sobreajuste decorrente das interações.

7.6 Filtragem por subconjunto

A entrada de texto **subset** aceita uma expressão R que filtra quais linhas são usadas no ajuste do modelo. Você pode digitar uma expressão diretamente (p. ex., `sale_age < 365 & area_id == 460`) ou usar o botão **Build filter...** para construí-la visualmente.

Caixa de diálogo Build Filter — Clique em “Build filter...” para abrir uma caixa de diálogo guiada. Cada linha de condição tem um menu suspenso de coluna, um operador (<, >, <=, >=, ==, !=) e uma entrada de valor que se adapta ao tipo da coluna: entrada numérica para números, seletor de data para datas e menu suspenso de valores distintos para colunas de caractere/fator. As condições são unidas por conectores AND (&) ou OR (|). Uma pré-visualização na parte inferior mostra a expressão e quantas linhas correspondem. Clique em **Apply** para inserir a expressão na entrada de texto.

Colunas de data precisam usar os invólucros `as.Date("...")` ou `as.POSIXct("...")` em expressões manuais. A caixa de diálogo Build Filter cuida disso automaticamente.

A filtragem por subconjunto não é destrutiva — as linhas excluídas permanecem no conjunto de dados e recebem previsões na exportação para Excel.

***Dica:** Use parênteses para agrupar condições OR ao combiná-las com AND. Sem parênteses, $A \& B \mid C$ é avaliado como $(A \& B) \mid C$, o que pode não ser o pretendido.*

7.7 Valores recomendados de parâmetros

O earthUI exibe um valor recomendado abaixo dos parâmetros principais da Seção 4. Essas recomendações são atualizadas reativamente com base no número de linhas usadas no ajuste (n) e de preditores selecionados (p). As fórmulas derivam do artigo de Friedman sobre MARS, dos algoritmos internos do earth e de testes empíricos. Os campos **nk**, **minspan** e **endspan** são pré-preenchidos com seus valores recomendados à medida que os dados são carregados; uma edição manual é preservada.

7.7.1 Parâmetros da passagem direta

nk (máximo de termos antes da poda):

$$\text{nk} = \min(100, \max(21, 2p + 1, \lfloor n/10 \rfloor))$$

n	$p=5$	$p=10$
30	21	21
50	21	21
100	21	21
200	21	21
500	50	50
1000	100	100
1500	100	100

O padrão do earth, $\min(200, \max(20, 2p)) + 1$, não leva em conta o tamanho do conjunto de dados e pode ser pequeno demais para conjuntos grandes, restringindo a passagem direta antes que ela explore adequadamente todos os preditores. O termo $\lfloor n/10 \rfloor$ permite aproximadamente um termo a cada 10 observações — uma regra prática padrão para evitar superparametrização. O teto de 100 evita retornos decrescentes.

minspan (mínimo de observações entre nós):

$$\text{minspan} = \min(16, \lfloor 5 + n/50 \rfloor)$$

endspan (mínimo de observações a partir das bordas dos dados):

$$\text{endspan} = \min(16, \lfloor 5 + n/28 \rfloor)$$

n	minspan	endspan
30	5	6
50	6	6
100	7	8
200	9	12
300	11	15
500	15	16
1500	16	16

O minspan autocalculado pelo earth usa a equação 43 de Friedman: $\lfloor (-\ln(-\ln 0.95) + \ln(p \cdot n)) / (2.5 \ln 2) \rfloor$, que cresce como $\ln(np)$ e aumenta lentamente demais para conjuntos de dados grandes. Em $n=1500$ ela produz apenas ≈ 7 , o que dá cerca de 245 posições candidatas a nó por preditor contínuo — muitas, permitindo que a passagem direta ajuste ruído. A fórmula recomendada visa aproximadamente 100 candidatos por preditor para n grande, deixando a cargo do autocálculo do earth os casos de n pequeno (nos quais ele é bem testado). O endspan é definido um pouco maior que o minspan para oferecer proteção adicional nas bordas, o que ajuda quando os dados têm caudas ralas (comum em imóveis).

penalty (penalidade GCV por nó):

$$\text{penalty} = \begin{cases} 3 & \text{se degree} > 1 \\ 2 & \text{se degree} = 1 \end{cases}$$

Este é o próprio padrão do earth — 2 para modelos aditivos, 3 para modelos com interações (a penalidade mais alta compensa o espaço de busca maior).

newvar.penalty (penalidade por introduzir um novo preditor):

Recomendado: **0,1** quando os preditores são correlacionados (p. ex., área construída com número de dormitórios, número de banheiros, tamanho do lote). Isso envia a passagem direta

em direção ao reaproveitamento de preditores já presentes no modelo, em vez de introduzir alternativas correlacionadas. O mecanismo: a melhora de RSS de cada novo preditor é multiplicada por $1/(1 + \text{penalty})$. Com `newvar.penalty = 0.1`, uma variável nova precisa melhorar a RSS em pelo menos 10% mais do que outro nó em uma variável já existente. Isso produz modelos mais simples, com menos preditores, sem afetar as estimativas finais dos coeficientes (a penalidade é removida após a seleção). Deixe em 0 se os preditores não forem correlacionados. Note que o `earth` não admite um `newvar.penalty` diferente de zero junto com pesos de caso — o `earthUI` desabilita a entrada (forçando 0) sempre que uma coluna de pesos é designada.

7.7.2 Parâmetros de poda

pmethod: Recomendado: **backward**. A poda regressiva baseada em GCV, que é o padrão, é determinística — os mesmos dados sempre produzem o mesmo modelo, sem dependência de sementes aleatórias. Isso é importante para a reprodutibilidade, especialmente em trabalhos de avaliação. O método `cv` usa validação cruzada para a poda, o que introduz dependência da semente.

nprune: Recomendado: **deixar vazio** (NULL). Deixe o GCV selecionar o número ótimo de termos. Definir `nprune` impõe um teto rígido que se sobrepõe ao critério do GCV.

7.7.3 Parâmetros de validação cruzada

nfold (dobras da VC): recomendado por faixa de tamanho amostral:

n (linhas do ajuste)	nfold
0–299	5
300–399	6
400–499	7
500–599	8
600–699	9
700+	10

Mais dobras treinam cada dobra com mais dados (estimativa menos enviesada), mas encolhem a dobra de teste retida. A tabela evolui até as 10 dobras padrão à medida que a amostra cresce, mantendo cada dobra de teste grande o bastante (cerca de 50+ linhas) para uma métrica por dobra estável, e recua para 5 dobras em amostras pequenas — nas quais a estimativa de CVR^2 ainda é ruidosa, de modo que convém aumentar `ncross` (p. ex. 15–20) para diluir a variância. O teto de 10 reflete o fato de que cada dobra adicional é um ajuste completo do `earth` com ganho diagnóstico desprezível. Com `pmethod = "backward"`, a validação cruzada não afeta o modelo — ela apenas calcula o CVR^2 diagnóstico e fornece resíduos para o modelo de variância.

ncross (repetições da VC): recomendado por faixa de tamanho amostral:

n (linhas do ajuste)	ncross
0–199	20
200–399	16
400–699	12
700–1199	10
1200–2999	7
3000+	5

As repetições diluem a sorte do particionamento em dobras e estabilizam o CVR^2 ; amostras maiores precisam de menos repetições porque uma única validação cruzada já é mais estável. Os

resíduos da VC repetida também alimentam o modelo de variância.

7.7.4 Modelo de variância

varmod.method: Recomendado: **lm** — o padrão de Milborrow: um modelo linear robusto de variância para intervalos de previsão (a dispersão é estimada a partir de resíduos ruidosos, de modo que um modelo simples generaliza melhor). Se **lm** não convergir com seus dados, experimente **rlm** ou **const** (homocedástico, sempre converge), ou recorra a **earth** (que também lida com dispersão não linear). Ele exige **nfold** > 1 e **ncross** > 1.

7.7.5 Arquivo de saída do Earth

Após cada ajuste bem-sucedido, o earthUI grava <filename>_earth_output_<timestamp>.txt na pasta de saída. Esse arquivo contém os termos do modelo, estatísticas de resumo (R^2 , GRSq, CVRSq), detalhes do modelo de variância e o log de rastreamento. Um arquivo é criado por ajuste, fornecendo um registro cumulativo para comparar configurações de parâmetros.

7.8 Padrões das configurações

Um botão de opção no topo da Seção 4 controla quais padrões são carregados:

- **Use last settings for input file** — restaura as configurações que você usou da última vez com esse arquivo específico
- **Use default settings** — aplica seus padrões personalizados salvos
- **Earth defaults** — redefine todos os parâmetros para os valores de fábrica

Clique em **Save current as default** na parte inferior da Seção 4 para armazenar os parâmetros atuais do earth e a matriz Allowed Interactions; o botão correspondente na Seção 3 armazena a resposta, a data de referência e a configuração dos preditores. Ambos são salvos por projeto e finalidade no banco de dados do projeto.

8 Ajustando o modelo Earth

8.1 O botão Fit

A Seção 5 da barra lateral contém o botão **Fit Earth Model** junto com um campo **Random seed**. A semente torna as atribuições de dobras da validação cruzada reprodutíveis — um menu suspenso recupera as últimas 10 sementes usadas, e a semente utilizada é registrada na aba Earth Output após o ajuste. Antes do ajuste, o earthUI valida sua configuração — é preciso selecionar uma resposta e incluir ao menos um preditor. Nos modos de avaliação/mercado, as colunas de latitude e longitude são arredondadas, e a idade da venda é calculada a partir da data de referência e da data de contrato (se designada).

8.2 Janela modal de ajuste e saída de rastreamento

Quando você clica em Fit, aparece uma sobreposição modal escura com:

- **Cabeçalho** — “Fitting Earth Model”, com um cronômetro de tempo decorrido contando os segundos e um botão **Abort** que cancela o ajuste em segundo plano
- **Log de rastreamento** — uma área de texto rolável, em fonte monoespaçada, exibindo em tempo real a saída da função `earth()`: informações do conjunto de dados, progresso da passagem direta, dobras da validação cruzada e tempo de conclusão
- **Código de cores** — linhas padrão em verde, linhas de validação cruzada em amarelo, erros em vermelho

O earthUI ajusta modelos de forma assíncrona usando `callr::r_bg()`, que executa o cálculo em um processo R separado. Isso mantém o aplicativo responsivo durante ajustes demorados. Um observador com sondagem de 300 ms lê a saída do processo em segundo plano e a transmite para o log de rastreamento.

Quando o ajuste termina, aparece a mensagem “Done in X.Xs”, um botão de fechar (X) é adicionado à janela modal e o fundo escurecido é removido, de modo que as abas de resultado possam ser renderizadas sob demanda. A janela modal então se fecha automaticamente assim que as saídas visíveis terminam de ser renderizadas — ou você pode fechá-la antes.

Em caso de sucesso, uma marca de verificação verde é acrescentada ao botão Fit. Se ocorrer um erro, a mensagem de erro é exibida no log de rastreamento e aparece uma notificação.

8.3 Log de ajuste

Todo ajuste (bem-sucedido ou não) grava um arquivo de log na sua pasta de saída:

- **Nome do arquivo:** `<datafile>_earth_log_<YYYYMMDD_HHMMSS>.txt`
- **Conteúdo:** carimbo de data/hora, toda a saída de rastreamento do processo de ajuste e quaisquer mensagens de erro
- **Local:** a pasta de saída do projeto ativo

***Dica:** Se o `callr` não estiver instalado, o earthUI recorre ao ajuste síncrono com uma barra de progresso simples. Instale o `callr` para ter a experiência completa de saída de rastreamento.*

9 Baixando os dados

Após o ajuste, a seção 6 da barra lateral — intitulada “Download Estimated Target Variable(s) & Residuals” no modo geral e “Download Estimated Sale Prices & Residuals” nos modos de avaliação e de mercado — baixa um arquivo Excel com previsões e diagnósticos. Essa saída é usada no Passo 7 (RCA) para atribuir uma classificação CQA (Condition-Quality-Appeal — estado, qualidade e atratividade) ao imóvel avaliando. A saída é ordenada por **residual_sf** e **cqa_sf** para ajudá-lo a avaliar em que ponto do ordenamento o avaliando se situa. Se o modelo for de boa qualidade, os imóveis devem ficar ordenados do menos atraente ao mais atraente, com base em características residuais que não entraram na regressão. O valor central deve ser aproximadamente 0, a metade inferior com valores negativos e a metade superior com valores positivos. Você deve encontrar as casas de pior qualidade, ou “para reformar”, perto do fim do ordenamento, e as melhores casas no topo. Costuma haver exceções decorrentes de anomalias como execuções hipotecárias, vendas por valor inferior à dívida, vendas de inventário (ligadas a herança) e vendas rápidas motivadas por mudança de emprego ou outras razões. A investigação das anomalias em geral revela uma razão pertinente para a anomalia de preço.

O rótulo do botão se adapta ao modo de finalidade:

- Geral/Mercado: **Download Output (Excel)**
- Avaliação: **Download Intermediate Output (Excel)**

O formato do nome de arquivo é `<datafile>_modified_<YYYYMMDD_HHMMSS>.xlsx`.

9.1 Colunas de saída

Para cada variável-resposta, são acrescentadas as seguintes colunas:

Coluna	Descrição
<code>est_<target></code>	Previsão do modelo, p. ex. <code>est_sale_price</code> (1 casa decimal)
<code>residual</code>	Observado menos previsto (1 casa decimal)
<code>cqa</code>	Escore Condition-Quality-Appeal (2 casas decimais, faixa 0–10)
<code>residual_sf</code>	Resíduo dividido pela área construída (se designada, 1 casa decimal)
<code>cqa_sf</code>	CQA calculado a partir do ordenamento por <code>residual_sf</code> (2 casas decimais)
<code><variable>_contribution</code>	Valor de contribuição por função g (1 casa decimal, uma coluna por função g)
<code>basis</code>	Contribuição do valor do intercepto, igual para todos os imóveis (1 casa decimal)
<code>calc_residual</code>	Observado menos (basis + todas as contribuições) — coluna de verificação (1 casa decimal)

Em modelos de múltiplas respostas, um sufixo `<i>` é acrescentado para distinguir as colunas de cada alvo.

9.2 Ordenação das colunas e formatação do Excel

As colunas de ordenamento são colocadas na **posição mais à esquerda** do arquivo de saída, nesta ordem: **residual_sf**, **cqa_sf**, **residual**, **cqa**. Isso facilita percorrer a lista ordenada e avaliar onde o avaliando se situa na distribuição de CQA.

Essas colunas recebem formatação numérica do Excel:

Coluna	Formato	Exemplo
<code>residual_sf</code>	Numérico, 2 casas decimais	12.34
<code>cqa_sf</code>	Numérico, 2 casas decimais	7.25
<code>residual</code>	Numérico, 0 casas decimais	15,200
<code>cqa</code>	Numérico, 2 casas decimais	6.80

9.3 Escores CQA

O escore CQA (Condition-Quality-Appeal) ordena o resíduo de cada linha em relação a todos os demais resíduos. Para uma dada linha, o CQA é a porcentagem de linhas com resíduo com sinal menor, multiplicada por 10. Isso produz uma escala de 0 a 10 em que:

- **CQA alto** ($\approx 9-10$) — o imóvel foi vendido por muito mais do que o modelo previa (resíduo positivo grande)
- **CQA baixo** ($\approx 0-1$) — o imóvel foi vendido por muito menos do que o previsto (resíduo negativo grande)
- **CQA ≈ 5** — o resíduo está próximo da mediana de todos os resíduos

Quando uma coluna `living_area` é designada, `cqa_sf` fornece o mesmo ordenamento com base nos resíduos por pé quadrado.

9.4 Ordenação e linha do avaliando

Nos modos de avaliação e de mercado, as linhas de comparáveis são ordenadas por `residual_sf` de forma decrescente (ou por `residual`, se nenhuma área construída for designada). A linha do avaliando (linha 1) permanece na posição 1 e não é ordenada. No modo geral, as linhas são exportadas em sua ordem original.

Essa ordenação, combinada com as colunas de ordenamento mais à esquerda, permite ao avaliador percorrer rapidamente os comparáveis, do mais superestimado ao mais subestimado, e avaliar onde o escore CQA atribuído ao avaliando se situa nessa distribuição.

Os níveis de fator nos dados de previsão são alinhados com os dados de treinamento antes da chamada a `predict()`. Linhas com níveis de fator não vistos produzirão previsões NA.

***Dica:** A coluna `calc_residual` deve sempre ser igual a `residual` (a menos de arredondamento). Se houver diferença significativa, isso indica um problema de cálculo que merece investigação.*

9.5 Exportação automática para o mgcvUI

A cada ajuste bem-sucedido com `degree` ≤ 2 , o earthUI salva automaticamente o objeto de resultado completo como um arquivo `.rds` na pasta de saída do projeto. O nome do arquivo segue o padrão `<datafile>_earthUI_result_<YYYYMMDD_HHMMSS>.rds`.

Esse arquivo pode ser carregado pelo Post Processing Module de mgcv — incluído no pacote earthUI e iniciado com `earthUI::launch_mgcv()` — usando `readRDS()`. A rotina GAM usa as posições dos nós e as funções de base do modelo earth como pontos de partida para os termos suavizados do GAM, permitindo uma transição fluida de MARS para GAM.

Modelos com `degree` > 2 são ignorados porque o mgcvUI só admite interações de duas variáveis. Quando o earthUI é iniciado no modo Trilogy (a partir da interface Trilogy), uma seção adicional **Lock Model Output** permite travar o `.rds` de um determinado ajuste — junto com sua base de preditores e as configurações de CQA compartilhadas — como o modelo earth canônico que o glmnetUI e o mgcvUI importam.

9.6 Importação no glmnetUI

O mesmo arquivo `.rds` salvo para a rotina GAM também pode ser importado no **Post Processing Module de elastic net** — igualmente incluído no pacote `earthUI` e iniciado com `earthUI::launch_glmnet()`. No aplicativo `glmnet`, navegue até **Section 2: Import from earthUI** e use o botão **Browse** para selecionar o arquivo `.rds`.

O `glmnetUI` usa a abordagem padrão `earth::model.matrix()`: as funções de base do modelo `earth` (dobradiças, interações e termos lineares) tornam-se as colunas da matriz de delineamento do `glmnet`. O `glmnet` então realiza regressão regularizada sobre essas colunas de base, selecionando-as e encolhendo-as via `lasso/elastic net`. Isso combina a construção adaptativa de base do `earth` com a regularização do `glmnet`.

9.6.1 Principais benefícios

- **Interações interpretáveis:** as interações do `earth` baseadas em dobradiças (inclusive termos de três vias) têm significado geométrico claro, ao contrário das interações nativas por produto cruzado do `glmnet`, difíceis de explicar em juízo ou em contextos de auditoria.
- **Não linearidade automática:** as funções dobradiça do `earth` capturam relações não lineares que um modelo `glmnet` padrão com preditores brutos deixaria escapar.
- **Seleção de variáveis:** o `glmnet` pode podar ainda mais a base do `earth`, descartando termos dobradiça que não contribuem.

9.6.2 Importante: a coluna de pesos precisa coincidir

Aviso

Se o `earthUI` usar uma coluna de pesos (p. ex., para excluir a linha do avaliando com `weight=0`), a mesma coluna de pesos precisa ser designada na configuração de variáveis do `glmnetUI`. Se as configurações de peso diferirem, as contagens de linhas não coincidirão e a base do `earth` não poderá ser anexada à matriz do modelo do `glmnet`.

9.6.3 Fluxo de trabalho

1. Ajuste um modelo `earth` no `earthUI` (qualquer degree).
2. O arquivo de resultado `.rds` é salvo automaticamente na pasta de saída do projeto.
3. No `glmnetUI`, abra **Section 2: Import from earthUI** e navegue até o arquivo `.rds`.
4. Verifique se o **mesmo arquivo de dados** e a **mesma coluna de pesos** são usados nos dois aplicativos.
5. Configure os parâmetros do `glmnet` na Seção 5 (a `Interaction Matrix` é desabilitada automaticamente quando há uma importação do `earthUI` ativa).
6. Clique em **Fit Model**. As colunas de base do `earth` aparecem nas abas `Equation`, `Coefficients` e `Contributions`.

10 Cálculos RCA e download

O botão **Calculate RCA Adjustments & Download** aparece na seção 7 da barra lateral, visível apenas no modo de avaliação depois que um modelo tenha sido ajustado. O RCA (Residual Constraint Approach) produz ajustes por comparável derivados do mercado, em relação ao imóvel avaliando.

10.1 Abrindo a caixa de diálogo RCA

Clicar no botão abre uma pequena caixa de diálogo modal com:

- **Score type** — botões de opção para escolher **CQA** ou **CQA per SF** (a opção por pé quadrado só está disponível quando uma coluna `living_area` está designada). Se você escolher “CQA per SF”, então, com base no escore CQA que atribuir ao avaliando, seu escore residual será sua área construída multiplicada pelo `residual_sf` correspondente ao escore CQA_SF informado.
- **CQA value** — entrada para o escore CQA do avaliando (0,00–10,00). Esse escore representa onde você acredita que o avaliando se situa na distribuição de qualidade dos comparáveis.
- Botão **Generate** — calcula a saída do RCA e a baixa como Excel. O botão permanece desabilitado até que um escore válido seja informado.

10.2 Interpolação do escore CQA

Quando você clica em Generate, o earthUI interpola o resíduo do avaliando a partir dos pares CQA/resíduo dos comparáveis:

1. Os escores CQA e os resíduos dos comparáveis são ordenados por CQA crescente
2. A interpolação linear (`stats::approx()`) mapeia o valor de CQA que você informou para um resíduo
3. Se estiver usando CQA por pé quadrado, o resíduo por pé quadrado é convertido de volta em resíduo total pela multiplicação pela área construída do avaliando
4. O valor estimado do avaliando é calculado como: previsão do modelo + resíduo interpolado

10.3 Colunas de saída

O arquivo Excel do RCA (`<datafile>_adjusted_<YYYYMMDD_HHMMSS>.xlsx`) inclui todas as colunas da saída intermediária mais:

Coluna	Descrição
<code>subject_value</code>	Previsão do modelo + resíduo interpolado (somente linha 1)
<code>subject_cqa</code>	O escore CQA que você informou (somente linha 1)
<code><variable>_adjustment</code>	Contribuição do avaliando menos contribuição do comparável (por função g)
<code>residual_adjustment</code>	Resíduo do avaliando menos resíduo do comparável
<code>net_adjustments</code>	Soma de todos os ajustes (contribuição + resíduo)
<code>gross_adjustments</code>	Soma dos valores absolutos de todos os ajustes
<code>residual_pct</code>	Ajuste residual como fração do preço de venda do comparável
<code>net_adj_pct</code>	Ajustes líquidos como fração do preço de venda do comparável

Coluna	Descrição
<code>gross_adj_pct</code>	Ajustes brutos como fração do preço de venda do comparável
<code>adjusted_sale_price</code>	Preço de venda do comparável + ajustes líquidos

As colunas de ajuste informam, para cada comparável, quanto da diferença o modelo atribui a cada variável. `adjusted_sale_price` é o preço de venda do comparável depois de aplicados todos os ajustes derivados do modelo — um conjunto de preços de venda ajustados que deve se agrupar em torno do valor estimado do avaliando.

10.4 A aba RCA Adjustments

Uma vez calculada a saída do RCA, a aba **RCA Adjustments** do painel principal exibe histogramas das distribuições de Residual Adj %, Net Adj % e Gross Adj % entre os comparáveis (um conjunto por alvo, em modelos de múltiplas respostas), oferecendo uma verificação visual rápida de quão concentrados estão os ajustes.

***Dica:** Um escore CQA de 5,00 posiciona o avaliando na mediana dos comparáveis. Escores acima de 5 indicam que o avaliando tem qualidade acima da média; abaixo de 5, qualidade abaixo da média.*

11 Grade de comparação de vendas

11.1 Visão geral e finalidade

A **Sales Comparison Grid** (grade de comparação de vendas) é uma pasta de trabalho do Excel gerada no modo de avaliação (seção 8 da barra lateral) depois do cálculo dos ajustes RCA. Ela apresenta o imóvel avaliando ao lado dos comparáveis selecionados em um formato de grade estruturado, com fórmulas do Excel que calculam automaticamente os ajustes e um preço de venda ajustado para cada comparável.

A grade foi concebida para o dossiê do avaliador — ela combina os ajustes derivados da regressão do modelo Earth com células editáveis nas quais o avaliador pode alocar o resíduo CQA a características específicas do imóvel. O nome do arquivo de saída é `SalesGrid_<YYYYMMDD_HHMMSS>.xlsx`.

11.2 Janela modal de seleção de comparáveis

Clicar em “Generate Sales Grid & Download” abre uma caixa de diálogo modal na qual você seleciona quais comparáveis incluir:

- Os **comparáveis recomendados** já vêm marcados. São comparáveis com percentual de ajuste bruto inferior a 25% do preço de venda, ordenados por idade da venda crescente (vendas mais recentes primeiro) e limitados a 30.
- Os **comparáveis adicionais** são listados abaixo, desmarcados por padrão. Eles têm ajustes brutos de 25% ou mais, mas continuam disponíveis para seleção (até 50 exibidos, ordenados por percentual de ajuste bruto crescente).
- Podem ser selecionados no máximo **30 comparáveis**, produzindo até **10 planilhas** (3 comparáveis por planilha).

Depois de confirmar sua seleção, os comparáveis escolhidos são ordenados por percentual de ajuste bruto crescente, de modo que os comparáveis mais semelhantes apareçam na Planilha 1.

11.3 Leiaute da grade

Cada planilha tem um leiaute de 20 colunas, acomodando o imóvel avaliando mais 3 comparáveis. As colunas de cada entidade são:

Entidade	Colunas (5 por entidade)
Avaliando	Rótulo, Valor factual 1, Valor factual 2, Valor factual 3, Contribuição de valor (VC)
Cada comparável	Valor factual 1, Valor factual 2, Valor factual 3, Contribuição de valor (VC), Ajuste

As linhas, de cima para baixo, são:

1. **Título** — nome da planilha (p. ex., “Intermediate Sales Comparable Grid — Sheet 1 of 3”)
2. **Cabeçalhos** — “Subject” e os cabeçalhos das colunas dos comparáveis, com endereço
3. **Address** — endereço completo do imóvel
4. **APN | MLS# | DOM | Subj.Prox** — número de matrícula, identificador do anúncio, dias no mercado e distância de Haversine (milhas) até o avaliando
5. **Sales Price | Concess. | Net SP** — preço de venda, concessões e fórmula do preço de venda líquido
6. Linha de cabeçalho **Regression Features**
7. **BASE VALUE** — o intercepto do modelo

8. **Date of Sale | OffMkt | OnMkt** — data do contrato, idade da venda e DOM
9. **Linhas agrupadas** (condicionais) — grupos Location, Site Size e/ou Age
10. **Linhas das variáveis do modelo** — uma linha por preditor (excluindo as variáveis agrupadas)
11. **Linha separadora em branco**
12. Linha de cabeçalho **Residual Features**
13. **CQA|Residual** — escore CQA + fórmula do resíduo remanescente
14. **Linhas de características residuais** — linhas nomeadas + linhas em branco para preenchimento pelo avaliador
15. **Total VC / Net Adjustment**
16. **Net Adjustment %**
17. **Gross Adjustment %**
18. **Adjusted Sale Price** — linha de fórmula
19. Rodapé de **copyright**

11.4 Preço de venda, concessões e Net SP

A linha 5 mostra três valores para cada comparável:

- **Sale Price** — o preço de venda do comparável, conforme os dados
- **Concessions** — o valor da coluna designada como **concessions** (0 se não designada)
- **Net SP** — uma fórmula do Excel: Sale Price – Concessions

A coluna do avaliando mostra “N/A” para o preço de venda (uma vez que ele é desconhecido) e qualquer valor de concessões, se disponível.

11.5 Linhas agrupadas (Location, Site, Age)

Quando certos tipos especiais são designados e essas variáveis aparecem no modelo, o earthUI cria linhas agrupadas que combinam variáveis relacionadas:

Loc: Long | Lat | Area — Aparece quando qualquer uma de **longitude**, **latitude** ou **area** está no modelo. Mostra os valores factuais de cada variável constituinte, uma contribuição de valor combinada (soma das VCs individuais) e, para os comparáveis, um ajuste (VC combinada do avaliando – VC combinada do comparável). Estilizada com fundo azul-claro nas células de VC.

Site Size | Dimensions — Aparece quando **lot_size** ou **site_dimensions** estão no modelo. Mesma estrutura do grupo Location.

Actual Age | Effective Age — Aparece quando **actual_age** ou **effective_age** estão no modelo. Mesma estrutura.

As variáveis consumidas pelas linhas agrupadas são **excluídas das linhas individuais de variáveis do modelo** abaixo, evitando dupla contagem nos totais de ajuste.

11.6 Linhas das variáveis do modelo

Abaixo das linhas agrupadas (ou diretamente abaixo de BASE VALUE, se não houver linhas agrupadas), uma linha por preditor restante do modelo mostra:

- **Valores factuais** do avaliando e de cada comparável (os valores efetivos dos dados)
- **Contribuição de valor (VC)** — a contribuição da função g para o valor previsto daquela linha
- **Ajuste** (somente comparáveis) — VC do avaliando menos VC do comparável

11.7 Linhas CQA|Residual e de características residuais

A linha **CQA|Residual** mostra o escore CQA de cada imóvel e contém uma fórmula para o **resíduo remanescente** — a parcela do resíduo ainda não alocada a características específicas. A fórmula é:

$$\text{Resíduo remanescente} = \text{Resíduo total} - \sum (\text{VCs das características residuais abaixo})$$

Abaixo disso vêm as **linhas de características residuais** — 5 linhas nomeadas (Location, View/Appeal, Condition, Quality, Other) mais 6 linhas em branco. São células de entrada nas quais o avaliador informa as contribuições de valor de características não captadas pelo modelo. À medida que o avaliador preenche valores, a fórmula do Resíduo remanescente diminui automaticamente.

Para cada comparável, a coluna de ajuste contém uma fórmula: VC da característica do avaliando – VC da característica do comparável.

11.8 Fórmula do preço de venda ajustado

A linha **Adjusted Sale Price** contém fórmulas do Excel:

- **Avaliando:** soma de todas as células de contribuição de valor, de BASE VALUE até a última linha de característica residual. Isso representa a previsão total do modelo para o avaliando mais quaisquer características residuais alocadas pelo avaliador.
- **Comparáveis:** Net SP + soma de todas as células de ajuste, da primeira linha agrupada ou de variável até a última linha de característica residual. Isso dá o preço de venda do comparável após todos os ajustes derivados da regressão e informados pelo avaliador.

11.9 Proteção da planilha e células editáveis

Cada planilha é **protegida** para evitar a modificação acidental de fórmulas e dados. A proteção bloqueia:

- Todas as células de fórmula (Net SP, resíduo remanescente, ajustes, preço de venda ajustado etc.)
- Todas as células de dados (endereços, preços de venda, valores factuais, contribuições de valor do modelo)
- Rótulos e cabeçalhos

As únicas células **desbloqueadas** (editáveis) são as entradas de contribuição de valor das características residuais — as células sob a linha CQA|Residual nas quais o avaliador informa os detalhes. Elas são estilizadas com fundo amarelo-claro para indicar que são editáveis.

11.10 Trabalhando com a grade no Excel

1. **Abra o arquivo** no Excel ou em um aplicativo de planilha compatível.
2. **Revise os ajustes da regressão** — os ajustes derivados do modelo já vêm preenchidos e bloqueados.
3. **Aloque o resíduo** — nas linhas de características residuais (células amarelas), informe sua avaliação de quanto do resíduo remanescente é atribuível a cada característica (Location, View, Condition, Quality etc.). A fórmula do Resíduo remanescente diminuirá conforme você aloca.
4. **Verifique o Adjusted Sale Price** — essa fórmula é atualizada automaticamente conforme você informa valores de características residuais.
5. **Várias planilhas** — se você selecionou mais de 3 comparáveis, navegue entre as planilhas. Cada planilha tem o mesmo avaliando na coluna da esquerda, com 3 comparáveis diferentes.

Dica: O objetivo é alocar o resíduo CQA entre as linhas de características até que o Resíduo remanescente fique próximo de zero. Isso demonstra que você consegue explicar a diferença entre a estimativa de regressão de cada comparável e seu preço de venda efetivo.

12 Baixando relatórios

Depois de ajustar um modelo, a geração de relatórios é um fluxo de trabalho Quarto em duas etapas: **Generate Quarto Report** (seção 9 da barra lateral no modo de avaliação, 7 nos demais) grava um pacote de relatório autocontido, e **Convert Quarto Report** (seção 10) o renderiza — ou qualquer outro arquivo Quarto — nos formatos finais.

12.1 Generate Quarto Report

Clicar em **Generate Quarto Report** grava um pacote Quarto autocontido na pasta de saída do projeto, em `<base>_qmd/`, contendo a fonte `.qmd` preenchida, todos os recursos de gráfico pré-gerados (PNG e PDF), os dados do relatório e um modelo de referência do Word. Nenhuma renderização ocorre nesta etapa — o pacote pode ser editado à mão, combinado com outras fontes Quarto por meio de `includes` do Quarto, ou pré-visualizado isoladamente com `quarto preview`.

12.2 Convert Quarto Report

A Seção 10 oferece um seletor de arquivos (com o pacote gerado mais recentemente no projeto atual como padrão — mas você pode navegar até qualquer arquivo `.qmd`) e caixas de seleção de formato:

- **HTML** — um arquivo `.html` com renderização matemática KaTeX, imagens embutidas e sumário. Usa o tema Bootstrap Flatly e a fonte Roboto Condensed.
- **Word** — um arquivo `.docx` adequado para edição e distribuição. Inclui sumário, números de página e usa um modelo de referência personalizado para estilo consistente.
- **PDF** — composto com LuaLaTeX para saída de qualidade profissional. O tamanho do papel segue a configuração de localidade (Letter ou A4). Inclui páginas em paisagem para matrizes de interação grandes e usa Roboto Condensed com Latin Modern Math para as equações. A opção de PDF só é oferecida quando uma distribuição LaTeX é detectada.

Clique em **Convert** para renderizar os formatos selecionados. A renderização roda em segundo plano — uma caixa de diálogo modal mostra o tempo decorrido e o progresso do Quarto, enquanto o aplicativo permanece responsivo. Uma notificação confirma a localização do arquivo ao término.

Todas as operações (ajuste do modelo, downloads de dados, cálculos RCA, geração da grade de vendas e renderização de relatórios) são registradas com carimbos de início/fim e tempos decorridos em `<datafile>_earthui_log.txt`, na pasta de saída, para solução de problemas e monitoramento de desempenho.

12.3 Conteúdo do relatório

Todo relatório gerado inclui as seguintes seções:

1. **Descrição do conjunto de dados** — número de observações, variável(is)-resposta, preditores utilizados, preditores categóricos
2. **Especificação do modelo** — degree, situação da validação cruzada, número de termos, preditores no modelo
3. **Allowed Interactions** (somente $\text{degree} \geq 2$) — a matriz de interações com marcas de verificação, formatada conforme o tipo de saída
4. **Resumo dos resultados** — R^2 , GRSq , GCV , RSS e $\text{CV } R^2$ (se a validação cruzada estiver habilitada)
5. **Equação do modelo** — a equação completa do modelo Earth em notação LaTeX

6. **Coeficientes e funções de base** — tabela de todos os termos, seus coeficientes e as funções de base que os definem
7. **Importância das variáveis** — gráfico de barras e tabela ordenada dos escores de importância dos preditores
8. **Contribuições das funções g** — gráficos para cada função g: gráficos de linha para termos univariados, gráficos de perspectiva 3D e de contorno para interações bivariadas
9. **Matriz de correlação** — mapa de calor das correlações entre preditores
10. **Diagnósticos** — gráficos de Resíduos vs. Ajustados, Q-Q normal e Observado vs. Previsto
11. **Decomposição ANOVA** — tabela mostrando a contribuição de cada função de base para a RSS
12. **Earth Output** — saída de texto bruta de `earth::summary()`, incluindo os detalhes da etapa de poda

Em modelos de múltiplas respostas, as seções 4–8 e 10 se repetem para cada variável-resposta.

***Dica:** Os relatórios exigem o pacote `R quarto` e uma instalação do Quarto. Para saída em PDF, também é preciso ter uma distribuição LaTeX instalada (p. ex., via `tinytex::install_tinytex()`) — caso contrário, a caixa de seleção de PDF fica oculta.*

13 Os Post Processing Modules: elastic net e GAM

O pacote earthUI traz dois Post Processing Modules — uma rotina de elastic net construída sobre `glmnet` e uma rotina de GAM construída sobre `mgcv`. Cada uma é tratada em profundidade por um artigo próprio nesta edição; esta seção explica *para que* elas servem no fluxo de trabalho do earthUI e como as três rotinas cooperam.

13.1 Por que ajustes de apoio?

O modelo earth produz a conclusão de valor. Os Post Processing Modules existem para **corroborá-la**: duas estimativas adicionais do mesmo avaliando, produzidas por motores metodologicamente distintos, ajustadas aos mesmos dados do projeto. Quando uma reestimação por elastic net da base do earth e um suavizador GAM independente chegam a valores materialmente iguais, a conclusão do avaliador deixa de repousar sobre as premissas de um único modelo — uma vantagem significativa em contextos de revisão e de litígio. Quando eles *discordam*, essa discordância é ela própria diagnóstica, apontando em geral para dados escassos, um termo excessivamente flexível ou um comparável atípico que merece reexame. Nenhum dos módulos se sobrepõe ao modelo earth; eles são evidência de apoio, deliberadamente apresentada sem ponderação.

13.2 Inicialização e estado compartilhado

Cada rotina é um aplicativo Shiny local separado:

- `earthUI::launch_glmnet()` — o aplicativo de elastic net, na porta 7879.
- `earthUI::launch_mgcv()` — o aplicativo de GAM, na porta 7880.

Os três aplicativos leem a mesma árvore de projetos e o mesmo banco de dados de configurações, de modo que o projeto aberto no earthUI fica imediatamente disponível nos módulos — junto com os campos herdados que o earthUI salva para ele: a Response (resposta), a Effective Date, o tipo e o valor do CQA do avaliando no RCA, e a designação de área construída. Você não os informa de novo; os módulos os leem, para que os mesmos valores percorram os três ajustes.

13.3 A rotina de elastic net

A rotina `glmnet` importa as funções de base do modelo earth — suas dobradiças, interações e termos lineares — como colunas de uma matriz de delineamento e as reestima sob regularização lasso/elastic net (veja “`glmnetUI Import`”, acima, para a mecânica e a ressalva quanto à coluna de pesos). O resultado preserva os termos interpretáveis e geometricamente significativos do earth, deixando que o regularizador encolha ou descarte aqueles que não se justificam. Veja o artigo sobre `glmnetUI` nesta edição para o tratamento completo.

13.4 A rotina de GAM

A rotina `mgcv` segue o caminho inverso: em vez de reaproveitar a base do earth, ela ajusta novos termos suavizados penalizados, usando as posições dos nós do modelo earth como pontos de partida (veja “`mgcvUI Auto-Export`”, acima). Como o GAM rederiva as formas funcionais sob um paradigma de suavização diferente, sua concordância com o modelo earth é corroboração genuinamente independente, e não um eco. São suportados modelos earth com $\text{degree} \leq 2$. Veja o artigo sobre `mgcvUI` nesta edição para o tratamento completo.

13.5 Um fluxo de trabalho prático

1. Ajuste e refine **primeiro o modelo earth**, no earthUI, até ficar satisfeito com ele. Todo ajuste bem-sucedido ($\text{degree} \leq 2$) exporta automaticamente o **.rds** que os módulos consomem.
2. Abra as rotinas **glmnet** e **mgcv** sobre o mesmo projeto, importe o resultado do earth e ajuste. Trabalhe em uma rotina de cada vez — os ajustes são demorados e as rotinas se coordenam por meio do projeto compartilhado, de modo que o ritmo sensato é sequencial, não paralelo.
3. Itere livremente. Cada aplicativo mantém seus resultados enquanto você alterna entre eles, e nada obriga a gerar relatórios pelo caminho.
4. Somente quando os três resultados o satisfizerem, volte a cada rotina e gere seu relatório Quarto. Cada método registra seu ajuste e sua conclusão de valor no registro trilogy do projeto, mantendo os três relatórios rastreáveis a uma única execução comparativa.

Dica: Os módulos são opcionais em todos os níveis: o fluxo de trabalho próprio do earthUI — ajuste, RCA, Sales Grids, relatórios — é completo sem eles. Recorra a eles quando uma conclusão exigir corroboração: um avaliando difícil, um segmento de mercado com poucos dados ou um trabalho com probabilidade de passar por revisão.

14 Conjunto de dados de demonstração: Appraisal_1.csv

O earthUI inclui um conjunto de dados MLS de demonstração para explorar o fluxo de trabalho de avaliação. Carregue-o programaticamente com:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
df <- import_data(demo_file)
```

Ou, no aplicativo Shiny, copie-o para a pasta `in/` de um projeto (ou anexe-o como arquivo de dados inicial na criação do projeto) e selecione-o na Seção 2 (Import Data).

14.1 Descrição

O arquivo contém 1.501 vendas residenciais mais 1 imóvel avaliando na linha 1 (1.502 linhas ao todo), de uma exportação simulada de MLS. Os dados representam vendas de casas unifamiliares em um mercado com várias regiões, com uma amplitude de tamanhos, idades e localizações de imóveis.

Não são dados reais, mas se baseiam em um bairro realista do Norte da Califórnia. Todas as informações de identificação foram alteradas ou removidas.

14.2 Colunas

Coluna	Tipo	Tipo especial	Descrição
weight	numeric	—	Peso da observação (0 = excluir do ajuste)
id	numeric	display_only	Identificador interno do registro
property_id	numeric	display_only	Identificador do imóvel no MLS
listing_id	character	display_only	Número do anúncio no MLS
parcel_number	character	display_only	Número de matrícula do cadastro municipal (APN)
street_address	character	display_only	Endereço do imóvel
city_name	character	display_only	Cidade
postal_code	character	display_only	CEP
county_name	character	display_only	Município
contract_date	Date	contract_date	Data do contrato de venda (calcula sale_age)
sale_age	numeric	—	Dias da data do contrato até a data de referência (pré-calculado)
coe_date	Date	display_only	Data de fechamento da escritura
listing_status	character	display_only	Situação do anúncio (p. ex., "Sold")
sale_price	numeric	(resposta)	Preço de venda — variável-resposta
rent	numeric	—	Aluguel mensal (para modelos de múltiplas respostas)
list_price	numeric	display_only	Preço de anúncio
original_list_price	numeric	display_only	Preço de anúncio original
living_sqft	numeric	living_area	Área construída bruta em pés quadrados
beds_total	integer	—	Número de dormitórios
baths_total	numeric	—	Contagem total de banheiros (p. ex., 2.5 = 2 completos + 1 lavabo)
lot_size	numeric	lot_size	Tamanho do lote em pés quadrados
area_id	integer	area	Identificador de região do MLS
area_text	character	display_only	Nome da região
age	numeric	actual_age	Idade do imóvel em anos
year_built	integer	display_only	Ano de construção
latitude	numeric	latitude	Latitude (arredondada para 3 casas decimais no modelo)
longitude	numeric	longitude	Longitude (arredondada para 3 casas decimais no modelo)
latitude6	numeric	display_only	Latitude com precisão total
longitude6	numeric	display_only	Longitude com precisão total
garage_spaces	integer	—	Número de vagas de garagem
fp_count	integer	—	Número de lareiras
no_of_stories	numeric	—	Número de pavimentos
style	character	—	Estilo arquitetônico
view	character	—	Tipo de vista (p. ex., "Neighborhood", "Hills")
days_on_market	integer	dom	Dias no mercado
listing_date	Date	listing_date	Data do anúncio
sale_concessions	numeric	concessions	Concessões do vendedor

14.3 Início rápido sugerido

1. Inicie o earthUI: `earthUI::launch()`
2. Crie um projeto com Purpose **For Appraisal** no assistente New Project, anexando `Appraisal_1.csv` como arquivo de dados inicial
3. Selecione `Appraisal_1.csv` na Seção 2 (Import Data)
4. Selecione `sale_price` como resposta
5. Atribua os tipos especiais conforme a tabela acima
6. Inclua os preditores: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (como fator), `age`, `latitude`, `longitude`, `garage_spaces`
7. Defina degree como 1 e clique em **Fit Earth Model**
8. Baixe a saída intermediária (Passo 6) e revise o ordenamento por CQA
9. Calcule os ajustes RCA (Passo 7) com um escore CQA de ~5,00
10. Gere a Sales Comparison Grid (Passo 8)

Dica: O conjunto de dados de demonstração é uma boa forma de explorar todos os recursos do earthUI antes de importar seus próprios dados de MLS. Os preditores sugeridos acima costumam produzir um modelo com R^2 acima de 0,85.

15 Experimental: extração de texto com Prolog e regras de derivação

A versão 0.10.0 introduz um recurso **opcional e genuinamente experimental** para extrair características prontas para modelagem a partir das colunas de texto descritivo de uma exportação de MLS — sobretudo a coluna “Public Remarks”, mas também as observações do corretor e listas de características delimitadas por vírgulas, como características dos cômodos ou estruturas existentes. O motor de extração é Prolog e, em particular, **Definite Clause Grammars (DCGs)**: as regras que reconhecem expressões como “move in ready”, “3 car garage” ou “owned solar” no texto das observações são escritas como regras gramaticais Prolog.

Este recurso ainda está em desenvolvimento. Ele já pode ser usado *parcialmente* para extração hoje — a gramática incluída já reconhece um conjunto útil de características — mas adaptá-lo ao vocabulário do seu próprio MLS exige ler e escrever Prolog, e DCGs em particular. Se esse não for o seu caso, pule este capítulo por completo: o recurso vem desligado por padrão, e o restante do earthUI funciona exatamente igual sem ele.

Aviso

Este é um recurso experimental em desenvolvimento ativo. Sua gramática de extração não é completa nem validada para trabalho de avaliação em produção — as colunas extraídas precisam ser conferidas contra o texto de origem antes de serem usadas em um modelo. Usar o recurso de forma eficaz exige conhecimento prático de Prolog e DCGs. Tudo o que ele faz é aditivo e reversível; se algo se comportar mal, desligue o seletor em Settings e siga normalmente.

15.1 O que ele faz

Os anúncios de MLS carregam uma grande quantidade de informação relevante para o valor em texto livre e em campos de lista delimitados. Esta etapa opcional transforma esse texto em colunas comuns que sua regressão pode usar, por meio de dois papéis especiais de coluna (Capítulo 6):

- **Colunas remarks** são interpretadas pela gramática DCG de Prolog em uma coluna de característica por atributo extraído. O prefixo da coluna é construído a partir das primeiras letras do nome da coluna de origem: `public_remarks` gera colunas `pr_*`, `agent_remark` gera `ar_*`, e assim por diante.
- **Colunas list** (p. ex. `existing_structures = "Shed, Workshop"`) são divididas em colunas `TRUE/FALSE <col>_<value>` one-hot. Essa parte é R puro e não precisa de Prolog algum.

A gramática incluída (`inst/prolog/mls_remarks.pl`) atualmente extrai, entre outras coisas: um nível graduado de estado de conservação com um `condition_score` de 1 a 5; indicadores e tipos de vista; indicadores de reforma de cozinha e banheiro; piso de madeira e outros revestimentos; uma lista e uma contagem de melhorias; vagas e área de garagem; piscina, spa e outras características externas; contagem de lareiras; edificações acessórias (galpão, celeiro, oficina, casita, casa de hóspedes, canil etc.) com tamanhos extraídos de dimensões como “40x50”; estacionamento e depósito para motorhome; propriedade versus arrendamento de sistema solar e armazenamento em baterias; tamanho do lote (incluindo conversão de acre para pé quadrado); valores ligados ao contexto da venda associados a um avaliando (redução de preço, taxas de condomínio, arrendamento ou quitação do sistema solar, Mello-Roos); indícios de localização; e um escore de sentimento de destaques/preocupações em $[-1, 1]$.

15.2 Requisitos: vProlog (Trealla), e não SWI-Prolog

A interpretação das observações roda sobre o pacote R **vProlog** — um motor Trealla Prolog autocontido. Ele não está no CRAN; instale-o a partir do repositório r-universe do autor, como mostrado em “Instalação do earthUI e do vProlog”, no início deste artigo (binários pré-compilados para Windows e macOS, código-fonte em github.com/wcraytor/vProlog). Nenhuma instalação de Prolog no sistema (SWI-Prolog ou outra) é exigida ou utilizada. Se o vProlog estiver ausente, a interpretação das observações é pulada com um aviso, enquanto a divisão de `list` ainda é executada. O auxiliar `vProlog_available()` informa se o motor e a gramática incluída estão ambos presentes.

15.3 Habilitando e executando

1. Abra **Settings** (engrenagem) e marque **Enable Prolog / list column processing**. O seletor vem desligado por padrão, traz uma observação “advanced feature — for users with Prolog expertise”, é lembrado por projeto e é exibido na caixa de diálogo Project Information.
2. Na tabela de preditores, defina o papel **Special** das colunas relevantes como `remarks` ou `list` (várias colunas podem ter cada papel).
3. Use a seção **Execute Prolog Processing**, situada entre as Seções 4 e 5 da barra lateral. Ela oferece duas etapas:

1. **Expand lists & remarks** — executa a gramática DCG sobre cada coluna `remarks` e divide em one-hot cada coluna `list`. A expansão é *incremental*: apenas as colunas designadas desde a última execução são processadas, e o botão fica desabilitado quando tudo o que foi designado já tiver sido expandido. Uma barra de progresso acompanha a análise.

2. **Execute derivation rules** — aplica suas próprias regras Prolog `derive/2` sobre as colunas expandidas (veja abaixo). As regras passam por verificação de sintaxe e por validação em execução simulada antes de rodar.

15.4 Regras de derivação

Para cada linha de dados, o earthUI declara as colunas da linha como fatos `cell(Name, Value)` e coleta todas as soluções de suas regras `derive(Col, Val)` — cada `Col` distinto torna-se uma nova coluna, e o *primeiro* valor encontrado por coluna prevalece, de modo que uma cláusula alternativa pode vir em seguida. Predicados auxiliares (`count_true/2`, `count_false/2`, `count_true_prefix/2`, `cell_or/3`, `is_na/1`) dão suporte a agregação e ao tratamento de NA, e fatos `drop/1` descartam colunas depois que as regras rodam. Por exemplo:

```
% Merge a collinear cluster of solar flags into one boolean.
derive(solar_any, B) :-
  ( cell(pr_has_solar, true)
  ; cell(pr_solar_owned, true)
  ; cell(pr_has_battery, true) ) -> B = true ; B = false.

% Count amenities instead of OR-ing them (numeric column).
derive(amenity_count, N) :- count_true_prefix(pr_has_, N).

% Threshold -> boolean, with a fallback clause.
derive(big_outbuilding, true) :- cell(pr_outbuilding_sqft, S), S >= 1000.
derive(big_outbuilding, false).

% Discard a raw granular column once consolidated.
drop(pr_price_solar_lease).
```

As regras ficam em um arquivo `<project>_rules.pl` por projeto, criado a partir de um modelo integralmente comentado, e são editadas por **Edit derivation rules...** em um editor de código flutuante e arrastável, que também pode ser acoplado à barra lateral. Um botão **Check** valida as regras (varredura de sintaxe, verificação de carga e uma execução simulada contra a linha 1 dos seus dados) sem executá-las.

15.5 Arquivo de dados ativo versionado

Depois de qualquer uma das etapas, o data frame aumentado é gravado na pasta de entrada do projeto como `<original_name>_<YYMMDD_HHMMSS>.xlsx` e torna-se o arquivo de dados **ativo** do projeto — reabrir o projeto carrega os dados aumentados, de modo que as expansões nunca são recalculadas, e esse mesmo arquivo é o que o `glmnetUI` e o `mgcvUI` consomem. As colunas de origem das observações brutas são automaticamente mantidas fora do modelo; as colunas geradas *não* são incluídas automaticamente como preditores. A análise de texto pode produzir muitas colunas, portanto marque como **Latent** as que você (ainda) não quiser na regressão, ou desabilite-as com `×` (Capítulo 6).

15.6 Interface programática e arquivos incluídos

Todo o pipeline é exportado para uso em lote: `execute_prolog_processing(df, remarks_cols, list_cols, rules_text)` orquestra as passagens, construído sobre `remarks_to_columns()`, `split_list_column()`, `apply_user_rules()` e `validate_rules()`. Os fontes Prolog incluídos em `inst/prolog/` são `mls_remarks.pl` (a gramática DCG e a agregação de características), `mls_analyze.pl` (um driver de E/S de arquivos para análise em lote), `mls_rules.pl` (os predicados auxiliares disponíveis às regras de derivação) e `climate_regions_ca.pl` (abaixo).

15.7 Experimento relacionado: prioris por região climática

Um recurso experimental complementar trata das características com *dados escassos* — aquelas que a regressão não consegue estimar por aparecerem em apenas um punhado de comparáveis (um climatizador evaporativo, digamos). Três auxiliares em R — `climate_region_for()`, `market_area_profile()` e `climate_feature_priors()` — classificam a localização de um projeto (município + cidade) em uma de sete regiões climáticas da Califórnia reconhecidas pelos compradores (uma redução prática das 16 zonas climáticas de edificação da California Energy Commission) e retornam prioris ordinais de valor contributivo para características de climatização e resfriamento (aquecimento, ar-condicionado, climatizador evaporativo, exaustor de casa inteira, filtragem) por região. O módulo Prolog `climate_regions_ca.pl` espelha a mesma classificação para uso dentro das regras de derivação; ambos os lados são gerados de uma única fonte canônica em R, de modo que não podem divergir. Os prioris destinam-se a *informar* a valoração de características e a servir de verificação de plausibilidade dos coeficientes estimados — as atribuições de região (em particular no caso de municípios que abrangem várias regiões) devem ser revisadas pelo avaliador.

15.7.1 Além da Califórnia: pacotes de jurisdição e módulos de região de mercado

A Califórnia é a primeira jurisdição, não o limite. O `earthUI` é usado internacionalmente, e o conhecimento de prioris climáticos está sendo reestruturado como **pacotes de jurisdição**: tabelas de fatos autocontidas e com formato de dados (definições de zona, atribuições de unidades administrativas, prioris de características) atreladas aos códigos de país e de nível administrativo do sistema de projetos, de modo que um pacote possa descrever um estado dos EUA, um Land alemão ou um país pequeno inteiro. Uma camada complementar de **módulos de região de mercado** capta o vocabulário local e prioris refinados para uma região de mercado nomeada.

Ambos foram concebidos para serem elaborados *em campo* por Engenheiros de Avaliação e avaliadores, em seu próprio idioma — e, realisticamente, com assistência de LLM: o mundo não tem muitos programadores Prolog, de modo que os artefatos de campo são deliberadamente tabelas, e não programas. O formato do pacote é um esquema documentado com um exemplo completo resolvido (a própria Califórnia), expressamente para que um assistente como o Claude possa redigir um pacote válido a partir de uma descrição de mercado em linguagem comum; o earthUI gera o Prolog subjacente a partir do pacote, e um validador bloqueante (verificações de esquema, cobertura, integridade referencial e execução simulada) precisa passar antes que um pacote possa ser habilitado. Todo pacote elaborado em campo carrega procedência obrigatória — o autor responsável, a data e a autoridade citada — de modo que o profissional responsável, e não a ferramenta de redação, respalda o conhecimento, e os relatórios podem citar exatamente o pacote e a versão utilizados.

Dica: Um padrão prático: expanda as observações, consolide os indicadores granulares `pr_*` em um punhado de contagens ou booleanos derivados com regras `derive/2`, aplique `drop/1` aos indicadores brutos que você consolidou e mantenha apenas as colunas derivadas como preditores candidatos. Isso mantém a lista de preditores curta e a colinearidade administrável.

Referências

- Friedman, Jerome H. (1991). “Multivariate Adaptive Regression Splines”. Em: *The Annals of Statistics* 19.1, pp. 1–141. DOI: [10.1214/aos/1176347963](https://doi.org/10.1214/aos/1176347963).
- Milborrow, Stephen (1 de out. de 2024). *earth: Multivariate Adaptive Regression Splines*. R package. URL: <https://CRAN.R-project.org/package=earth>.
- (s.d.[a]). *Notes on the earth Package*. URL: <http://www.milbo.org/doc/earth-notes.pdf>.
- (s.d.[b]). *Variance Models in earth*. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.

glmnetUI: uma interface interativa para o pacote glmnet (Lasso e Elastic Net)

Tradução para o português, elaborada com auxílio de IA e revisada pela redação; a versão em inglês é a autorizada.

William Bert Craytor

Volume 1, Número 1 · Verão de 2026

Publicado em 1 de julho de 2026

DOI: [10.67330/ja8s2k39](https://doi.org/10.67330/ja8s2k39)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da [licença Creative Commons Atribuição 4.0](#).

Resumo. glmnetUI é uma interface gráfica de usuário para o pacote **glmnet** do R, que ajusta modelos lineares generalizados regularizados por meio das penalizações Lasso, ridge e elastic-net. Oferece três modos de uso — modelagem preditiva geral, avaliação de imóveis e análise de região de mercado — e conduz o usuário pela importação de dados, pela configuração do modelo, pelo ajuste com validação cruzada, pelas trajetórias dos coeficientes e pelos relatórios baixáveis. Este artigo documenta os requisitos de formato de dados do glmnetUI, o fluxo de trabalho de modelagem, as telas de saída e a referência completa de funcionalidades.

Palavras-chave: Lasso, regressão ridge, elastic net, regressão regularizada, glmnet, R, interface gráfica de usuário, avaliação de imóveis

Nota do editor (julho de 2026): Pouco antes da publicação, o glmnetUI foi consolidado dentro do earthUI (versão 0.11.0 e posteriores), que agora oferece o fluxo de trabalho de elastic net aqui descrito juntamente com os fluxos de trabalho de MARS e GAM. O glmnetUI permanece disponível em sua forma final publicada (0.5.0) em <https://github.com/wcraytor/glmnetUI>; o desenvolvimento futuro continua no earthUI. Este artigo se refere ao glmnetUI tal como foi publicado.

Sumário

1	Introdução	3
1.1	Histórico e antecedentes	3
1.1.1	O Lasso	3
1.1.2	Regressão ridge e elastic net	4
1.1.3	O pacote glmnet	4
1.1.4	Comparação com MARS (earth) e GAM (mgcv)	4
1.1.5	Fluxo de trabalho recomendado	5
1.2	O que é o glmnetUI?	6
1.3	Três modos de uso	6
1.4	Recursos específicos do setor imobiliário	7
1.5	Primeiros passos	7
2	Requisitos dos dados de entrada do MLS	9
2.1	Formato e estrutura do arquivo	9
2.2	Colunas exigidas no modo de avaliação	9
2.3	Convenções de nomenclatura das colunas especiais	10
2.4	Qualidade e completude dos dados	10
2.5	Posição da linha do avaliando	10

3	Modo General Purpose	12
3.1	Visão geral	12
3.1.1	Skip First Row	12
3.1.2	Configurações por propósito	12
3.2	O fluxo de trabalho da barra lateral	12
3.3	Projetos	13
3.4	Importação do earthUI (opcional)	13
3.5	Abas do painel principal	14
3.6	Persistência das configurações	14
3.7	Modo escuro	15
3.8	Localidade e configurações regionais	15
3.8.1	Salvar padrões	15
4	Modo Appraisal	16
4.1	Tratamento da linha do avaliando	16
4.2	Data-base e sale age	16
4.3	Designações especiais de colunas	16
4.4	Visão geral dos ajustes de RCA	16
5	Modo Market Area Analysis	17
5.1	Diferenças em relação ao modo Appraisal	17
5.2	Quando usar o modo Market	17
6	Seleção de variáveis	18
6.1	Variável resposta	18
6.2	A tabela de preditores	18
6.3	Deteção de tipos de dados e sobreposições	18
6.4	Sinais esperados e sua imposição	18
6.5	Referência dos tipos especiais de coluna	19
7	Seleção de parâmetros	20
7.1	Alpha (parâmetro de mistura)	20
7.2	Seleção de lambda	20
7.3	Family	20
7.4	Standardize	21
7.5	Restrições de sinal	21
7.6	Relaxed Lasso	21
7.7	Semente aleatória	21
7.8	Matriz de interações	22
7.9	Bloquear o efeito principal (somente interação)	22
7.10	Parâmetros avançados	22
7.11	Padrões das configurações	23
8	Ajuste do modelo	24
8.1	O botão de ajuste	24
8.2	O que acontece durante o ajuste	24
9	Abas de resultados	25
9.1	Data Preview	25
9.2	Equation	25
9.3	Correlation	25
9.4	Summary	25

9.5	Coefficients	26
9.6	Variable Importance	26
9.7	Contributions	26
9.8	ANOVA	26
9.9	Diagnostics	26
9.10	Glmnet Output	27
9.11	Report	27
10	Download dos dados	28
10.1	Colunas de saída	28
10.2	Escores CQA	28
11	Cálculos e download do RCA	29
11.1	Abertura da caixa de diálogo do RCA	29
11.2	Interpolação do escore CQA	29
11.3	Colunas de saída	29
12	Grade de comparação de vendas	30
12.1	Caixa de diálogo de seleção de comparáveis	30
12.2	Estrutura da pasta de trabalho	30
12.3	Proteção das planilhas e células residuais	30
13	Download de relatórios	31
13.1	O fluxo Quarto em duas etapas	31
13.2	Formatos de relatório	31
13.3	Conteúdo dos relatórios	31
14	Comparação com o earthUI	33
14.1	Principais diferenças	33
14.2	Recursos compartilhados	33
14.3	Quando usar cada um	33
15	Conjunto de dados de demonstração: Appraisal_1.csv	35
15.1	Descrição	35
15.2	Colunas	35
15.3	Início rápido sugerido	35
16	Requisitos de sistema e solução de problemas	37
16.1	Plataformas suportadas	37
16.2	Dependências opcionais	37
16.3	Notas por plataforma	37
16.4	Degradação graciosa	37
16.5	Solução de problemas	37

1 Introdução

1.1 Histórico e antecedentes

1.1.1 O Lasso

O **Lasso** (Least Absolute Shrinkage and Selection Operator) foi introduzido por Robert Tibshirani em 1996. O nome é um jogo de palavras deliberado com o laço do vaqueiro — uma

corda que restringe e seleciona. Tibshirani o escolheu porque o método literalmente “laça” os coeficientes, restringindo seus valores absolutos e puxando alguns exatamente para zero, de modo a selecionar quais variáveis permanecem no modelo.

A intuição de Tibshirani foi acrescentar uma penalização L_1 (soma dos valores absolutos dos coeficientes) ao objetivo de mínimos quadrados. Diferentemente da penalização L_2 da regressão ridge (Hoerl e Kennard, 1970), que contrai os coeficientes em direção a zero mas nunca o atinge, a penalização L_1 produz soluções *esparsas* nas quais muitos coeficientes são exatamente zero. Isso torna o Lasso ao mesmo tempo um método de regularização e um método de seleção de variáveis.

1.1.2 Regressão ridge e elastic net

A **regressão ridge** foi proposta por Arthur Hoerl e Robert Kennard em 1970. Ela acrescenta uma penalização L_2 (soma dos quadrados dos coeficientes) que contrai todos os coeficientes proporcionalmente, mas mantém todas as variáveis no modelo. A ridge é eficaz quando os preditores estão correlacionados (multicolinearidade), mas não simplifica o modelo.

O **elastic net**, introduzido por Hui Zou e Trevor Hastie em 2005, combina as duas penalizações. O parâmetro de mistura α controla a combinação: $\alpha = 1$ é Lasso puro, $\alpha = 0$ é ridge puro, e os valores intermediários resultam em um compromisso. O elastic net supera uma limitação do Lasso: quando os preditores são altamente correlacionados, o Lasso tende a selecionar um e ignorar os demais, ao passo que o elastic net agrupa os preditores correlacionados.

1.1.3 O pacote glmnet

O pacote de R `glmnet` foi desenvolvido por Jerome Friedman, Trevor Hastie e Robert Tibshirani na Universidade Stanford. Ele implementa a regularização elastic net para modelos lineares generalizados por meio de um algoritmo de descida por coordenadas extremamente eficiente (Friedman, Hastie e Tibshirani 2010), posteriormente estendido aos modelos de riscos proporcionais de Cox (Simon et al. 2011) e a todas as famílias de modelos lineares generalizados (Tay, Narasimhan e Hastie 2023). O pacote trata das famílias gaussian (linear), binomial (logística), poisson e outras famílias de GLM.

Entre os recursos principais do `glmnet` estão:

- **Trajetória de regularização:** ajusta o modelo ao longo de uma malha de valores de λ em uma única passagem, muito mais rápido do que ajustá-lo separadamente em cada λ .
- **Validação cruzada:** `cv.glmnet()` seleciona automaticamente a intensidade ótima de regularização por validação cruzada de k dobras.
- **Limites dos coeficientes:** admite limites superiores e inferiores para coeficientes individuais (restrições de sinal).
- **Relaxed lasso:** reajusta o modelo sobre as variáveis selecionadas com penalização reduzida ou nula, diminuindo o viés.

1.1.4 Comparação com MARS (earth) e GAM (mgcv)

O aplicativo `earth` e seus dois Post Processing Modules (módulos de pós-processamento; `glmnet` e `mgcv`) utilizam mecanismos de modelagem distintos. A tabela a seguir resume as principais diferenças:

Característica	glmnet (Elastic Net)	earth (MARS)	mgcv (GAM)
Tipo de modelo	Linear (com regularização)	Linear por partes (splines adaptativos)	Não linear suave (splines penalizados)
Não linearidade	Somente por interações ou expansão de base	Automática, por funções dobradiça com nós guiados pelos dados	Automática, por funções suaves (thin plate, cúbicas etc.)
Seleção de variáveis	Nativa, pela penalização L_1 (Lasso)	Nativa, por passos para frente/para trás com poda por GCV	Opcional, por contração de dupla penalização (<code>select=TRUE</code>)
Interações	Produtos cruzados aos pares (difíceis de interpretar)	Produtos baseados em dobradiças (interpretáveis, até grau 3)	Suavizadores de produto tensorial (<code>te()</code> , <code>ti()</code>)
Interpretabilidade	Os coeficientes são pesos lineares; fáceis de explicar individualmente	As funções g mostram efeitos parciais lineares por partes para cada variável	Curvas suaves de efeito parcial; muito intuitivas
Mais indicado para	Dados de alta dimensão, seleção de variáveis, amostras pequenas	Deteção automática de não linearidade e de interações	Relações não lineares suaves, modelagem flexível
Ajustes do RCA	Ajustes lineares por variável	Ajustes lineares por partes via funções g	Ajustes suaves via efeitos parciais
Controle do sobreajuste	Lambda validado por validação cruzada	Poda baseada em GCV	Estimação do parâmetro de suavização por REML/GCV

O **MARS** (Multivariate Adaptive Regression Splines) foi introduzido por Jerome Friedman em 1991. Ele constrói modelos lineares por partes selecionando de forma adaptativa funções dobradiça $\max(0, x - k)$ e as posições de seus nós a partir dos dados. A implementação em R é o pacote **earth**, de Stephen Milborrow. O MARS se destaca por descobrir automaticamente relações não lineares e interações sem que o usuário precise especificá-las. Suas funções g (termos de base agrupados por variável) fornecem curvas de efeito parcial altamente interpretáveis.

Os **GAM** (Generalized Additive Models, modelos aditivos generalizados) foram formalizados por Hastie e Tibshirani (1990). A implementação em R **mgcv**, de Simon Wood, oferece splines de regressão penalizados com seleção automática da suavização por REML ou GCV. Os GAM produzem curvas suaves de efeito parcial que são intuitivas de visualizar e de explicar. Quando o **earthUI** exporta as posições dos nós para o **mgcvUI**, os nós derivados do **earth** servem de ponto de partida para os termos suaves do GAM, combinando a colocação adaptativa de nós do MARS com a estimação suave do GAM.

1.1.5 Fluxo de trabalho recomendado

Para a avaliação de imóveis e aplicações similares que exigem modelos interpretáveis e defensáveis:

1. **Comece pelo earthUI** para descobrir as variáveis importantes, as relações não lineares e as interações presentes em seus dados. A seleção automática de base e as funções g do **earth** fornecem excelentes percepções iniciais.

2. **Refine com o mgcvUI** se você quiser efeitos parciais mais suaves. Importe os nós do `earthUI` para o `mgcvUI` para uma transição fluida de modelos lineares por partes para modelos suaves.
3. **Use o glmnetUI** quando precisar de uma seleção de variáveis agressiva (Lasso), quando o número de preditores se aproximar ou exceder o número de observações, ou quando quiser um modelo puramente linear com regularização por motivos de defensabilidade.

1.2 O que é o glmnetUI?

`glmnetUI` é uma interface gráfica de usuário para o pacote `glmnet` do R. Ele é executado como um aplicativo Shiny local — não há login, nem servidor, nem contas. Você o inicia a partir do R, importa um conjunto de dados (CSV ou Excel), configura seu modelo e o ajusta de forma interativa.

A partir da versão atual, o `glmnetUI` deixou de ser um pacote instalado separadamente: ele é distribuído *dentro* do pacote `earthUI` como seu Post Processing Module de elastic net, iniciado com `earthUI::launch_glmnet()`. O aplicativo em si — e tudo o que é descrito neste artigo — não mudou; apenas a instalação e o comando de inicialização são diferentes. As três rotinas (`earth`, `glmnet`, `mgcv`) compartilham uma mesma árvore de projetos, um mesmo banco de dados de configurações e uma mesma cópia da infraestrutura de apoio.

O aplicativo oferece um fluxo de trabalho completo: importação de dados, configuração de variáveis, ajuste do modelo, gráficos de diagnóstico, importância das variáveis, equações do modelo e relatórios baixáveis em formato Word, PDF ou HTML.

A **regressão elastic net** combina duas técnicas de regularização:

- **Regressão ridge** ($\alpha = 0$): contrai proporcionalmente todos os coeficientes em direção a zero. Adequada quando se acredita que todas as variáveis são relevantes.
- **Regressão Lasso** ($\alpha = 1$): pode contrair coeficientes exatamente até zero, realizando seleção automática de variáveis. Produz modelos mais simples e mais interpretáveis.
- **Elastic net** ($0 < \alpha < 1$): uma combinação das duas abordagens. Útil quando os preditores estão correlacionados.

1.3 Três modos de uso

Todo projeto do `glmnetUI` tem um **purpose** (propósito, ou modo de uso), escolhido no momento da criação (Seção 1, *Project*). O propósito do projeto ativo determina quais ferramentas e elementos de interface aparecem; para trabalhar em outro modo, crie ou abra um projeto com aquele propósito. Os três modos são:

- **General** (geral) — regressão elastic net para qualquer tipo de população ou conjunto de dados. Este é o modo padrão. Oferece o fluxo de trabalho completo de regressão regularizada, sem acréscimos específicos de qualquer domínio.
- **For Appraisal** (para avaliação) — regressão elastic net adaptada à avaliação de imóveis. Acrescenta recursos específicos para a avaliação de um único imóvel, incluindo o tratamento do imóvel avaliando, as designações especiais de colunas e o Residual Constraint Approach (RCA).
- **Market Area Analysis** (análise de região de mercado) — regressão elastic net adaptada aos estudos de região de mercado. Acrescenta recursos para analisar grupos de imóveis em um mercado definido, com exclusão opcional da linha do avaliando.

Nos três modos, o mecanismo de modelagem central é idêntico — você está sempre ajustando um modelo elastic net (`glmnet`). A configuração do propósito controla quais ferramentas e elementos de interface adicionais ficam disponíveis.

1.4 Recursos específicos do setor imobiliário

Quando se seleciona **For Appraisal** ou **Market Area Analysis**, o glmnetUI ativa vários recursos concebidos para a análise imobiliária:

- **Designações especiais de colunas** — cada preditor pode ser marcado com um papel especial, como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Essas designações controlam como a coluna é tratada durante o ajuste e na saída.
- **Arredondamento de latitude e longitude** — as colunas designadas como latitude ou longitude são automaticamente arredondadas para 3 casas decimais para evitar sobreajuste.
- **Coluna Sale Age** — quando uma coluna é designada como `contract_date` e uma data-base (Effective Date) é informada, o glmnetUI calcula uma coluna `sale_age` (dias entre a data da venda e a data-base) e a substitui como preditor.
- **Cálculos do RCA (somente avaliação)** — no modo de avaliação, depois de ajustar o modelo o glmnetUI pode calcular a saída do Residual Constraint Approach (RCA), que produz os ajustes por comparável, os resumos de ajuste líquido/bruto e um preço de venda ajustado para o imóvel avaliando.

***Dica:** o modo General funciona para qualquer conjunto de dados — financeiro, científico, de engenharia ou imobiliário. Os modos de avaliação e de mercado simplesmente acrescentam recursos de conveniência para os profissionais do setor imobiliário.*

1.5 Primeiros passos

Para usar o glmnetUI:

1. **Instale o earthUI** no R: `install.packages("earthUI")` — a rotina glmnet está incluída. (Veja “Installing earthUI and vProlog”, no artigo sobre o earthUI desta edição, para a versão de desenvolvimento e os componentes opcionais.)
2. **Inicie o aplicativo:** execute `earthUI::launch_glmnet()` no R. O aplicativo abre em seu navegador web na porta 7879. Você também pode acessá-lo diretamente navegando até `http://localhost:7879` no navegador. O aplicativo lembra o último modo de uso empregado e o restaura automaticamente.
3. **Crie ou abra um projeto** na Seção 1 da barra lateral. Um projeto define o propósito e o local de trabalho (país, estado, condado, cidade) e suas pastas de entrada/saída. Os projetos são armazenados sob uma raiz `regProj` comum e são compartilhados com os aplicativos irmãos earthUI e mgcvUI. Abrir um projeto com um propósito diferente é a forma de alternar entre os modos.
4. **Importe seus dados** na Seção 2, selecionando um arquivo CSV ou Excel da pasta de entrada do projeto ativo. Adicione arquivos a essa pasta e clique em **Refresh** para listá-los.
5. **Importe do earthUI (opcional)** — a Seção 3 permite importar um arquivo `.rds` de resultado do earthUI para usar as funções de base dobradiça do earth com a regularização do glmnet. Pule esta etapa se você não tiver um resultado do earthUI.
6. **Configure as variáveis** — escolha a variável resposta e os preditores, defina os tipos de dados e os sinais esperados e atribua os papéis especiais de coluna.
7. **Defina os parâmetros do glmnet** — alpha, lambda, family, restrições de sinal, interações e outras opções.
8. **Ajuste o modelo** — clique em “Fit Glmnet Model” e examine os resultados no painel principal.
9. **Exporte** — baixe as previsões em Excel, gere e converta um relatório Quarto ou (no modo de avaliação) calcule os ajustes do RCA e uma grade de comparação de vendas

(Sales Comparison Grid).

As configurações são automaticamente persistidas no armazenamento local do seu navegador e restauradas quando você recarrega o mesmo arquivo de entrada.

2 Requisitos dos dados de entrada do MLS

Para os fluxos de trabalho de avaliação de imóveis e de análise de mercado, seus dados de entrada geralmente vêm de uma exportação de um Multiple Listing Service (MLS). Este capítulo descreve a estrutura de arquivo esperada e as colunas que o glmnetUI pode utilizar.

2.1 Formato e estrutura do arquivo

O glmnetUI aceita arquivos **CSV** e **Excel** (.xlsx, .xls). Na importação, os nomes das colunas são automaticamente convertidos para **snake_case** — por exemplo, “Living SqFt” torna-se **living_sqft**, “Contract Date” torna-se **contract_date** e “Sale Price” torna-se **sale_price**. Essa normalização garante referências de coluna consistentes ao longo de todo o fluxo de trabalho. O separador de CSV e o separador decimal usados na importação são determinados pelas configurações de localidade (veja o Capítulo 3, “Locale & Regional Settings”).

Seu arquivo de dados deve ser uma tabela plana, com uma linha por imóvel e uma coluna por atributo. A primeira linha do arquivo deve conter os cabeçalhos das colunas.

2.2 Colunas exigidas no modo de avaliação

Embora o glmnetUI funcione com qualquer conjunto de colunas, o fluxo de trabalho completo de avaliação se beneficia da presença das seguintes colunas:

Coluna	Tipo especial	Finalidade
Sale Price	(resposta)	Variável resposta do modelo
Contract Date	contract_date	Usada para calcular sale_age (dias a partir da data-base)
Listing Date	listing_date	Usada com a data do contrato para calcular o DOM quando não há coluna de DOM
Days on Market	dom	Dias em mercado; exibida nas exportações
Concessions	concessions	Concessões na venda; SP líquido = preço de venda – concessões
Living Area (SF)	living_area	Habilita resíduos por SF (residual_sf , cqa_sf)
Lot Size	lot_size	Coluna de tamanho do terreno
Site Dimensions	site_dimensions	Agrupada com o tamanho do terreno
Latitude	latitude	Arredondada para 3 casas decimais
Longitude	longitude	Arredondada para 3 casas decimais
Area ID	area	Identificador da região de mercado / bairro
Actual Age	actual_age	Idade do imóvel
Effective Age	effective_age	Idade efetiva do imóvel
Address	display_only	Mostrada nas exportações; excluída do modelo

Os nomes das colunas da planilha podem estar em um idioma estrangeiro — os nomes “especiais” estão em inglês para que o programa em R possa lhes dar tratamento especial. Fora isso, os nomes de coluna informados aparecem nos modelos de regressão, nos gráficos e, no caso de avaliações, nos relatórios de saída.

Nem todas as colunas são obrigatórias. O glmnetUI se adapta — se uma coluna estiver ausente, o recurso correspondente é simplesmente omitido. Entretanto, para modelos de precificação imobiliária certas colunas são altamente recomendadas para que se obtenha um ajuste aceitável:

1. **Sale Age** — o número de dias entre a data do contrato de venda e a data-base da avaliação ou da análise. Se estiver sendo usado um histórico de vendas de vários anos,

especialmente para períodos superiores a 5 anos, **sale_age** frequentemente desempenha papel central na estimação do preço de venda.

2. **Living Area** (área construída) — também aparece com nomes como “Living Sqft”, “GLA” (gross living area) e assim por diante. Este é outro determinante de primeira ordem do preço de venda.
3. **Total Bath Count** (número total de banheiros) — o total de banheiros completos, de um quarto, de meio banheiro e de três quartos. Por exemplo, dois banheiros completos e um lavabo corresponderiam ao valor 2,5.
4. **Garage Bays** ou **Garage Area** — o número de vagas de garagem ou a área da garagem.
5. **Lot Size** (área do terreno) — a área de terreno do imóvel, tipicamente em pés quadrados ou acres.
6. **Longitude**, **Latitude** e, se disponível, **Area ID**. Essas variáveis de localização ajudam o modelo a levar em conta a variação geográfica de preços.

2.3 Convenções de nomenclatura das colunas especiais

O glmnetUI identifica as colunas pela **designação de tipo especial**, e não pelo nome da coluna. Você pode dar às suas colunas qualquer nome que queira na exportação do MLS — o que importa é que você atribua o tipo especial correto na tabela de configuração de variáveis (Capítulo 6).

Por exemplo, seu MLS pode exportar a área construída como “GLA”, “Living SqFt”, “liv_area” ou “gross_living_area”. Depois da importação (na qual ela se torna snake_case), basta designá-la como **living_area** no menu suspenso Special. O glmnetUI então a usará para os cálculos de resíduo por SF, independentemente de seu nome original.

2.4 Qualidade e completude dos dados

- **Valores ausentes (NA)**: as linhas com valores NA em qualquer coluna de preditor ou de resposta são removidas automaticamente antes do ajuste.
- **Colunas de data**: as colunas de caracteres que correspondem a formatos de data comuns são analisadas automaticamente na importação. Os formatos de ano com 2 dígitos têm prioridade quando não se detecta um ano de 4 dígitos.
- **Colunas numéricas**: preço de venda, área construída, tamanho do terreno, concessões e campos semelhantes precisam ser numéricos. Se o seu MLS exportar preços com símbolos de moeda ou separadores de milhar (por exemplo, “\$350,000” ou “350.000,00”), pode ser necessário limpá-los antes da importação.
- **Colunas de fator**: as colunas de texto, TRUE/FALSE e de fator do R são tratadas automaticamente como categóricas — uma regressão só consegue tratá-las como fatores. Uma coluna **numérica** (por exemplo, um código numérico de região ou de estilo) só se torna fator quando você marca sua caixa **Factor**; o glmnetUI nunca infere que algo é categórico a partir da faixa de valores de uma coluna numérica, de modo que preditores numéricos discretos como **bath_count** permanecem contínuos. O papel Special é independente da decisão sobre fator.

***Dica:** examine a coluna NA na tabela de preditores após a importação. Colunas com muitos valores ausentes podem fazer com que linhas sejam descartadas. Considere excluir colunas com muitos NA ou limpar os dados antes da importação.*

2.5 Posição da linha do avaliando

No modo **Appraisal**, a **linha 1** deve ser o **imóvel avaliando**. Todas as demais linhas são vendas comparáveis. A linha do avaliando é excluída do ajuste do modelo. Após o ajuste, o

modelo ainda assim gera previsões para a linha do avaliando.

No modo **Market Area Analysis**, colocar o avaliando na linha 1 é opcional.

No modo **General** não há tratamento especial de linhas — todas as linhas são tratadas de forma idêntica.

3 Modo General Purpose

3.1 Visão geral

O modo General Purpose é o padrão quando você inicia o glmnetUI. Ele oferece o fluxo de trabalho completo de regressão elastic net para qualquer conjunto de dados — não apenas imóveis. Você pode usar o glmnetUI para dados científicos, análise financeira, estudos de engenharia ou qualquer problema de regressão.

No modo General, a interface omite os recursos específicos do setor imobiliário (colunas especiais, sale age, arredondamento de coordenadas, RCA). A barra lateral é enxugada para se concentrar na seleção de variáveis, na configuração de parâmetros, no ajuste do modelo e na exportação.

3.1.1 Skip First Row

Uma caixa de seleção **Skip first row** (pular a primeira linha) aparece no topo da barra lateral assim que um projeto General ou Market é aberto. Quando marcada, a linha 1 é excluída do ajuste do modelo. Isso é útil quando a linha 1 contém uma observação-alvo ou de referência que não deve influenciar o modelo. No modo Appraisal, a linha 1 (o imóvel avaliando) é sempre excluída automaticamente.

3.1.2 Configurações por propósito

As configurações (seleção de preditores, parâmetros do modelo, interações) são salvas separadamente para cada combinação de arquivo de entrada e modo de uso. Alternar entre os modos General, Appraisal e Market preserva as configurações de cada modo de forma independente.

3.2 O fluxo de trabalho da barra lateral

A barra lateral está organizada em seções numeradas e recolhíveis, que o conduzem da seleção do projeto até a exportação do relatório. A maioria das seções aparece somente depois que um projeto ativo é aberto, e várias delas são específicas do propósito do projeto. A pasta de saída é derivada do projeto ativo — não há um campo separado de pasta de saída.

1. Project — crie ou abra um projeto. Um projeto define o propósito e o local de trabalho (país, estado, condado, cidade) e suas pastas de entrada (`<os>_in/`) e de saída. Os projetos são armazenados sob uma raiz `regProj` comum (por padrão `~/regProj`, sobrescrevível pela variável de ambiente `REGPROJ_ROOT`) e são compartilhados com os aplicativos irmãos `earthUI` e `mgcvUI`, de modo que os modelos dos três convivem lado a lado. Clique em **Close Project** para trocar de projeto.

2. Import Data — selecione um arquivo CSV ou Excel da pasta de entrada do projeto ativo. Adicione arquivos à pasta pelo Finder/Explorer e clique em **Refresh** para listá-los. Para arquivos Excel com várias planilhas, aparece um seletor de planilha. Os nomes das colunas são automaticamente convertidos para `snake_case`.

3. Import from earthUI (opcional) — importe um arquivo `.rds` de resultado do `earthUI` para usar as funções de base dobradiça do `earth` com a regularização elastic net do `glmnet`. O botão de navegação tem o mesmo estilo do seletor de arquivos da Seção 2. Esta etapa é opcional — pule-a se você não tiver um resultado do `earthUI`.

4. Variable Configuration — seletor da variável resposta, tabela de preditores com caixas de seleção para Include, Factor e Force, além dos menus suspensos Sign. A coluna Special (para designar datas de contrato, coordenadas etc.) aparece apenas nos modos Appraisal e Market. Veja o Capítulo 6 para todos os detalhes.

5. glmnet Call Parameters — toda a configuração do modelo: alpha, lambda, family, standardize, restrições de sinal, relaxed lasso, semente aleatória, matriz de interações e parâmetros avançados (lambda.min.ratio, nlambda, métrica de perda da validação cruzada, limiar de convergência, número máximo de iterações, intercepto). Veja o Capítulo 7 para a referência completa dos parâmetros.

6. Fit Glmnet Model — o botão que executa o modelo. Os resultados aparecem nas abas à direita.

7. Download Output — exporta previsões, resíduos, escores CQA e contribuições por variável como arquivo Excel para a pasta de saída do projeto. Disponível em todos os modos de uso depois que o modelo é ajustado.

8. Calculate RCA Adjustments & Download — calcula os ajustes de RCA por variável para cada comparável e interpola o resíduo do avaliando por meio do CQA. Aparece somente nos modos Appraisal e Market, após o ajuste. Veja o Capítulo 11.

9. Generate Sales Grid & Download — gera a grade intermediária de comparação de vendas, ordenando os comparáveis pelo percentual de ajuste bruto. Aparece somente no modo Appraisal, depois que os ajustes de RCA são calculados. Veja o Capítulo 12.

10. Generate Quarto Report — grava um pacote de relatório Quarto (.qmd) autocontido — com todos os gráficos e tabelas — na pasta de saída do projeto. Disponível assim que o modelo é ajustado.

11. Convert Quarto Report — renderiza o pacote .qmd gerado em HTML, Word ou PDF. O PDF exige uma instalação LaTeX; se nenhuma for detectada, a opção de PDF fica oculta. Veja o Capítulo 13.

Nos modos Market e General, as etapas ocultas exclusivas da avaliação são puladas e as duas seções de relatório são renumeradas em conformidade.

3.3 Projetos

O glmnetUI organiza o trabalho em **projetos** sob uma pasta raiz **regProj** comum (a mesma árvore usada pelo earthUI e pelo mgcvUI). Cada projeto possui uma pasta de entrada (<os>_in) e pastas de saída por mecanismo (por exemplo, <os>_out_glmnet).

Para criar um projeto, clique em **+ New...** na Seção 1 e informe:

- **Purpose** — General, Appraisal ou Market Area.
- **Location** — a hierarquia político-administrativa do projeto. O conjunto de níveis que você preenche depende do país escolhido, de modo que as convenções de nomenclatura estrangeiras são automaticamente contempladas — um projeto dos EUA usa estado/condado/cidade, um projeto francês usa região/département/commune, um projeto japonês usa prefeitura/cidade-ou-distrito/bairro, e assim por diante.
- **Project Name** — um rótulo curto, limitado a 8 caracteres (letras, dígitos, _ e -).

O aplicativo não usa literalmente o nome digitado: ele constrói um nome de pasta único e datado, antepondo os códigos de localização e a data-hora de criação, na forma <country>_<admin level codes>_<YYYYMMDD-HHMMSS>_<your name> — por exemplo, um projeto dos EUA chamado burl torna-se **us_ca_081_burlin_20260604-131500_burl**. O limite de 8 caracteres aplica-se somente à parte que você digita. Apenas os projetos recém-criados recebem o prefixo de data-hora; os projetos já existentes em disco mantêm seus nomes originais e continuam plenamente acessíveis.

3.4 Importação do earthUI (opcional)

A Seção 3 da barra lateral — **Import from earthUI** — permite importar um arquivo .rds de resultado do earthUI. Isso permite que o glmnetUI use as funções de base dobradiça do earth (termos lineares por partes) com a regularização elastic net do glmnet, combinando a detecção

automática de não linearidade do earth com a seleção de variáveis e as restrições de coeficientes do glmnet.

Quando um resultado do earthUI é importado, são os preditores do modelo earth — e não as caixas Include — que orientam o ajuste do glmnet, porque o glmnet ajusta sobre a expansão de base do earth. A tabela Variable Configuration deixa isso explícito: **Include** e **Type** ficam bloqueados, as linhas dos preditores que o modelo earth não utilizou ficam desabilitadas e uma nota explica que o conjunto de preditores vem do modelo earth. **Force** e **Sign** permanecem editáveis para os preditores do earth, pois continuam a se aplicar à base do earth no momento do ajuste.

O botão de navegação tem o mesmo estilo do seletor de arquivos da Seção 2. Esta etapa é inteiramente opcional — se você não tiver um arquivo de resultado do earthUI, basta pular a Seção 3 e seguir para a Seção 4.

3.5 Abas do painel principal

Após o ajuste, o painel principal oferece as seguintes abas:

Aba	Conteúdo
Data Preview	Prévia dos dados importados. No modo de avaliação, dividida nas tabelas Subject Property e Comparable Sales.
Equation	A equação do modelo ajustado, exibida em LaTeX/MathJax, mostrando cada coeficiente não nulo e os termos de interação.
Correlation	Mapa de calor das correlações entre preditores numéricos (disponível antes do ajuste).
Summary	Métricas principais (R^2 , R^2 aj., GR^2 , R^2 de VC, RMSE, MAE) e métricas de avaliação (COD, PRD, razão mediana).
Coefficients	Tabela de coeficientes não nulos, com avisos de violação de sinal.
Variable Importance	Gráfico de barras e tabela ordenada da importância dos preditores ($ \beta \times sd(x)$). Interatividade com plotly ao passar o cursor, quando disponível.
Contributions	Gráficos de efeito parcial por variável, com rótulos de inclinação. Dispersão + reta de ajuste para preditores numéricos, box plots para fatores, superfície 3D/dispersão/mapa de calor para interações.
ANOVA	Tabela de decomposição da variância: SQ por variável, % da SQ do modelo, coeficiente.
Diagnostics	Quatro gráficos: trajetória dos coeficientes, erro de validação cruzada, observado vs. previsto e resíduos vs. ajustados.
Glmnet Output	Saída bruta do modelo: semente aleatória utilizada, impressão do modelo, lambda selecionado, lambda.min/1se, gamma (se relaxado) e vetor completo de coeficientes.
Report	Campos de configuração e exportação do relatório.
RCA Adjustments	Resultados da análise RCA (modo de avaliação, após o cálculo).

3.6 Persistência das configurações

O glmnetUI salva automaticamente sua configuração no armazenamento local do navegador, indexada pelo nome do arquivo de entrada. Quando você recarrega o mesmo arquivo, todas as configurações são restauradas: seleção da variável resposta, caixas dos preditores, tipos de dados, sinais esperados, parâmetros do glmnet e matriz de interações. O último modo de uso empregado também é persistido globalmente e restaurado quando o aplicativo é reiniciado. Há três opções disponíveis:

- **Use last settings for input file** — restaura as configurações salvas para este arquivo específico
- **Use default settings** — aplica os padrões globais que você salvou
- **glmnet defaults** — redefine para os padrões de fábrica ($\alpha = 1$, λ_{1se} por validação cruzada, 10 dobras, gaussian, standardize ativado)

Um botão **Save current as default** salva todas as configurações atuais como padrão global.

3.7 Modo escuro

Clique no ícone de lua/sol no canto superior direito para alternar entre os temas claro e escuro. A preferência de tema é salva no armazenamento local e persiste entre sessões. Todos os elementos da interface (tabelas, gráficos, cartões, botões) se adaptam ao tema selecionado por meio de variáveis CSS.

3.8 Localidade e configurações regionais

O glmnetUI oferece suporte a convenções internacionais de formatação numérica e de CSV por meio de um sistema de localidade baseado no país. O menu suspenso **Settings**, na barra de título, oferece os seletores **Country** e **Paper** para 31 países suportados. Cada predefinição configura:

- **Separador do CSV** — vírgula (,) para EUA/Reino Unido/Japão ou ponto e vírgula (;) para a maior parte da Europa, onde a vírgula é usada como separador decimal.
- **Separador decimal** — ponto (.) ou vírgula (,).
- **Separador de milhar** — vírgula (EUA/Reino Unido/Japão), ponto (Alemanha/Itália/Espanha), espaço (Finlândia/França/Polônia/Bálticos/Ucrânia/Rússia) ou apóstrofo (Suíça).
- **Tamanho do papel** — Letter (EUA/Canadá/México) ou A4 (em todos os demais casos).

3.8.1 Salvar padrões

Clique em **Save as my default** para armazenar globalmente suas preferências de localidade. Esses padrões se aplicam a todas as sessões futuras, independentemente do arquivo de dados carregado.

***Dica:** a formatação numérica nos eixos dos gráficos e nos rótulos de inclinação se adapta automaticamente à sua localidade. A localidade alemã usa pontos para os milhares (200.000), a finlandesa usa espaços (200 000) e a suíça usa apóstrofos (200'000). Nenhum símbolo monetário é exibido — o glmnetUI é agnóstico quanto à moeda.*

4 Modo Appraisal

Quando você seleciona **For Appraisal** como propósito, o glmnetUI se configura para a avaliação de um único imóvel. Todos os recursos descritos no Capítulo 3 permanecem disponíveis; este capítulo trata apenas dos acréscimos específicos da avaliação.

4.1 Tratamento da linha do avaliando

No modo de avaliação, **a linha 1 do seu conjunto de dados é o imóvel avaliando** e todas as demais linhas são vendas comparáveis. Seu arquivo de entrada precisa estar organizado de acordo (veja o Capítulo 2). O preço de venda do avaliando pode ser deixado em branco ou receber qualquer valor — o glmnetUI o trata automaticamente como NA durante o ajuste.

Após a importação, a aba Data Preview se divide em duas seções: **Subject Property** (linha 1) e **Comparable Sales** (linhas 2 em diante). A linha 1 é sempre excluída do ajuste do modelo. Após o ajuste, o modelo ainda assim gera previsões para a linha do avaliando.

4.2 Data-base e sale age

Nos modos de avaliação e de mercado, um campo **Effective Date** (data-base) aparece na seção Variable Configuration (com a data de hoje como padrão). Se você designar uma coluna como **contract_date** no menu suspenso da coluna Special, o glmnetUI calcula uma coluna **sale_age** — o número inteiro de dias entre a data do contrato de cada venda e a data-base. Essa coluna é acrescentada como preditor.

Quando a data-base muda, **sale_age** é recalculada automaticamente.

4.3 Designações especiais de colunas

Nos modos de avaliação e de mercado, um menu suspenso **Special** aparece para cada preditor na tabela Variable Configuration. Veja o Capítulo 6 para a referência completa dos tipos especiais e de seus efeitos.

4.4 Visão geral dos ajustes de RCA

O botão **Calculate RCA Adjustments & Download** (seção 8 da barra lateral, visível nos modos Appraisal e Market após o ajuste) calcula ajustes derivados do mercado para cada comparável em relação ao avaliando. O fluxo de trabalho completo do RCA é descrito no Capítulo 11.

Dica: o *escore CQA* que você atribui ao avaliando controla quanto da distribuição dos resíduos é atribuído a ele. Um *escore* de 5,00 coloca o avaliando na mediana dos comparáveis.

5 Modo Market Area Analysis

Quando você seleciona **Market Area Analysis** como propósito, o glmnetUI oferece os mesmos recursos específicos do setor imobiliário do modo de avaliação (colunas especiais, sale age, arredondamento de coordenadas), mas está orientado à análise de um grupo de imóveis, e não à avaliação de um único avaliando.

5.1 Diferenças em relação ao modo Appraisal

- **O tratamento da linha do avaliando é flexível** — no modo de mercado, a linha 1 não é automaticamente tratada como imóvel avaliando.
- **Sem grade de comparação de vendas** — a etapa “Calculate RCA Adjustments & Download” (Seção 8) está disponível no modo de mercado, mas a etapa Sales Grid (Seção 9) é exclusiva do modo de avaliação.

5.2 Quando usar o modo Market

O modo Market Area Analysis é apropriado quando você está:

- Construindo um modelo de regressão para um bairro ou região de mercado a fim de compreender os fatores determinantes de valor
- Analisando como variáveis como área, idade, tamanho do terreno e localização afetam os preços de venda em um grupo de imóveis
- Preparando o embasamento para uma análise de condições de mercado ou para a delimitação de um bairro
- Trabalhando com um conjunto de dados no qual você quer os recursos de coluna especial (sale age, arredondamento de coordenadas), mas não precisa do fluxo de trabalho de ajustes do RCA

Dica: o modo Market também é útil para regressão imobiliária de caráter geral, quando você quer os recursos de coluna especial mas não precisa do fluxo de trabalho de ajustes do RCA.

6 Seleção de variáveis

A Seção 4 da barra lateral — **Variable Configuration** — é onde você escolhe quais colunas participam do modelo e como elas são tratadas.

6.1 Variável resposta

O menu suspenso **Target (response) variable**, no topo da Seção 4, lista todas as colunas do seu conjunto de dados. Selecione uma coluna como variável resposta (por exemplo, o preço de venda). A coluna resposta é automaticamente excluída da lista de preditores.

6.2 A tabela de preditores

Abaixo do seletor da variável resposta, uma tabela lista todas as colunas restantes com os seguintes campos:

Coluna	Descrição
Variable	Nome da coluna
Type	Menu suspenso de tipo de dado: numeric , integer , character , Date , POSIXct
Include	Caixa de seleção — inclui esta coluna como preditor no modelo
Factor	Caixa de seleção — trata a variável como fator categórico (cria colunas dummy na matriz do modelo)
Force	Caixa de seleção — força a entrada no modelo (fator de penalização = 0, nunca descartada pelo lasso)
Special	Menu suspenso (somente avaliação/mercado) — veja a referência de tipos especiais de coluna adiante
Sign	Sinal esperado do coeficiente: positivo, negativo ou qualquer um
NAs	Contagem de valores ausentes

6.3 Detecção de tipos de dados e sobreposições

O glmnetUI detecta automaticamente os tipos de dados na importação. Colunas numéricas, inteiras e de data são reconhecidas. As colunas de caracteres que se parecem com datas (padrões comuns de formato de data) são classificadas como **Date**.

Você pode sobrepor qualquer detecção alterando o menu suspenso **Type**. Mudar os tipos afeta a forma como a coluna é codificada na matriz do modelo.

As colunas de texto, TRUE/FALSE e de fator do R são tratadas automaticamente como categóricas. Uma coluna numérica só se torna fator quando você marca sua caixa **Factor** — o glmnetUI nunca infere que algo é categórico a partir da faixa de valores de uma coluna numérica, de modo que preditores numéricos discretos, como uma contagem de banheiros, permanecem contínuos, a menos que você indique o contrário. A configuração Factor persiste via localStorage.

6.4 Sinais esperados e sua imposição

Para cada preditor, o menu suspenso **Sign** especifica se você espera que seu coeficiente seja positivo, negativo ou de qualquer direção. Após o ajuste:

- Os coeficientes com sinais inesperados são **sinalizados na tabela Coefficients**
- Os avisos de sinal aparecem na aba Summary

Quando a caixa **Enforce Sign Constraints** está habilitada nos parâmetros do glmnet (Capítulo 7), o modelo é forçado a produzir coeficientes compatíveis com os sinais esperados. Isso é implementado por meio de `upper.limits` e `lower.limits` no glmnet.

6.5 Referência dos tipos especiais de coluna

Nos modos de avaliação e de mercado, o menu suspenso **Special** oferece as seguintes opções:

Ponderação:

- `weight` — coluna de peso das observações (disponível em todos os modos; apenas uma é permitida; as linhas com peso = 0 são excluídas do ajuste)

Tipos de data e hora:

- `contract_date` — dispara o cálculo automático de `sale_age` a partir da data-base
- `listing_date` — usada como alternativa para calcular os dias em mercado
- `dom` — identifica a coluna de dias em mercado
- `sale_age` — designa uma coluna de sale age já existente, de modo que qualquer coluna pode servir como sale age, em vez de o glmnetUI calcular uma a partir de `contract_date`

Tipos monetários:

- `concessions` — identifica as concessões na venda (créditos do vendedor, incentivos ao comprador etc.)

Tipos de transação:

- `sale_type` — identifica a coluna de tipo de venda/transação

Tipos de tamanho e localização:

- `latitude` — valores automaticamente arredondados para 3 casas decimais
- `longitude` — mesmo tratamento de arredondamento da latitude
- `area` — identificador de região de mercado ou bairro (por exemplo, `area_id`, `neighborhood`, `market_area`)
- `living_area` — habilita os cálculos de resíduo por pé quadrado (`residual_sf` e `cqa_sf`)
- `lot_size` — coluna de tamanho do terreno
- `site_dimensions` — agrupada com o tamanho do terreno (por exemplo, “75x120”)

Tipos de idade:

- `actual_age` — coluna de idade do imóvel
- `effective_age` — idade efetiva do imóvel

Tipos de exibição:

- `display_only` — a coluna é incluída nas exportações para Excel, mas totalmente excluída do ajuste do modelo. Use isso para campos de endereço, números de MLS, identificadores de matrícula ou outros dados de referência. Várias colunas podem ter essa designação.

Dica: as colunas designadas como *display_only* permanecem no seu conjunto de dados e aparecem na exportação para Excel, mas são inteiramente excluídas da tabela de preditores. Use isso para colunas de identificação, endereços ou outros campos de referência.

7 Seleção de parâmetros

A Seção 5 da barra lateral — **glmnet Call Parameters** — dá acesso a todas as opções de configuração do modelo elastic net. Cada parâmetro tem um ícone azul de ajuda (?) com uma explicação em dica de tela.

7.1 Alpha (parâmetro de mistura)

O parâmetro alpha controla o tipo de regularização:

Valor de alpha	Tipo	Comportamento
0	Ridge	Contraí todos os coeficientes proporcionalmente; nunca leva nenhum a zero
1	Lasso	Pode levar coeficientes exatamente a zero (seleção de variáveis)
0 a 1	Elastic Net	Combinação de ridge e lasso

Estão disponíveis dois métodos de seleção de alpha:

- **Fixed** — define um único valor de alpha por meio do controle deslizante (padrão: 1,0)
- **Grid Search** — testa vários valores de alpha e seleciona aquele com o menor erro de validação cruzada. Configure a faixa e o número de valores.

7.2 Seleção de lambda

Lambda controla a intensidade da regularização:

Método	Descrição
Cross-validation	Recomendado. Testa muitos valores de lambda usando validação cruzada de k dobras.
Manual	Informe um valor específico de lambda, se você tiver um motivo para isso.

Ao usar a validação cruzada, há duas escolhas de lambda disponíveis:

- **lambda.1se** (padrão) — o maior lambda dentro de 1 erro-padrão do mínimo. Produz modelos mais simples e mais regularizados. Recomendado para trabalhos de avaliação.
- **lambda.min** — o lambda que minimiza o erro de validação cruzada. Produz o modelo de melhor ajuste, porém potencialmente mais complexo.

O parâmetro **Number of CV Folds** controla como os dados são divididos para a validação cruzada (padrão: 10 dobras).

7.3 Family

Escolha a família de distribuição da sua variável resposta:

Family	Caso de uso
gaussian	Respostas contínuas (por exemplo, preço de venda). A mais comum.

Family	Caso de uso
binomial	Desfechos binários (por exemplo, vendido/não vendido).
poisson	Dados de contagem (por exemplo, número de vendas).

7.4 Standardize

Quando marcada (padrão), todos os preditores são escalonados para ter média 0 e desvio padrão 1 antes do ajuste. Isso garante que a penalização trate todos os preditores igualmente, independentemente de sua escala original. Os coeficientes são devolvidos na escala original. Em geral, deve ser mantida ativada.

7.5 Restrições de sinal

Quando **Enforce Sign Constraints** está habilitada, os coeficientes são restringidos para corresponder aos sinais esperados definidos na tabela de variáveis (Capítulo 6):

- Um sinal “positivo” força o coeficiente a ser ≥ 0
- Um sinal “negativo” força o coeficiente a ser ≤ 0
- As variáveis definidas como “either” (qualquer um) ficam sem restrição

Isso é implementado por meio dos parâmetros `upper.limits` e `lower.limits` em `glmnet()`.

7.6 Relaxed Lasso

O lasso comum usa a mesma penalização para a seleção de variáveis E para a estimação dos coeficientes, o que pode contrair demais coeficientes importantes em direção a zero. O **relaxed lasso** separa essas etapas:

1. Primeiro, selecionam-se as variáveis por lasso (a penalização determina quais coeficientes sobrevivem)
2. Em seguida, reajustam-se as variáveis sobreviventes com penalização menor ou nula, o que produz estimativas menos enviesadas

O controle deslizante **Gamma** regula o grau de relaxamento:

Gamma	Efeito
0	Totalmente relaxado (reajuste por MQO, sem nenhuma contração)
1	Sem relaxamento (igual ao <code>glmnet</code> comum)
Intermediário	Combina os ajustes relaxado e penalizado

Com a validação cruzada, gamma é escolhido automaticamente para obter o desempenho ótimo.

7.7 Semente aleatória

Um campo de texto **Random seed**, pré-preenchido com um inteiro aleatório, torna reprodutível a atribuição das dobras da validação cruzada — `set.seed()` é chamada antes do ajuste, e a semente utilizada é mostrada na aba `Glmnet Output` e na mensagem de status do ajuste. Após cada ajuste, uma nova semente é gerada automaticamente (usando o relógio do sistema, e não o gerador de números aleatórios do R). As últimas 5 sementes usadas para o arquivo atual são mostradas como links clicáveis, para fácil recuperação; o histórico de sementes é reiniciado quando um novo arquivo é carregado.

7.8 Matriz de interações

Uma matriz triangular superior interativa permite controlar quais pares de variáveis podem interagir. Cada célula contém uma caixa de seleção — marcada significa que o termo de interação $x1:x2$ é incluído na matriz do modelo.

- **Allow All** — marca todos os pares de interação
- **Clear All** — desmarca todos os pares (o padrão para um arquivo novo é tudo desmarcado)
- **Clique no nome de uma variável** — alterna todas as interações daquela variável

As seleções de interação são salvas no localStorage do navegador para cada arquivo de entrada e restauradas quando você recarrega o mesmo arquivo.

Quando qualquer interação está habilitada, o relatório exportado inclui uma matriz **Enabled Interactions** (preditores nos dois eixos, com uma marca de verificação onde o termo produto $x1:x2$ foi incluído), renderizada em paisagem e espelhando a matriz Allowed Interactions do earthUI. Observe que o lasso ainda pode ter contraído a zero o coeficiente de um termo habilitado.

***Dica:** com muitos preditores, o lasso zera automaticamente os termos de interação desnecessários. Começar com "Allow All" e deixar o modelo selecionar é uma abordagem razoável.*

Aviso

Habilitar interações no glmnet produzirá ajustes que podem ser difíceis de interpretar e de explicar em juízo ou em auditoria. Use interações com cautela e verifique se os ajustes resultantes são razoáveis.

7.9 Bloquear o efeito principal (somente interação)

Clique com o botão direito no nome de uma variável na matriz de interações para impedi-la de entrar no modelo como efeito principal. Um indicador “1” em negrito aparece após o nome da variável. Quando bloqueada:

- A variável é **removida** dos termos principais da fórmula
- A variável **ainda pode** aparecer em termos de interação ($x1:x2$)
- Clique novamente com o botão direito para desbloquear
- O estado persiste via localStorage para cada arquivo de entrada

Um caso de uso típico: ao modelar ajustes de tempo que devem escalonar com o tamanho do imóvel, bloqueie `sale_age` como efeito principal e permita-a apenas em uma interação com `living_area`. Isso cria um termo de interação `sale_age:living_area` em vez de um coeficiente `sale_age` constante — a taxa de valorização no tempo varia conforme o tamanho da casa.

7.10 Parâmetros avançados

Uma seção recolhível “Advanced”, no final de glmnet Call Parameters, expõe configurações adicionais. Elas têm padrões sensatos, mas ficam visíveis para que todas as configurações do modelo possam ser documentadas para fins judiciais ou de auditoria.

- **Lambda Min Ratio** — razão entre o menor e o maior lambda da trajetória de regularização. Padrão: 0,00001. Para conjuntos de dados pequenos (menos de 50 observações), um valor maior (por exemplo, 0,01) pode melhorar o desempenho.
- **Number of Lambda Values** — número de valores de lambda na trajetória. Padrão: 100.
- **CV Loss Metric** — função de perda da validação cruzada: MSE (padrão para gaussian), MAE (robusta a valores atípicos) ou Deviance.

- **Convergence Threshold** — limiar de convergência da descida por coordenadas. Padrão: $1e-07$.
- **Max Iterations** — número máximo de iterações da descida por coordenadas. Padrão: 100.000. Aumente-o se aparecerem avisos de convergência.
- **Fit Intercept** — inclui um termo de intercepto (constante). Padrão: ativado. O intercepto é o “valor-base” no RCA de avaliação.

7.11 Padrões das configurações

Três opções controlam como os parâmetros são inicializados quando um arquivo é carregado:

- **Use last settings for input file** — restaura exatamente as configurações usadas da última vez com este arquivo
- **Use default settings** — aplica os padrões globais que você salvou
- **glmnet defaults** — redefine para as configurações de fábrica ($\alpha = 1$, λ_{1se} por validação cruzada, 10 dobras, gaussian, standardize ativado)

Clique em **Save current as default** para armazenar todas as configurações atuais como padrão global para os arquivos futuros.

8 Ajuste do modelo

8.1 O botão de ajuste

A Seção 6 da barra lateral contém o botão **Fit Glmnet Model**. Clicar nele executa o modelo com a sua configuração atual.

8.2 O que acontece durante o ajuste

Quando você clica em Fit, o glmnetUI:

1. **Prepara os dados** — remove as linhas com NA, aplica os pesos (se designados), exclui a linha do avaliando (no modo de avaliação) e codifica as variáveis de fator como colunas dummy.
2. **Constrói a matriz do modelo** — cria a matriz de delineamento, incluindo quaisquer termos de interação permitidos na matriz de interações.
3. **Aplica as restrições** — se a imposição de sinal estiver habilitada, define `upper.limits` e `lower.limits` para os coeficientes.
4. **Executa a validação cruzada** — se a VC estiver selecionada, chama `cv.glmnet()` para encontrar o lambda ótimo. Se o relaxed lasso estiver habilitado, chama `cv.glmnet(..., relax = TRUE)`.
5. **Ajusta o modelo final** — armazena o modelo ajustado, os coeficientes, as previsões e os resíduos.
6. **Atualiza todas as abas de resultados** — Summary, Coefficients, Equation, Variable Importance, Contributions, ANOVA e Diagnostics são preenchidas.

Uma mensagem de status abaixo do botão Fit mostra o número de observações, as linhas excluídas e o lambda selecionado.

9 Abas de resultados

9.1 Data Preview

Mostra os dados importados como uma DataTable interativa. Nos modos de avaliação/mercado, a prévia é dividida em duas tabelas: **Subject Property** (linha 1) e **Comparable Sales** (linhas 2 em diante).

Cada célula é mantida em uma única linha e truncada na largura da coluna, de modo que campos largos de texto livre (como observações sobre o imóvel) não estiquem uma linha por toda a página. **Dê um duplo clique em qualquer célula** para abrir uma janela pop-up mostrando seu conteúdo completo, sem truncamento.

9.2 Equation

Exibe a equação do modelo ajustado, renderizada em LaTeX via MathJax. Mostra todos os coeficientes não nulos com formatação matemática adequada:

- Coeficientes positivos exibidos com sinal +
- Coeficientes negativos exibidos com sinal −
- Termos de interação exibidos como $x_1 \times x_2$
- Números formatados com separadores e casas decimais apropriadas

9.3 Correlation

Uma matriz em mapa de calor mostrando as correlações de Pearson entre todos os preditores numéricos e a variável resposta. Disponível imediatamente após a importação dos dados — não requer o ajuste do modelo.

- Gradiente de cores: azul (−1) → branco (0) → vermelho (+1)
- Valores de correlação impressos em cada célula
- O tamanho do texto e dos eixos se adapta ao número de variáveis
- Útil para identificar multicolinearidade antes do ajuste

9.4 Summary

Estatísticas de ajuste do modelo, exibidas em cartões:

- **R²** — proporção da variância explicada
- **R² ajustado** — ajustado pelo número de preditores
- **GR²** — R² generalizado, a partir da razão de deviance do glmnet no lambda selecionado
- **R² de VC** — R² de validação cruzada (quando se usa VC)
- **RMSE** — raiz do erro quadrático médio
- **MAE** — erro absoluto médio

Nos modos de avaliação/mercado, cartões adicionais mostram:

- **Median Ratio** — mediana de (previsto / observado)
- **COD** — coeficiente de dispersão (quanto menor, melhor)
- **PRD** — diferencial relacionado ao preço

Um **aviso de sobreajuste** aparece quando o R² de treinamento excede o R² de VC em mais de 0,1.

Abaixo dos cartões, uma tabela de coeficientes com avisos de sinal duplica a aba Coefficients, por conveniência.

9.5 Coefficients

Uma tabela de todos os coeficientes não nulos, mostrando:

- Nome da variável
- Valor do coeficiente
- Sinal esperado (definido na configuração de variáveis)
- Marcador de aviso de sinal, quando a direção do coeficiente não corresponde às expectativas

9.6 Variable Importance

Magnitude padronizada do coeficiente: $|\beta| \times \text{sd}(x)$ para cada preditor, agregada entre as colunas dummy no caso de variáveis de fator. Dois modos de exibição:

- **plotly interativo** (quando o pacote `plotly` está instalado): gráfico de barras horizontais com dicas de tela que mostram a importância exata, o coeficiente e o percentual relativo.
- **ggplot2 estático** (alternativa): gráfico de barras horizontais com rótulos de eixo formatados com separadores.

Abaixo do gráfico, uma `DataTable` mostra Variable, Importance, Coefficient e Relative %.

9.7 Contributions

Gráficos de efeito parcial por variável, mostrando a contribuição de cada preditor às previsões do modelo. Selecione uma variável no menu suspenso.

Para preditores numéricos: um gráfico de dispersão dos valores de x da variável vs. sua contribuição ($\beta \times x$), com um segmento de reta vermelho e um rótulo de inclinação. O rótulo de inclinação mostra o efeito marginal por unidade (por exemplo, $+1,234.56/\text{unit}$), com unidades adaptativas para variáveis de faixa estreita, como latitude/longitude (por exemplo, $+0.12/0.001$). O subtítulo mostra o valor do intercepto (base).

Para variáveis de fator: um box plot mostrando a distribuição dos valores de contribuição por nível do fator.

Para interações de duas variáveis: três visualizações empilhadas — um gráfico de superfície 3D (`plotly` interativo quando disponível; caso contrário, `persp()` estático) da contribuição conjunta, um gráfico de dispersão com os pontos coloridos pela contribuição e um mapa de calor da contribuição média ao longo das faixas das duas variáveis.

9.8 ANOVA

Tabela de decomposição da variância, mostrando para cada preditor:

- **SQ** — soma de quadrados: $\sum(\text{contribution}_j - \overline{\text{contribution}_j})^2$
- **% da SQ do modelo** — percentual da soma de quadrados total do modelo
- **Coefficient** — o coeficiente dominante daquele preditor

Inclui uma linha de intercepto e uma linha TOTAL MODEL. Os termos de interação (que contêm “:”) são apresentados como grupos separados.

9.9 Diagnostics

Quatro gráficos de diagnóstico com fontes grandes (tamanho base de 16 pt), 15 marcas nos eixos e rótulos formatados com separadores (sem notação científica):

1. **Coefficient Path** — mostra como todos os coeficientes mudam à medida que $\log(\lambda)$ aumenta. As linhas são coloridas por variável. Ajuda a visualizar a trajetória de regularização e a seleção de variáveis.

2. **CV Error** — erro de validação cruzada em função de $\log(\lambda)$. Os pontos mostram o erro médio de VC; as barras de erro mostram ± 1 erro-padrão. As linhas verticais tracejadas marcam $\lambda_{\min}$ (azul) e λ_{1se} (verde).
3. **Actual vs Predicted** — gráfico de dispersão dos valores observados vs. previstos. Os pontos devem se agrupar em torno da reta de referência tracejada de 45 graus. Uma dispersão maior indica mais erro de previsão.
4. **Residuals vs Fitted** — gráfico de dispersão dos resíduos vs. valores ajustados. Deve mostrar dispersão aleatória em torno da linha do zero (tracejada). Padrões sugerem má especificação do modelo (por exemplo, não linearidade, heterocedasticidade).

Cada gráfico tem um botão **Download PNG** para salvá-lo a 150 DPI.

9.10 Glmnet Output

Saída bruta do modelo, mostrando: a semente aleatória utilizada, a impressão do modelo, o λ selecionado, $\lambda_{\min}/\lambda_{1se}$, γ (se o relaxed lasso estiver habilitado) e o vetor completo de coeficientes.

9.11 Report

Configure os campos do relatório (nome do avaliador, endereço do imóvel, data do relatório, número do processo) e exporte para Word (.docx) ou PDF diretamente na aba. Para o fluxo completo de relatórios baseado em Quarto (HTML, Word ou PDF), veja o Capítulo 13.

10 Download dos dados

Após o ajuste, o botão **Download Output (Excel)** (seção 7 da barra lateral, disponível em todos os modos de uso) exporta um arquivo Excel com previsões e diagnósticos para a pasta de saída do projeto. O título da seção é “Download Estimated Sale Prices & Residuals” nos modos de avaliação e de mercado e “Download Estimated Target Variable(s) & Residuals” no modo General.

10.1 Colunas de saída

Coluna	Descrição
<code>est_<target></code>	Previsão do modelo (por exemplo, <code>est_sale_price</code>)
<code>residual</code>	Observado – previsto
<code>cqa</code>	Escore de condição-qualidade-atratividade (escala de 0 a 10)
<code>residual_sf</code>	Resíduo / área construída (se <code>living_area</code> for designada)
<code>cqa_sf</code>	CQA calculado a partir da ordenação por <code>residual_sf</code>
<code><var>_contribution</code>	Contribuição de cada variável à previsão
<code>basis</code>	Valor do intercepto (o mesmo para todas as linhas)

Uma marca de verificação branca aparece no botão de download após a conclusão bem-sucedida.

10.2 Escores CQA

O CQA ordena o resíduo de cada linha em relação a todos os demais, em uma escala de 0 a 10:

- **CQA alto (~9–10):** vendeu por muito mais do que o previsto — provavelmente condição/qualidade/atratividade superiores
- **CQA baixo (~0–1):** vendeu por muito menos do que o previsto — provavelmente condição inferior ou venda forçada
- **CQA ~5:** próximo da mediana

Nos modos de avaliação/mercado, as linhas são ordenadas por `residual_sf` em ordem decrescente quando uma coluna `living_area` é designada.

11 Cálculos e download do RCA

O fluxo de trabalho do RCA (Residual Constraint Approach) está disponível nos modos **Appraisal** e **Market Area Analysis**, após o ajuste do modelo. A grade de comparação de vendas subsequente (Capítulo 12) é exclusiva do modo de avaliação.

11.1 Abertura da caixa de diálogo do RCA

Clique no botão **Calculate RCA Adjustments & Download** na seção 8 da barra lateral. Escolha:

- O tipo de escore: **CQA** ou **CQA per SF**. Se “CQA per SF” for selecionado, o resíduo do avaliando é escalonado pela área construída.
- Informe o escore CQA do avaliando (0,00–9,99; padrão 5,00)

11.2 Interpolação do escore CQA

1. Os escores CQA e os resíduos dos comparáveis são ordenados
2. A interpolação linear mapeia o seu valor de CQA para um resíduo
3. Valor do avaliando = previsão do modelo + resíduo interpolado

11.3 Colunas de saída

Coluna	Descrição
subject_value	Previsão do modelo + resíduo interpolado
<var>_adjustment	Contribuição do avaliando – contribuição do comparável
residual_adjustment	Resíduo do avaliando – resíduo do comparável
net_adjustments	Soma de todos os ajustes
gross_adjustments	Soma dos ajustes em valor absoluto
adjusted_sale_price	Preço de venda do comparável + ajustes líquidos

Uma marca de verificação branca aparece no botão após o cálculo bem-sucedido.

12 Grade de comparação de vendas

A grade de comparação de vendas está disponível somente no modo **Appraisal**, após o cálculo dos ajustes do RCA (Etapa 8). Ela gera um arquivo Excel formatado adequado para inclusão em laudos de avaliação.

12.1 Caixa de diálogo de seleção de comparáveis

Clicar no botão **Generate Sales Grid & Download**, na seção 9 da barra lateral, abre uma caixa de diálogo modal que lista todas as vendas comparáveis da saída do RCA. Os comparáveis são divididos em dois grupos:

- **Recommended comps** (pré-marcados): ajuste bruto < 25% do preço de venda, ordenados por sale age (dos mais recentes aos mais antigos).
- **Other comps** (desmarcados por padrão): os demais comparáveis, ordenados por sale age.

Selecione até 30 comparáveis e clique em **Generate Sales Grid** para criar a pasta de trabalho.

12.2 Estrutura da pasta de trabalho

A pasta de trabalho gerada contém até 10 planilhas, cada uma com 3 comparáveis lado a lado:

- **Linhas de cabeçalho:** endereços do avaliando e dos comparáveis, preços de venda, datas de venda e dias em mercado
- **Linhas agrupadas:** localização (latitude, longitude, área), tamanho do terreno (lot size, site dimensions) e idade (idade real, idade efetiva), com cabeçalhos de grupo e subitens
- **Linhas das variáveis do modelo:** uma linha por coeficiente não nulo do modelo, mostrando a contribuição do avaliando, as contribuições dos comparáveis e os percentuais de ajuste
- **Linhas de características residuais:** células vazias e desbloqueadas para a entrada do avaliador quanto a características não captadas pelo modelo (por exemplo, condição, qualidade, vista)
- **Linhas de resumo:** % de ajuste líquido, % de ajuste bruto e preço de venda ajustado, com fórmulas ativas do Excel

12.3 Proteção das planilhas e células residuais

Cada planilha é protegida por senha para evitar edições acidentais, mas as células de valor das características residuais ficam explicitamente desbloqueadas. Isso permite que os avaliadores insiram valores para características que não estão no modelo, preservando ao mesmo tempo os cálculos de ajuste conduzidos por fórmulas.

13 Download de relatórios

13.1 O fluxo Quarto em duas etapas

A exportação de relatórios pela barra lateral é um fluxo Quarto em duas etapas:

1. **Generate Quarto Report** (seção 10 da barra lateral) — grava um pacote Quarto autocontido (o fonte `.qmd` mais os gráficos, `report_data.rds` e `reference.docx`) na pasta de saída `glmnet` do projeto.
2. **Convert Quarto Report** (seção 11 da barra lateral) — renderiza um arquivo `.qmd` nos formatos de saída escolhidos. O caminho aponta por padrão para o pacote gerado mais recentemente, mas você pode navegar até qualquer `.qmd`.

13.2 Formatos de relatório

Três formatos estão disponíveis na etapa de conversão:

- **HTML** — documento HTML autocontido. Funciona em todas as plataformas sem software adicional além do pandoc. Padrão recomendado.
- **Word (.docx)** — documento formatado. Adequado para edição e distribuição. Funciona em todas as plataformas.
- **PDF** — gerado pelo Quarto com LaTeX. O tamanho do papel segue a configuração de localidade (Letter ou A4). **Exige uma instalação LaTeX** (veja o apêndice de requisitos de sistema). Se nenhum LaTeX for detectado, a opção de PDF é automaticamente ocultada das caixas de seleção de formato.

Os relatórios são gerados com o Quarto quando este está disponível, com recurso alternativo ao rmarkdown. A aba Report oferece ainda botões diretos de Export to Word e Export to PDF.

13.3 Conteúdo dos relatórios

Os relatórios incluem todo o conteúdo das abas:

- Descrição do conjunto de dados (nome do arquivo, observações, variável resposta, preditores, propósito)
- Especificação do modelo (α , λ , family, standardize, relaxed, coeficientes não nulos)
- Resumo dos resultados (R^2 , R^2 ajustado, R^2 generalizado, R^2 de VC, RMSE, MAE)
- Equação do modelo (renderizada em LaTeX)
- Tabela de coeficientes
- Importância das variáveis (gráfico de barras e tabela)
- Gráficos de contribuição por variável, correspondentes à aba Contributions:
 - **Preditores numéricos**: gráfico de dispersão com reta de ajuste linear
 - **Variáveis de fator**: box plot por nível
 - **Interações**: dispersão colorida pela contribuição, mapa de calor da contribuição média e imagem estática da superfície 3D (persp)
- Matriz Enabled Interactions (quando alguma interação está habilitada) — preditores nos dois eixos, com uma marca de verificação onde o termo produto `x1:x2` foi incluído, renderizada em paisagem
- Mapa de calor da matriz de correlação
- Tabela de decomposição ANOVA
- Gráficos de diagnóstico:
 - Trajetória dos coeficientes

- Erro de validação cruzada
- Observado vs. previsto
- Resíduos vs. ajustados
- Gráfico Q-Q normal
- Saída do modelo

Todos os rótulos de eixo e legendas de cor usam números com separadores (sem notação científica).

Uma marca de verificação branca aparece no botão do relatório após a geração bem-sucedida, e um cronômetro é exibido enquanto o relatório está sendo renderizado.

14 Comparação com o earthUI

O aplicativo earth e seu Post Processing Module glmnet formam um único conjunto de ferramentas para modelagem de regressão. Eles compartilham o mesmo formato de dados, os mesmos tipos especiais de coluna, o mesmo fluxo de trabalho do RCA e os mesmos conjuntos de dados de demonstração, mas usam mecanismos de modelagem diferentes.

14.1 Principais diferenças

Característica	glmnetUI	earthUI
Método	Elastic net (glmnet)	MARS/Earth (earth)
Seleção de variáveis	Automática, pela penalização lasso	Automática, por poda para frente/para trás
Relações entre coeficientes	Somente efeitos lineares	Lineares por partes (funções dobradiça)
Interações	Aos pares, pela matriz de interações	Até grau 3
Sinais dos coeficientes	Impostos por restrições	Não impostos diretamente
Ajuste relaxado	Relaxed lasso disponível	Não se aplica
Sintonia de alpha	Busca em grade disponível	Não se aplica (sem parâmetro de mistura)
Grade de comparação de vendas	Disponível (Etapa 9)	Disponível (Etapa 8)
Gráficos de funções g	Não se aplica	Visualização de termos agrupados
Formatos de relatório	HTML, Word, PDF	HTML, Word, PDF

14.2 Recursos compartilhados

As duas ferramentas oferecem:

- Importação de CSV e Excel com conversão dos nomes de coluna para snake_case
- Suporte a localidades de 31 países, com separador de CSV, separador decimal e tamanho do papel
- Designações especiais de colunas (contract_date, living_area, area, latitude, longitude etc.)
- Cálculo automático de sale_age a partir da data-base
- Alternância entre modo claro e escuro
- Persistência das configurações via localStorage
- Mapa de calor de correlação, importância das variáveis, contribuições, ANOVA, diagnósticos
- Escores CQA e fluxo de trabalho de ajustes do RCA
- Grade de comparação de vendas, com seleção de comparáveis e saída em Excel formatada
- Compatibilidade com os mesmos conjuntos de dados de demonstração (Appraisal_1.csv)

14.3 Quando usar cada um

- Use o **glmnetUI** quando quiser um modelo linear simples e interpretável, com seleção automática de variáveis. O elastic net é particularmente adequado quando há muitos

preditores e se deseja que o modelo selecione os mais importantes.

- **Use o earthUI** quando esperar relações não lineares (por exemplo, o efeito da área construída sobre o preço mudando conforme o tamanho) ou quando precisar de funções dobradiça para captar efeitos de limiar. Os modelos earth também captam interações de ordem superior (graus 2 e 3). O earthUI oferece adicionalmente um recurso experimental de extração de observações baseado em Prolog/DCG, que não tem equivalente no glmnetUI.
- **Use os dois** para comparar modelos lineares e não lineares sobre os mesmos dados. Se ambos produzirem valores de R^2 semelhantes, o modelo glmnet, mais simples, pode ser preferível pela interpretabilidade.

15 Conjunto de dados de demonstração: Appraisal_1.csv

15.1 Descrição

O conjunto de dados de demonstração acompanha o glmnetUI (e é compartilhado com o earthUI). Localize-o com:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "glmnetUI")
```

O arquivo contém 1.502 vendas residenciais (mais 1 imóvel avaliando na linha 1) de uma exportação simulada de MLS. Os dados representam vendas de casas unifamiliares em um mercado com várias regiões, abrangendo uma faixa de tamanhos, idades e localizações de imóveis.

Não se trata de dados reais, mas eles se baseiam em um bairro realista do Norte da Califórnia. Todas as informações de identificação foram alteradas ou removidas.

15.2 Colunas

Coluna	Tipo	Tipo especial	Descrição
weight	numeric	weight	Peso da observação (0 = excluir do ajuste)
id	numeric	display_only	Identificador interno do registro
property_id	numeric	display_only	Identificador do imóvel no MLS
listing_id	character	display_only	Número do anúncio no MLS
parcel_number	character	display_only	Número da inscrição cadastral (APN)
street_address	character	display_only	Endereço do imóvel
city_name	character	display_only	Cidade
postal_code	character	display_only	Código postal
county_name	character	display_only	Condado
contract_date	Date	contract_date	Data do contrato de venda (calcula sale_age)
sale_age	numeric	—	Dias da data do contrato até a data-base
sale_price	numeric	(resposta)	Preço de venda — variável resposta
living_sqft	numeric	living_area	Área construída bruta em pés quadrados
beds_total	integer	—	Número de dormitórios
baths_total	numeric	—	Número total de banheiros (por exemplo, 2,5 = 2 completos + 1 lavabo)
lot_size	numeric	lot_size	Área do terreno em pés quadrados
area_id	integer	area	Identificador de região do MLS
age	numeric	actual_age	Idade do imóvel em anos
latitude	numeric	latitude	Latitude (arredondada para 3 casas decimais no modelo)
longitude	numeric	longitude	Longitude (arredondada para 3 casas decimais no modelo)
garage_spaces	integer	—	Número de vagas de garagem
days_on_market	integer	dom	Dias em mercado
listing_date	Date	listing_date	Data do anúncio
sale_concessions	numeric	concessions	Concessões do vendedor

15.3 Início rápido sugerido

1. Inicie a rotina glmnet: `earthUI::launch_glmnet()`

2. Crie um novo projeto (Seção 1) com o propósito **For Appraisal**
3. Coloque **Appraisal_1.csv** na pasta de entrada do projeto e selecione-o na Seção 2 (clique em **Refresh**, se necessário)
4. Selecione **sale_price** como variável resposta
5. Atribua os tipos especiais conforme mostrado na tabela acima
6. Inclua os preditores: **sale_age**, **living_sqft**, **baths_total**, **lot_size**, **area_id** (como fator), **age**, **latitude**, **longitude**, **garage_spaces**
7. Mantenha $\alpha = 1$ (Lasso) e validação cruzada com $\lambda.1se$
8. Clique em **Fit Glmnet Model**
9. Examine as abas Summary, Coefficients e Diagnostics
10. Baixe a saída (Etapa 7) e examine a ordenação por CQA
11. Calcule os ajustes do RCA (Etapa 8) com um escore CQA de ~5,00
12. Gere uma grade de comparação de vendas (Etapa 9) com os comparáveis recomendados

16 Requisitos de sistema e solução de problemas

16.1 Plataformas suportadas

O glmnetUI é executado em **macOS**, **Windows** e **Linux**. O RStudio Desktop (2023.06 ou posterior) é fortemente recomendado — ele já traz o **pandoc**, necessário para os relatórios.

16.2 Dependências opcionais

- **LaTeX** (TinyTeX, MiKTeX ou MacTeX): somente para a geração de relatórios em PDF. Se não for detectado, a opção de PDF é automaticamente ocultada. Instale com: `tinytex::install_tinytex()`
- **sysfonts + showtext**: fonte Roboto Condensed para os gráficos. Recorre automaticamente à fonte sem serifa do sistema se estiver indisponível ou offline.
- **earth** ($\geq 5.3.0$): necessário apenas para o pipeline de importação do earthUI.

16.3 Notas por plataforma

macOS: o que apresenta menos problemas. Instale o TinyTeX para o PDF: `tinytex::install_tinytex()`

Windows: funciona bem com o RStudio. Máquinas corporativas ou com restrições podem ter limitações no diretório temporário — o aplicativo emite um aviso no console, mas continua a funcionar sem persistência das configurações.

Linux: pode ser necessário instalar bibliotecas do sistema: `sudo apt install libcurl4-openssl-dev libssl-dev libxml2-dev libsqlite3-dev libfontconfig1-dev`

16.4 Degradação graciosa

Componente ausente	Comportamento
Sem LaTeX	Opção de PDF oculta. HTML e Word continuam disponíveis.
Sem internet / falha nas fontes	Usa-se a fonte sem serifa do sistema. Mensagem registrada no console.
Sem RSQLite / sistema de arquivos somente leitura	As configurações não persistem. O aplicativo funciona normalmente.
Diretório temporário sem permissão de escrita	A geração do relatório falha com um erro claro. Defina TMPDIR para um local com permissão de escrita.

16.5 Solução de problemas

“PDF option not available”: execute `tinytex::install_tinytex()` no R e reinicie o aplicativo.

“Settings will not persist”: verifique as permissões do diretório de dados do usuário (macOS: `~/Library/Application Support/R/glmnetUI/`, Linux: `~/.local/share/R/glmnetUI/`, Windows: `%APPDATA%/R/data/glmnetUI/`).

Porta 7879 já em uso: execute `lsof -ti:7879 | xargs kill` (macOS/Linux) ou `Stop-Process -Id (Get-NetTCPConnection -LocalPort 7879).OwningProcess` (Windows PowerShell).

Referências

- Friedman, Jerome, Trevor Hastie e Robert Tibshirani (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent”. Em: *Journal of Statistical Software* 33.1, pp. 1–22. DOI: [10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).
- Simon, Noah et al. (2011). “Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent”. Em: *Journal of Statistical Software* 39.5, pp. 1–13. DOI: [10.18637/jss.v039.i05](https://doi.org/10.18637/jss.v039.i05).
- Tay, J. Kenneth, Balasubramanian Narasimhan e Trevor Hastie (2023). “Elastic Net Regularization Paths for All Generalized Linear Models”. Em: *Journal of Statistical Software* 106.1, pp. 1–31. DOI: [10.18637/jss.v106.i01](https://doi.org/10.18637/jss.v106.i01).

mgcvUI: uma interface interativa para o pacote mgcv (modelos aditivos generalizados)

Tradução para o português do Brasil, elaborada com assistência de IA e revisada pela redação; a

versão em inglês é a autorizada.

William Bert Craytor

Volume 1, Número 1 · Verão de 2026

Publicado em 1 de julho de 2026

DOI: [10.67330/xtv1pa25](https://doi.org/10.67330/xtv1pa25)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artigo de acesso aberto é distribuído sob os termos da [licença Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Resumo. mgcvUI é uma interface gráfica de usuário para o pacote de R mgcv, que ajusta modelos aditivos generalizados (GAM) por meio de splines de regressão penalizada com seleção automática do grau de suavização. Oferece três modos de uso — modelagem preditiva geral, avaliação de imóveis e análise de região de mercado — e conduz o usuário pela importação de dados, pela especificação de termos suaves, pelo ajuste do modelo, pelos gráficos de diagnóstico e de efeitos e pelos relatórios passíveis de download. Este artigo documenta os requisitos de formato de dados do mgcvUI, o fluxo de trabalho de modelagem, as telas de saída e a referência completa de funcionalidades.

Palavras-chave: modelos aditivos generalizados, GAM, splines penalizadas, suavização, mgcv, R, interface gráfica de usuário, avaliação de imóveis

Nota do editor (julho de 2026): Pouco antes da publicação, o mgcvUI foi consolidado dentro do earthUI (versão 0.11.0 e posteriores), que agora oferece o fluxo de trabalho de GAM aqui descrito ao lado dos fluxos de trabalho de MARS e de elastic net. O mgcvUI continua disponível em sua forma final publicada (0.3.0) em <https://github.com/wcraytor/mgcvUI>; o desenvolvimento futuro prossegue no earthUI. Este artigo se refere ao mgcvUI tal como foi publicado.

Sumário

1	Introdução	4
1.1	Histórico e antecedentes	4
1.1.1	Modelos aditivos generalizados	4
1.1.2	O pacote mgcv	4
1.1.3	Comparação com MARS (earth) e elastic net (glmnet)	5
1.1.4	O fluxo earthUI-mgcvUI	5
1.1.5	Fluxo de trabalho recomendado	6
1.2	O que é o mgcvUI?	6
1.3	Três modos de uso	6
1.4	Recursos específicos do setor imobiliário	7
1.5	Primeiros passos	7
2	Requisitos dos dados de entrada do MLS	9
2.1	Formato e estrutura do arquivo	9
2.2	Colunas necessárias no modo de avaliação	9
2.3	Convenções especiais de nomenclatura de colunas	10
2.4	Qualidade e completude dos dados	10
2.5	Posicionamento da linha do avaliando	10

3	Modo General	11
3.1	Visão geral	11
3.2	O fluxo de trabalho da barra lateral	11
3.3	Importar do earthUI (opcional)	12
3.4	Abas do painel principal	12
3.5	Persistência das configurações	13
3.6	Modo escuro	13
3.7	Localidade e configurações regionais	13
4	Modo Appraisal	15
4.1	Tratamento da linha do avaliando	15
4.2	Data de referência e idade da venda	15
4.3	Designações especiais de colunas	15
4.4	Visão geral dos ajustes de RCA	15
5	Modo Market Area Analysis	16
5.1	Diferenças em relação ao modo Appraisal	16
5.2	Quando usar o modo de mercado	16
6	Seleção de variáveis	17
6.1	Variável alvo	17
6.2	Transformação da resposta	17
6.3	A tabela de preditores	17
6.4	Detecção de tipos de dados e substituições	18
6.5	Referência dos tipos especiais de coluna	18
7	Seleção de parâmetros	19
7.1	Predefinições de parâmetros	19
7.2	Family	19
7.3	Method	19
7.4	Gamma (multiplicador de suavização)	20
7.5	Validação cruzada	20
7.6	Select (encolhimento)	20
7.7	Base padrão	20
7.8	k padrão (dimensão da base)	20
7.9	Tipo de tensor	21
7.10	Interações permitidas	21
7.11	Interações fator por termo suave	21
7.12	Parâmetros avançados	21
8	Ajuste do modelo	23
8.1	O botão de ajuste	23
8.2	O que acontece durante o ajuste	23
8.3	Ajuste em segundo plano e progresso	23
9	Abas de resultados	24
9.1	Data	24
9.2	Equation	24
9.3	Summary	24
9.4	Variable Importance	24
9.5	Gráficos de contribuição	25
9.5.1	Termos suaves univariados — $s(x)$	25

9.5.2	Termos fator por suave — $s(x, \text{by} = \text{factor})$	25
9.5.3	Termos de interação — $ti(x_1, x_2)$ e $te(x_1, x_2)$	25
9.5.4	Termos paramétricos — lineares e de fator	25
9.6	Correlation	25
9.7	Diagnostics	26
9.8	RCA Adjustments	26
9.9	Anova	26
9.10	Mgev Output	26
9.11	Sign Check	26
9.12	Concurvity	26
9.13	Basis Check	27
10	Download dos dados	28
10.1	Colunas de saída	28
10.2	Escores de CQA	28
11	Cálculos de RCA e download	29
11.1	Abertura da caixa de diálogo do RCA	29
11.2	Interpolação do escore de CQA	29
11.3	Colunas de saída	29
12	Grade de comparação de vendas	30
12.1	Caixa de diálogo de seleção de comparáveis	30
12.2	Estrutura da pasta de trabalho	30
12.3	Proteção da planilha e células residuais	30
13	Download de relatórios	31
13.1	Etapa 1: gerar o relatório Quarto	31
13.2	Etapa 2: converter o relatório Quarto	31
13.3	Conteúdo do relatório	31
14	Exportação das funções do modelo	33
14.1	Funções em R	33
14.2	Funções em Python	33
14.3	Cabeçalho C++	33
14.4	Banco de dados SQLite	33
15	Comparação com earthUI e glmnetUI	34
15.1	Principais diferenças	34
15.2	Recursos compartilhados	34
15.3	Fluxo de trabalho recomendado	34
16	Conjunto de dados de demonstração: Appraisal_1.csv	36
16.1	Descrição	36
16.2	Colunas	36
16.3	Roteiro rápido sugerido	37
17	Requisitos de sistema e solução de problemas	38
17.1	Plataformas suportadas	38
17.2	Versão do R	38
17.3	Pacotes de R necessários	38
17.4	Dependências opcionais	39
17.5	Observações por plataforma	39

17.5.1	macOS	39
17.5.2	Windows	39
17.5.3	Linux	39
17.6	Degradação graciosa	40
17.7	Solução de problemas	40
17.7.1	“PDF option is not available”	40
17.7.2	“Settings will not persist”	40
17.7.3	“Diagnostics plot unavailable”	41
17.7.4	“Report error: number of items to replace...”	41
17.7.5	Porta 7880 já em uso	41
17.7.6	Falha na instalação de pacotes no Linux	41

1 Introdução

1.1 Histórico e antecedentes

1.1.1 Modelos aditivos generalizados

Os **modelos aditivos generalizados** (GAM) foram formalizados por Trevor Hastie e Robert Tibshirani em sua monografia de 1990. Os GAM estendem o modelo linear substituindo cada termo linear $\beta_j x_j$ por uma função suave $f_j(x_j)$, de modo que o modelo fica assim:

$$y = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p) + \varepsilon$$

Cada f_j é estimada a partir dos dados por meio de uma spline de regressão penalizada. As funções suaves permitem que o modelo capte relações não lineares sem que o usuário precise especificar de antemão uma forma funcional. A penalização controla a ondulação de cada termo suave, evitando o sobreajuste sem deixar de permitir a flexibilidade necessária para acompanhar os padrões reais.

1.1.2 O pacote mgcv

O pacote de R `mgcv` (Mixed GAM Computation Vehicle) foi desenvolvido por Simon Wood na Universidade de Bath (Wood 2017). Lançado pela primeira vez em 2001, é hoje a implementação padrão dos GAM em R e acompanha toda instalação de R. Seus métodos subjacentes são descritos em uma série de artigos que abrangem a estimação estável de múltiplos parâmetros de suavização (Wood 2004), as splines de regressão thin-plate (Wood 2003), a estimação REML rápida e estável (Wood 2011) e a seleção geral de modelos suaves (Wood, Pya e Säfken 2016). Entre seus recursos principais estão:

- **Seleção automática do grau de suavização:** o parâmetro de suavização de cada termo é estimado como parte do ajuste do modelo, por meio de REML (máxima verossimilhança restrita), GCV (validação cruzada generalizada) ou ML (máxima verossimilhança).
- **Múltiplos tipos de base:** splines de regressão thin-plate (tp), splines de regressão cúbica (cr), P-splines (ps) e B-splines (bs), cada uma com propriedades distintas.
- **Termos suaves de produto tensorial:** `te()` e `ti()` permitem interações suaves entre variáveis sem incorrer na maldição da dimensionalidade.
- **Seleção de variáveis:** a opção `select = TRUE` acrescenta uma penalização adicional que pode encolher termos suaves até zero, realizando de fato uma seleção de variáveis.
- **Diagnósticos de concurvidade:** medem o análogo da multicolinearidade para os termos suaves.

1.1.3 Comparação com MARS (earth) e elastic net (glmnet)

O aplicativo earth e seus dois módulos de pós-processamento (glmnet e mgcv) usam motores de modelagem diferentes. A tabela a seguir resume as principais diferenças:

Característica	mgcv (GAM)	earth (MARS)	glmnet (elastic net)
Tipo de modelo	Não linear suave (splines penalizadas)	Linear por partes (splines adaptativas)	Linear (com regularização)
Não linearidade	Automática, por funções suaves (thin plate, cúbica etc.)	Automática, por funções de dobradiça com nós determinados pelos dados	Somente por meio de interações ou expansão de base
Seleção de variáveis	Opcional, por encolhimento de dupla penalização (<code>select=TRUE</code>)	Nativa, por passos progressivos/regressivos com poda por GCV	Nativa, pela penalização L_1 (Lasso)
Interações	Termos suaves de produto tensorial (<code>te()</code> , <code>ti()</code>)	Produtos baseados em dobradiças (interpretáveis, até grau 3)	Produtos cruzados aos pares (difíceis de interpretar)
Interpretabilidade	Curvas suaves de efeito parcial; muito intuitivas	As funções <code>g</code> mostram efeitos parciais lineares por partes para cada variável	Os coeficientes são pesos lineares; fáceis de explicar individualmente
Mais indicado para	Relações não lineares suaves, modelagem flexível	Deteção automática de não linearidade e de interações	Dados de alta dimensão, seleção de variáveis, amostras pequenas
Ajustes de RCA	Ajustes suaves por efeitos parciais	Ajustes lineares por partes por meio das funções <code>g</code>	Ajustes lineares por variável
Controle do sobreajuste	Estimação do parâmetro de suavização por REML/GCV	Poda baseada em GCV	Lambda validado por validação cruzada

O **MARS** (Multivariate Adaptive Regression Splines) foi apresentado por Jerome Friedman em 1991. Ele constrói modelos lineares por partes selecionando de forma adaptativa funções de dobradiça e as posições de seus nós a partir dos dados. A implementação em R é o pacote **earth**, de Stephen Milborrow.

A regressão **elastic net**, implementada no pacote **glmnet** por Friedman, Hastie e Tibshirani (2010), combina as penalizações Lasso (L_1) e ridge (L_2) para realizar simultaneamente a seleção de variáveis e o encolhimento dos coeficientes.

1.1.4 O fluxo earthUI–mgcvUI

Um fluxo de trabalho poderoso combina a descoberta automática de nós do earthUI com a estimação suave do mgcvUI. Quando o earthUI exporta um arquivo de resultados `.rds` e o mgcvUI o importa:

1. **O earth descobre os nós** — o MARS encontra posições de dobradiça (pontos de mudança) determinadas pelos dados para cada variável.

2. **O mgcvUI as usa como âncoras** — os nós do earth passam a ser a base para as splines de regressão cúbica (base `cr`) no GAM, dando aos termos suaves uma vantagem inicial, com posições de nós que refletem a estrutura real dos dados.
3. **O mgcv refina o ajuste** — a estimação por splines penalizadas suaviza o ajuste linear por partes do earth, produzindo curvas suaves de efeito parcial e preservando a forma geral descoberta pelo earth.

Esse fluxo faz a ponte entre a modelagem exploratória (earth) e a modelagem confirmatória (GAM), reunindo o melhor das duas abordagens.

1.1.5 Fluxo de trabalho recomendado

Para avaliação de imóveis e aplicações semelhantes que exigem modelos interpretáveis e defensáveis:

1. **Comece pelo earthUI** para descobrir as variáveis importantes, as relações não lineares e as interações presentes em seus dados. A seleção automática de base e as funções `g` do earth fornecem excelentes indicações iniciais.
2. **Refine com o mgcvUI** se quiser efeitos parciais mais suaves. Importe os nós do earthUI para o mgcvUI e faça uma transição contínua de modelos lineares por partes para modelos suaves.
3. **Use o glmnetUI** quando precisar de uma seleção de variáveis agressiva (Lasso), quando o número de preditores se aproximar ou ultrapassar o número de observações, ou quando quiser um modelo puramente linear com regularização, em nome da defensabilidade.

1.2 O que é o mgcvUI?

O mgcvUI é uma interface gráfica de usuário para o pacote de R `mgcv`. Ele é executado como um aplicativo Shiny local — não há login, não há servidor e não há contas. Você o inicia a partir do R, importa um conjunto de dados (CSV ou Excel), configura seu modelo e ajusta um GAM de forma interativa.

Na versão atual, o mgcvUI já não é um pacote instalado separadamente: ele é distribuído *dentro* do pacote earthUI, como seu módulo de pós-processamento de GAM, iniciado com `earthUI::launch_mgcv()`. O aplicativo em si — e tudo o que se descreve neste artigo — permanece inalterado; apenas a instalação e o comando de início são diferentes. As três rotinas (earth, glmnet, mgcv) compartilham uma mesma árvore de projetos, um mesmo banco de dados de configurações e uma mesma cópia da infraestrutura de apoio.

O aplicativo oferece um fluxo de trabalho completo: importação de dados, configuração de variáveis com especificação de termos suaves, ajuste do modelo com processamento em segundo plano, gráficos de diagnóstico, curvas suaves de efeito parcial e relatórios passíveis de download em formato Word, PDF ou HTML.

Os **modelos aditivos generalizados** substituem a relação linear βx por uma função suave $f(x)$ para cada preditor. Isso significa:

- **Não é preciso especificar formas funcionais** — os dados determinam a forma de cada relação.
- **Efeitos parciais suaves** — a contribuição de cada preditor é uma curva suave, e não uma reta, o que torna fácil ver como uma variável afeta a resposta em diferentes valores.
- **Seleção automática do grau de suavização** — o mgcv determina o quanto cada curva deve ondular, por REML ou GCV, equilibrando ajuste e sobreajuste.

1.3 Três modos de uso

Todo projeto tem um **purpose** (propósito ou modo de uso), escolhido ao criá-lo na Seção 1 da barra lateral (**Project**) e fixado a partir de então. O propósito do projeto ativo determina quais

ferramentas e elementos da interface aparecem; para trabalhar em outro modo, crie ou abra um projeto com esse propósito. Os três modos são:

- **General** (geral) — regressão GAM para qualquer tipo de população ou conjunto de dados. Esse é o modo padrão. Oferece o fluxo de trabalho completo de GAM sem acréscimos específicos de nenhum domínio.
- **For Appraisal** (para avaliação) — regressão GAM adaptada à avaliação de imóveis. Acrescenta recursos específicos para a avaliação de um único imóvel, incluindo o tratamento do imóvel avaliando, as designações especiais de colunas e o Residual Constraint Approach (RCA).
- **Market Area Analysis** (análise de região de mercado) — regressão GAM adaptada aos estudos de região de mercado. Acrescenta recursos para analisar grupos de imóveis dentro de um mercado definido, com tratamento opcional da linha do avaliando.

Nos três modos o motor de modelagem subjacente é idêntico — você sempre ajusta um GAM por meio de `mgcv::gam()`. A definição do propósito controla quais ferramentas e elementos de interface adicionais ficam disponíveis.

1.4 Recursos específicos do setor imobiliário

Quando se seleciona **For Appraisal** ou **Market Area Analysis**, o `mgcvUI` ativa vários recursos concebidos para a análise imobiliária:

- **Designações especiais de colunas** — cada preditor pode ser marcado com um papel especial, como `contract_date`, `dom`, `concessions`, `latitude`, `longitude`, `living_area`, `lot_size`, `actual_age`, `effective_age`, `area`, `site_dimensions` ou `display_only`. Essas designações controlam como a coluna é tratada durante o ajuste e nas saídas.
- **Coluna Sale Age** — quando uma coluna é designada como `contract_date` e é informada uma data de referência, o `mgcvUI` calcula uma coluna `sale_age` (dias entre a data da venda e a data de referência) e a substitui como preditor.
- **Cálculos de RCA (somente em avaliação)** — no modo de avaliação, após ajustar o modelo, o `mgcvUI` pode calcular a saída do Residual Constraint Approach (RCA), que produz ajustes por comparável, resumos de ajuste líquido/bruto e um preço de venda ajustado para o imóvel avaliando.

***Dica:** O modo General funciona com qualquer conjunto de dados — financeiro, científico, de engenharia ou imobiliário. Os modos de avaliação e de mercado apenas acrescentam recursos de conveniência para os profissionais do setor imobiliário.*

1.5 Primeiros passos

Para usar o `mgcvUI`:

1. **Instale o earthUI** no R: `install.packages("earthUI")` — a rotina `mgcv` está incluída. (Veja “Installing earthUI and vProlog” no artigo sobre o `earthUI` nesta edição para a versão de desenvolvimento e os componentes opcionais.)
2. **Inicie o aplicativo:** execute `earthUI::launch_mgcv()` no R. O aplicativo abre em seu navegador na porta 7880.
3. **Crie ou abra um projeto** na Seção 1 da barra lateral. Um projeto define o propósito (General, For Appraisal ou Market Area Analysis) e o local de trabalho (país, estado, município, cidade), além de suas pastas de entrada e saída. Os projetos são armazenados sob uma raiz `regProj` compartilhada e são compartilhados com os aplicativos irmãos `earthUI` e `glmnetUI`. Abrir um projeto com outro propósito é a forma de alternar entre os modos.

4. **Importe seus dados** na Seção 2, selecionando um arquivo CSV ou Excel da pasta de entrada do projeto ativo. Acrescente arquivos a essa pasta e clique em **Refresh** para listá-los.
5. **Importe do earthUI (opcional)** — a Seção 3 permite importar um arquivo `.rds` de resultados do earthUI para semear o GAM com os nós descobertos pelo earth. Pule esta etapa se você não tiver um resultado do earthUI.
6. **Configure as variáveis** — escolha o alvo e os preditores, defina os tipos de dados, atribua eventuais papéis especiais de coluna e marque as variáveis como fator ou linear.
7. **Defina os parâmetros do mgcv** — família, método, gamma, tipo de base, k, tipo de tensor, interações e demais opções.
8. **Ajuste o modelo** — clique em “Fit Mgcvc GAM Model” e examine os resultados no painel principal.
9. **Exporte** — baixe as predições em Excel, gere e converta um relatório Quarto ou (no modo de avaliação) calcule os ajustes de RCA e uma grade de vendas.

As configurações são automaticamente persistidas em um banco de dados SQLite e restauradas quando você recarrega o mesmo arquivo de entrada.

2 Requisitos dos dados de entrada do MLS

Para os fluxos de trabalho de avaliação de imóveis e de análise de mercado, seus dados de entrada normalmente vêm de uma exportação de um Multiple Listing Service (MLS). Este capítulo descreve a estrutura de arquivo esperada e as colunas que o mgcvUI é capaz de utilizar.

2.1 Formato e estrutura do arquivo

O mgcvUI aceita arquivos **CSV** e **Excel** (.xlsx, .xls). Na importação, os nomes das colunas são automaticamente convertidos para **snake_case** — por exemplo, “Living SqFt” passa a **living_sqft**, “Contract Date” passa a **contract_date** e “Sale Price” passa a **sale_price**. Essa normalização garante referências de coluna consistentes ao longo de todo o fluxo de trabalho. O separador de CSV e o marcador decimal usados na importação são determinados pelas configurações de localidade (veja o Capítulo 3, “Localidade e configurações regionais”).

Seu arquivo de dados deve ser uma tabela plana, com uma linha por imóvel e uma coluna por atributo. A primeira linha do arquivo deve conter os cabeçalhos das colunas.

2.2 Colunas necessárias no modo de avaliação

Embora o mgcvUI funcione com qualquer conjunto de colunas, o fluxo de trabalho completo de avaliação se beneficia da presença das seguintes colunas:

Coluna	Tipo especial	Finalidade
Sale Price	(alvo)	Variável resposta do modelo
Contract Date	contract_date	Usada para calcular sale_age (dias desde a data de referência)
Listing Date	listing_date	Usada com a data do contrato para calcular o DOM caso não exista coluna de DOM
Days on Market	dom	Dias no mercado; exibida nas exportações
Concessions	concessions	Concessões na venda; SP líquido = preço de venda – concessões
Living Area (SF)	living_area	Habilita resíduos por SF (residual_sf , cqa_sf)
Lot Size	lot_size	Coluna de tamanho do terreno
Site Dimensions	site_dimensions	Agrupada com o tamanho do terreno
Latitude	latitude	Arredondada para 3 casas decimais
Longitude	longitude	Arredondada para 3 casas decimais
Area ID	area	Identificador de região de mercado / bairro
Actual Age	actual_age	Idade do imóvel
Effective Age	effective_age	Idade efetiva do imóvel
Address	display_only	Exibida nas exportações; excluída do modelo

Os nomes das colunas da planilha podem estar em outro idioma — os nomes “especiais” estão em inglês para que o programa em R possa lhes dar tratamento especial. No mais, os nomes de coluna informados aparecem nos modelos de regressão, nos gráficos e, em trabalhos de avaliação, nos relatórios de saída.

Nem todas as colunas são obrigatórias. O mgcvUI se adapta — se uma coluna estiver ausente, o recurso correspondente é simplesmente omitido. Ainda assim, para modelos de precificação imobiliária, certas colunas são altamente recomendáveis:

1. **Sale Age** — o número de dias entre a data do contrato de venda e a data de referência. Quando se usa um histórico de vendas de vários anos, **sale_age** costuma desempenhar papel central.

2. **Living Area** — também aparece sob nomes como “Living Sqft” ou “GLA” (gross living area, área bruta habitável).
3. **Total Bath Count** — o número total de banheiros completos, de um quarto, de meio banheiro e de três quartos.
4. **Garage Bays** ou **Garage Area** — vagas de garagem ou área em pés quadrados.
5. **Lot Size** — a área de terreno do imóvel.
6. **Longitude**, **Latitude** e, se disponível, **Area ID**, para a variação geográfica de preços.

2.3 Convenções especiais de nomenclatura de colunas

O mgcvUI identifica as colunas pela **designação de tipo especial**, e não pelo nome da coluna. Você pode dar às suas colunas os nomes que quiser na exportação do MLS — o que importa é atribuir o tipo especial correto na tabela de configuração de variáveis (Capítulo 6).

2.4 Qualidade e completude dos dados

- **Valores ausentes (NA)**: linhas com valores NA em qualquer coluna de preditor ou de alvo são removidas automaticamente antes do ajuste. O mgcvUI restringe os dados apenas às colunas usadas no modelo, de modo que NAs em colunas não utilizadas não causam a exclusão de linhas.
- **Colunas numéricas**: preço de venda, área habitável, tamanho do terreno e campos semelhantes devem ser numéricos.
- **Colunas de fator**: colunas de texto, TRUE/FALSE e de fator do R são tratadas automaticamente como categóricas — só podem entrar no modelo como fatores. Uma coluna **numérica** (por exemplo, um código numérico de região ou de estilo) só se torna fator quando você marca sua caixa de seleção **Factor**; o mgcvUI nunca infere natureza categórica a partir da faixa de valores de uma coluna numérica, de modo que preditores numéricos discretos como `bath_count` permanecem contínuos (suavizados). O papel especial (Special) é independente da decisão sobre fator.

***Dica:** Examine a coluna NA na tabela de preditores após a importação. Colunas com muitos valores ausentes podem provocar a exclusão de linhas. Colunas com $\geq 50\%$ de NAs são marcadas em vermelho e automaticamente excluídas.*

2.5 Posicionamento da linha do avaliando

No modo **Appraisal**, a **linha 1 deve ser o imóvel avaliando**. Todas as demais linhas são vendas comparáveis. A linha do avaliando é excluída do ajuste do modelo. Após o ajuste, o modelo ainda assim gera previsões para a linha do avaliando.

No modo **Market Area Analysis**, colocar o avaliando na linha 1 é opcional.

No modo **General** não há tratamento especial de linhas — todas as linhas são tratadas igualmente.

3 Modo General

3.1 Visão geral

O modo General é o padrão quando você inicia o mgcvUI. Ele oferece o fluxo de trabalho completo de GAM para qualquer conjunto de dados — não apenas imobiliário. Você pode usar o mgcvUI para dados científicos, análise financeira, estudos de engenharia ou qualquer problema de regressão em que se esperem relações não lineares suaves.

No modo General, a interface omite os recursos específicos do setor imobiliário (colunas especiais, idade da venda, RCA). A barra lateral é simplificada para concentrar-se na seleção de variáveis, na configuração de parâmetros, no ajuste do modelo e na exportação.

3.2 O fluxo de trabalho da barra lateral

A barra lateral é organizada em seções numeradas e recolhíveis que o conduzem da seleção do projeto até a exportação do relatório. A maioria das seções aparece somente depois que há um projeto ativo aberto, e várias delas são específicas do propósito do projeto. A pasta de saída deriva do projeto ativo — não há um campo separado de pasta de saída.

1. Project — crie ou abra um projeto. Um projeto define o propósito e o local de trabalho (país, estado, município, cidade), além de suas pastas de entrada (`<os>_in/`) e de saída. Os projetos são armazenados sob uma raiz `regProj` compartilhada (por padrão `~/regProj`, substituível pela variável de ambiente `REGPROJ_ROOT` ou pelo menu suspenso `Settings`) e são compartilhados com os aplicativos irmãos `earthUI` e `glmnetUI`, de modo que os modelos dos três convivem lado a lado. Clique em **Close Project** para trocar.

2. Import Data — selecione um arquivo CSV ou Excel da pasta de entrada do projeto ativo. Acrescente arquivos à pasta pelo Finder/Explorer e clique em **Refresh** para listá-los. Para arquivos Excel com várias planilhas, aparece um seletor de planilha. Os nomes das colunas são automaticamente convertidos para `snake_case`.

3. Import from earthUI (opcional) — importe um arquivo `.rds` de resultados do `earthUI` para semear o GAM com as posições de nós descobertas pelo `earth` (e para semear as seleções de variáveis). Esta etapa é opcional — pule-a se você não tiver um resultado do `earthUI` para importar.

4. Variable Configuration — seletor da variável alvo, transformação da resposta (`none`, `log`, `log10`) e tabela de preditores com caixas de seleção para `Include`, `Factor` e `Linear`. A coluna `Special` aparece somente nos modos `Appraisal` e `Market`. Veja o Capítulo 6 para todos os detalhes.

5. MgcV Call Parameters — toda a configuração do modelo: predefinições de parâmetros, `family`, `method`, `gamma`, validação cruzada, `select`, tipo de base, `k`, tipo de tensor, matriz de interações e parâmetros avançados. Veja o Capítulo 7 para a referência completa de parâmetros.

6. Fit MgcV GAM Model — o botão que executa o modelo. Os resultados aparecem nas abas à direita.

7. Download Output — exporta predições, resíduos, escores de CQA e contribuições por variável como um arquivo Excel para a pasta de saída do projeto. Disponível em todos os modos de uso, depois que o modelo é ajustado.

8. Calculate RCA Adjustments & Download — calcula os ajustes de RCA por variável para cada comparável e interpola o resíduo do avaliando por meio do CQA. Aparece somente nos modos `Appraisal` e `Market`, depois do ajuste.

9. Generate Sales Grid & Download — gera a grade de comparação de vendas, recomendando comparáveis pelo percentual de ajuste bruto. Aparece somente no modo `Appraisal`.

10. Generate Quarto Report — grava um pacote de relatório Quarto (`.qmd`) autocontido — com todos os gráficos e tabelas — na pasta de saída do projeto. Disponível depois que o modelo é ajustado.

11. Convert Quarto Report — renderiza o pacote `.qmd` gerado em HTML, Word ou PDF. O PDF exige uma instalação de LaTeX; se nenhuma for detectada, a opção de PDF fica oculta. Veja o Capítulo 13.

As etapas específicas de cada propósito são omitidas (e as demais etapas renumeradas) nos outros modos: no modo Market a etapa da grade de vendas está ausente e, no modo General, tanto a etapa de RCA quanto a da grade de vendas estão ausentes.

3.3 Importar do earthUI (opcional)

A Seção 3 da barra lateral — **Import from earthUI** — permite importar um arquivo `.rds` de resultados do earthUI. Isso habilita o fluxo earth-mgcv: as posições de nós determinadas pelos dados no earth tornam-se pontos de ancoragem para os termos suaves do GAM.

Quando um resultado do earthUI é importado, o mgcvUI:

- Exibe um resumo com a variável alvo, o R quadrado, o número de termos e as posições de nós por variável.
- Seleciona automaticamente a predefinição de parâmetros “Earth Pipeline” (base de spline de regressão cúbica, $\gamma = 1.4$).
- Define a variável alvo de modo a coincidir com o alvo do earth.
- Marca previamente Include para as variáveis preditoras do earth.
- Marca previamente Factor para as variáveis categóricas do earth e Linear para os preditores lineares do earth.
- Preenche previamente as interações detectadas pelo earth (destacadas em amarelo na matriz de interações e bloqueadas contra alterações manuais).
- Oferece um botão **Export Knots CSV** para baixar os nós do earth como arquivo CSV.

A tabela Variable Configuration passa então a exibir uma nota explicando que as seleções de variáveis (Include, Linear, Factor) e os nós das splines foram *semeados* a partir do modelo earth. Diferentemente do glmnetUI, em que o ajuste glmnet fica atrelado à base do earth, o mgcvUI ajusta o GAM a partir de suas seleções atuais — o earth apenas dá uma vantagem inicial, e todos os controles permanecem editáveis.

Um botão **Clear** remove os dados importados do earth e restaura o modo autônomo.

3.4 Abas do painel principal

Após a importação dos dados, o painel principal oferece as seguintes abas (as abas que dependem do modelo são preenchidas depois do ajuste):

Aba	Conteúdo
Data	DataTable interativa dos dados importados. Nos modos de avaliação e de mercado, dividida em tabelas de imóvel avaliando e de vendas comparáveis.
Equation	Fórmula completa do GAM renderizada em MathJax, mais uma tabela com as definições das funções suaves.
Summary	Métricas principais (R^2 , R^2 de VC, desviância explicada, AIC, n, número de termos suaves) e tabelas de termos suaves e paramétricos.
Variable Importance	Gráfico de barras horizontais das estatísticas F (termos suaves) e dos valores $ t $ (paramétricos), mais tabela ordenada.
Contribution	Curvas suaves de efeito parcial interativas em plotly, com faixas de confiança e marcadores dos nós do earth.
Correlation	Mapa de calor das correlações entre preditores numéricos (disponível antes do ajuste).
Diagnostics	Resíduos versus ajustados, gráfico Q-Q, histograma, resposta versus ajustados (via gratia). Dispersão de observados versus preditos.
RCA Adjustments	Histogramas da análise de RCA (modo de avaliação/mercado, após o cálculo).
Anova	Tabela de ANOVA combinando termos paramétricos e suaves.
Mgcv Output	Texto bruto: tempos de execução, fórmula, resumo do modelo, família, concurvidade.
Sign Check	Verificação da consistência de direção entre earth e GAM (quando há importação do earth).
Concurvity	Matrizes de concurvidade global e aos pares, com codificação por cores.
Basis Check	Testes de adequação da dimensão da base por <code>gam.check()</code> .

3.5 Persistência das configurações

O mgcvUI salva automaticamente sua configuração em um banco de dados SQLite, indexado pelo nome do arquivo de entrada. Quando você recarrega o mesmo arquivo, todas as configurações são restauradas: seleção do alvo, caixas de seleção dos preditores, tipos de dados, parâmetros do mgcv e matriz de interações. O propósito é uma propriedade do projeto ativo, e não uma configuração global, e o último arquivo de entrada utilizado é lembrado por projeto e carregado automaticamente quando o projeto é reaberto.

3.6 Modo escuro

Clique no ícone de lua/sol no canto superior direito para alternar entre os temas Nord Light e Nord Dark. A preferência de tema é salva no localStorage e persiste entre as sessões. Todos os elementos da interface se adaptam ao tema selecionado.

3.7 Localidade e configurações regionais

O mgcvUI dá suporte a convenções internacionais de formatação de números e de CSV por meio de um sistema de localidades baseado em país. O menu suspenso **Settings** na barra de título oferece os seletores **Country** e **Paper** para mais de 30 países. Cada predefinição configura:

- **Separador de CSV** — vírgula (,) para EUA/Reino Unido/Japão ou ponto e vírgula (;) para a maior parte da Europa.
- **Marcador decimal** — ponto (.) ou vírgula (,).
- **Separador de milhares** — vírgula (EUA/Reino Unido/Japão), ponto (Alemanha/Itália/Espanha), espaço (Finlândia/França/Polônia/países bálticos) ou apóstrofo (Suíça).

- **Tamanho do papel** — Letter (EUA/Canadá/México) ou A4 (nos demais países).

Clique em **Save** para gravar suas preferências de localidade globalmente. A seção Import Data também tem um seletor **Locale** por arquivo, para importações pontuais que usem outras convenções. O mesmo menu suspenso Settings contém o campo **regProj root folder**, que realoca a árvore de projetos compartilhada usada por mgcvUI, earthUI e glmnetUI.

***Dica:** A formatação numérica nos eixos dos gráficos e nos rótulos de inclinação se adapta automaticamente à sua localidade. A localidade alemã usa pontos para os milhares (200.000), a finlandesa usa espaços (200 000) e a suíça usa apóstrofes (200'000). Nenhum símbolo monetário é exibido — o mgcvUI é agnóstico quanto à moeda.*

4 Modo Appraisal

Quando o propósito do projeto ativo é **For Appraisal**, o mgcvUI se configura para a avaliação de um único imóvel. Todos os recursos descritos no Capítulo 3 continuam disponíveis; este capítulo trata apenas dos acréscimos específicos da avaliação.

4.1 Tratamento da linha do avaliando

No modo de avaliação, **a linha 1 do seu conjunto de dados é o imóvel avaliando** e todas as demais linhas são vendas comparáveis. Seu arquivo de entrada deve estar organizado de acordo (veja o Capítulo 2). O preço de venda do avaliando pode ficar em branco ou receber qualquer valor — o mgcvUI automaticamente o trata como NA durante o ajuste.

Após o ajuste, o modelo gera previsões para a linha do avaliando, o que é a base do fluxo de trabalho de ajustes de RCA.

4.2 Data de referência e idade da venda

Nos modos de avaliação e de mercado, um campo **Effective Date** aparece na seção Variable Configuration (com a data de hoje como padrão). Se você designar uma coluna como **contract_date** no menu suspenso Special, o mgcvUI calcula uma coluna **sale_age** — o número inteiro de dias entre a data do contrato de cada venda e a data de referência. Essa coluna é acrescentada como preditor.

Quando a data de referência muda, **sale_age** é recalculada automaticamente.

4.3 Designações especiais de colunas

Nos modos de avaliação e de mercado, um menu suspenso **Special** aparece para cada preditor na tabela Variable Configuration. Veja o Capítulo 6 para a referência completa dos tipos especiais e de seus efeitos.

4.4 Visão geral dos ajustes de RCA

O botão **Calculate RCA Adjustments & Download** (visível nos modos de avaliação e de mercado após o ajuste) calcula ajustes derivados do mercado para cada comparável em relação ao avaliando. O fluxo de trabalho completo do RCA está descrito no Capítulo 11.

***Dica:** O score de CQA que você atribui ao avaliando controla quanto da distribuição de resíduos é atribuído ao avaliando. Um score de 5,00 coloca o avaliando na mediana dos comparáveis.*

5 Modo Market Area Analysis

Quando o propósito do projeto ativo é **Market Area Analysis**, o mgcvUI oferece os mesmos recursos específicos do setor imobiliário que o modo de avaliação (colunas especiais, idade da venda), mas é orientado à análise de um grupo de imóveis, e não à avaliação de um único avaliando.

5.1 Diferenças em relação ao modo Appraisal

- **O tratamento da linha do avaliando é flexível** — no modo de mercado, a linha 1 não é automaticamente tratada como imóvel avaliando.
- **Não há grade de vendas** — a etapa “Generate Sales Grid & Download” não está disponível.

5.2 Quando usar o modo de mercado

O modo Market Area Analysis é apropriado quando você está:

- Construindo um modelo GAM para um bairro ou região de mercado a fim de compreender os determinantes de valor não lineares
- Analisando como variáveis como área construída, idade, tamanho do terreno e localização afetam os preços de venda em um grupo de imóveis
- Preparando fundamentação para uma análise de condições de mercado ou para a delimitação de um bairro
- Trabalhando com um conjunto de dados em que você quer os recursos de colunas especiais (idade da venda), mas não precisa do fluxo de trabalho da grade de vendas

***Dica:** O modo de mercado também é útil para regressões imobiliárias em geral, quando você quer os recursos de colunas especiais mas não precisa do fluxo de trabalho de avaliação de um único imóvel.*

6 Seleção de variáveis

A Seção 4 da barra lateral — **Variable Configuration** — é onde você escolhe quais colunas participam do modelo e como elas são tratadas.

6.1 Variável alvo

O menu suspenso **Target (response) variable**, no alto da Seção 4, lista todas as colunas numéricas do seu conjunto de dados. Selecione uma coluna como variável resposta (por exemplo, o preço de venda). A coluna alvo é automaticamente excluída da lista de preditores.

6.2 Transformação da resposta

Abaixo do seletor de alvo, um menu suspenso **Response Transform** oferece três opções:

- **None (raw)** — usa a resposta como está. Adequado quando a relação entre preditores e resposta é aproximadamente aditiva na escala original.
- **Log (natural)** — aplica a transformação logarítmica natural antes do ajuste. Ao ser retrotransformada, ela torna os ajustes proporcionais (%) em vez de absolutos (\$). É particularmente útil para preços de venda, em que um ajuste de \$10.000 significa coisas muito diferentes em distintos patamares de preço.
- **Log10** — aplica a transformação logarítmica de base 10.

***Dica:** A transformação logarítmica torna os ajustes temporais proporcionais (%) em vez de absolutos (\$). Em modelos imobiliários que abrangem uma ampla faixa de preços, as transformações logarítmicas costumam produzir resíduos mais bem comportados.*

Quando se seleciona uma transformação logarítmica, os valores ≤ 0 na resposta são automaticamente filtrados. Todas as saídas (predições, contribuições, resíduos) são automaticamente retrotransformadas para a escala original.

6.3 A tabela de preditores

Abaixo do seletor de alvo, uma tabela rolável lista todas as colunas restantes. As colunas de caixas de seleção usam cabeçalhos verticais girados, por compacidade. A ordem das colunas é:

Coluna	Descrição
Variable	Nome da coluna (fonte monoespçada, truncado com reticências). Célula com borda.
Type	Menu suspenso de tipo de dado: numeric, integer, character, factor, logical, Date, POSIXct. Menu suspenso com borda.
Include	Caixa de seleção — incluir esta coluna como preditor no modelo. Cabeçalho vertical.
Factor	Caixa de seleção — tratar como fator categórico (cria um coeficiente por nível). Cabeçalho vertical.
Linear	Caixa de seleção — forçar uma relação linear (sem termo suave, apenas βx). Cabeçalho vertical.
Special	Menu suspenso (apenas nos modos de avaliação/mercado) — veja a referência de tipos especiais de coluna adiante. Menu suspenso com borda.
NAs	Contagem e percentual de valores ausentes, com código de cores: vermelho ($\geq 50\%$), laranja ($\geq 20\%$), cinza (> 0). Célula com borda.

Fator vs. suave vs. linear: por padrão, as variáveis numéricas incluídas recebem um termo suave $f(x)$. Marcar **Factor** cria um termo categórico (um coeficiente por nível). Marcar **Linear** cria um termo linear simples βx em vez de um termo suave. Variáveis marcadas como Factor e Linear ao mesmo tempo são tratadas como Factor.

6.4 Detecção de tipos de dados e substituições

O mgcvUI detecta automaticamente os tipos de dados na importação. Colunas numéricas, inteiras, lógicas, de fator e de data são reconhecidas. Você pode substituir qualquer detecção alterando o menu suspenso **Type**. Alterar os tipos afeta o modo como a coluna é tratada no modelo.

6.5 Referência dos tipos especiais de coluna

Nos modos de avaliação e de mercado, o menu suspenso **Special** oferece as seguintes opções:

Ponderação:

- **weight** — coluna de peso das observações (apenas uma é permitida; linhas com peso = 0 são excluídas do ajuste)

Tipos de data e hora:

- **contract_date** — dispara o cálculo automático de **sale_age** a partir da data de referência
- **sale_age** — identifica uma coluna de idade da venda já calculada (dias desde a data de referência), usada como está
- **listing_date** — usada como alternativa para calcular os dias no mercado
- **dom** — identifica a coluna de dias no mercado

Tipo de registro:

- **sale_type** — identifica a coluna de tipo de registro que distingue o avaliando das vendas comparáveis

Tipos monetários:

- **concessions** — identifica as concessões na venda (créditos do vendedor, incentivos ao comprador etc.)

Tipos de tamanho e localização:

- **latitude** — valores automaticamente arredondados para 3 casas decimais
- **longitude** — mesmo tratamento de arredondamento da latitude
- **area** — identificador de região de mercado ou de bairro
- **living_area** — habilita os cálculos de resíduo por pé quadrado (**residual_sf** e **cqa_sf**)
- **lot_size** — coluna de tamanho do terreno
- **site_dimensions** — agrupada com o tamanho do terreno

Tipos de idade:

- **actual_age** — coluna de idade do imóvel
- **effective_age** — idade efetiva do imóvel

Tipos de exibição:

- **display_only** — a coluna é incluída nas exportações para Excel, mas fica inteiramente excluída do ajuste do modelo. Use este tipo para campos de endereço, números de MLS, identificadores de matrícula ou outros dados de referência.

***Dica:** As colunas designadas como **display_only** permanecem em seu conjunto de dados e aparecem na exportação para Excel, mas ficam inteiramente fora da tabela de preditores. Use-as para colunas de identificação, endereços ou outros campos de referência.*

7 Seleção de parâmetros

A Seção 5 da barra lateral — **Mgcv Call Parameters** — dá acesso a todas as opções de configuração do GAM. Cada parâmetro tem um ícone azul de ajuda (?) com uma explicação em dica de tela.

7.1 Predefinições de parâmetros

Um menu suspenso no alto oferece duas predefinições que configuram valores padrão sensatos para fluxos de trabalho comuns:

Predefinição	Base	Configurações
Standalone (descoberta)	tp (thin plate)	k = auto (10), gamma = 1.2, select = TRUE
Earth Pipeline (refinamento)	cr (regressão cúbica)	k = auto (10), gamma = 1.4, select = TRUE

A predefinição **Earth Pipeline** é selecionada automaticamente quando se importam nós do earthUI. A base de spline de regressão cúbica (cr) é necessária para a integração dos nós do earth, porque permite especificar posições exatas de nós.

7.2 Family

Escolha a família de distribuição para sua variável resposta:

Família	Caso de uso
gaussian	Respostas contínuas (por exemplo, preço de venda). A mais comum.
Gamma	Respostas contínuas positivas com variância proporcional à média.
poisson	Dados de contagem (por exemplo, número de vendas).
binomial	Desfechos binários.
inverse.gaussian	Dados positivos com forte assimetria à direita.

7.3 Method

O método de estimação do parâmetro de suavização:

Método	Descrição
REML	Máxima verossimilhança restrita (padrão, recomendado). O mais robusto contra o sobreajuste.
GCV.Cp	Validação cruzada generalizada. Tende a selecionar modelos um pouco mais complexos.
ML	Máxima verossimilhança. Semelhante ao REML, mas pode subsuavizar.
P-REML	Variante REML de Pearson.
P-ML	Variante ML de Pearson.
GACV.Cp	Validação cruzada aproximada generalizada.

7.4 Gamma (multiplicador de suavização)

O parâmetro gamma multiplica os graus de liberdade efetivos no critério de suavização, favorecendo ajustes mais suaves (menos ondulados). O padrão é 1.2.

Valor	Efeito
1.0	Suavização padrão
1.2–1.4	Um pouco mais suave que o padrão (recomendado)
2.0+	Muito mais suave — bom para dados ruidosos ou amostras pequenas

Valores mais altos de gamma protegem contra o sobreajuste ao penalizar a complexidade mais fortemente. A predefinição Earth Pipeline usa 1.4 para obter suavização adicional ao refinar os nós do earth.

7.5 Validação cruzada

Quando marcada (padrão), o mgcvUI executa validação cruzada de 10 dobras após o ajuste para calcular um R quadrado de validação cruzada. Isso fornece uma estimativa honesta do poder preditivo fora da amostra.

7.6 Select (encolhimento)

Quando marcada (padrão), acrescenta a cada termo suave uma penalização extra que pode encolhê-lo até zero. Isso habilita a seleção automática de variáveis — os termos suaves sem importância são efetivamente removidos do modelo. Implementado por meio de `select = TRUE` em `gam()`.

7.7 Base padrão

O tipo de função de base spline para os termos suaves:

Base	Nome completo	Descrição
tp	Spline de regressão thin plate	Padrão. Isotrópica (não é preciso posicionar nós). Boa escolha de uso geral.
cr	Spline de regressão cúbica	Permite o posicionamento explícito de nós. Necessária para a integração dos nós do earth.
ps	P-spline	Base de B-splines igualmente espaçadas com penalização por diferenças. Computacionalmente eficiente.
bs	B-spline	Base de B-splines flexível.

7.8 k padrão (dimensão da base)

A dimensão da base k controla a ondulação máxima de cada termo suave. O valor 0 (o padrão) significa “automático” — o mgcvUI usa $k = 10$ ou o número de nós do earth, o que for apropriado.

- **k pequeno (3–5)**: curvas muito suaves, quase lineares.
- **k médio (8–15)**: flexibilidade moderada (em geral suficiente).
- **k grande (20+)**: muito flexível — risco de sobreajuste sem regularização forte.

Dica: Os graus de liberdade efetivos (EDF) na aba Summary indicam quantos graus de liberdade cada termo suave de fato usa. Se o EDF estiver próximo de $k - 1$, a base pode ser pequena demais — aumente k . Se o EDF for bem menor que k , o termo suave está bem penalizado.

7.9 Tipo de tensor

Para interações entre variáveis contínuas, há dois tipos de produto tensorial disponíveis:

Tipo	Função	Descrição
ti	ti(x1, x2)	Interação tensorial — modela apenas o efeito de interação, com os efeitos principais tratados separadamente por termos <code>s()</code> univariados. Preferível pela interpretabilidade.
te	te(x1, x2)	Produto tensorial — modela todo o efeito conjunto, inclusive os efeitos principais. Mais difícil de decompor para os ajustes de RCA.

7.10 Interações permitidas

Uma seção recolhível **Allowed Interactions** exibe uma matriz triangular superior de caixas de seleção para todos os preditores suaves incluídos (não lineares e não fatores). Cada caixa de seleção habilita um termo suave de produto tensorial entre o par de variáveis correspondente.

- **Allow All / Clear All** — botões de alternância em massa.
- **Clicar no nome de uma variável** — alterna todas as interações daquela variável.
- **Clicar com o botão direito no nome de uma variável** — “Block from main effect” — marca a variável para uso apenas em interações (sem termo suave univariado).

Quando se importam nós do earthUI, as interações detectadas pelo earth ficam previamente marcadas e bloqueadas (destacadas com fundo amarelo). Se o modelo earth tiver usado `degree = 1` (sem interações), exibe-se uma mensagem informativa.

Aviso

Habilitar interações tensoriais aumenta substancialmente a complexidade do modelo. Use interações quando houver conhecimento de domínio ou evidência detectada pelo earth que as sustente. Verifique se os ajustes resultantes são razoáveis e explicáveis.

7.11 Interações fator por termo suave

Uma matriz separada, **Factor-by-Smooth Interactions**, aparece quando há variáveis de fator e variáveis suaves incluídas. Cada caixa de seleção cria um termo `s(x, by = factor)` — uma curva suave distinta para cada nível do fator.

7.12 Parâmetros avançados

Uma seção recolhível “Advanced” expõe configurações adicionais:

- **Optimizer** — algoritmo de estimação do parâmetro de suavização: `outer/newton` (padrão), `outer/bfgs` ou `efs` (extended Fellner-Schall).

- **Scale** — parâmetro de escala. 0 (padrão) = estimar a partir dos dados. Defina um valor conhecido para modelos de escala fixa.
- **Discrete covariate method** — ajuste discretizado rápido para conjuntos de dados grandes (`discrete = TRUE`).
- **Threads** — número de threads paralelas (1–32). Só tem efeito quando `discrete = TRUE`.

8 Ajuste do modelo

8.1 O botão de ajuste

A Seção 6 da barra lateral contém o botão **Fit MgcV GAM Model**. Clicar nele executa o modelo com sua configuração atual.

8.2 O que acontece durante o ajuste

Quando você clica em Fit, o mgcvUI:

1. **Valida a configuração** — verifica se há pelo menos um preditor e pelo menos 10 linhas completas. Avisa se menos de 50% das linhas estiverem completas (listando as colunas responsáveis pela maior parte dos NAs).
2. **Prepara os dados** — restringe às colunas relevantes para o modelo, remove linhas com NAs, aplica os pesos (se designados), exclui a linha do avaliando (no modo de avaliação) e aplica as transformações da resposta.
3. **Constrói a fórmula do GAM** — monta termos suaves $s(x, bs="cr", k=10)$ para os preditores numéricos, termos de fator para os categóricos, termos lineares para as variáveis marcadas como Linear e produtos tensoriais para as interações selecionadas. Os nós do earth são incorporados como posições explícitas de nós nos termos de base **cr**.
4. **Trata os casos-limite** — descarta variáveis constantes, converte preditores binários em fatores, limita k ao número de valores únicos menos 1 e complementa os nós do earth com quantis dos dados quando são necessários mais nós.
5. **Ajusta o GAM** — chama `mgcv::gam()` com a fórmula construída, a família, o método, gamma, select e os demais parâmetros.
6. **Executa a validação cruzada** (se habilitada) — VC de 10 dobras, calculando o R quadrado fora da amostra.
7. **Atualiza todas as abas de resultados** — Summary, Equation, gráficos de Contribution, Diagnostics e todas as demais abas são preenchidos.

8.3 Ajuste em segundo plano e progresso

Quando o pacote `callr` está disponível, o ajuste do modelo é executado em um processo em segundo plano, de modo que o aplicativo permanece responsivo. Uma sobreposição de progresso em estilo de terminal, com tema escuro, mostra:

- Um **cronômetro** exibindo os segundos/minutos decorridos.
- Um **registro de execução com código de cores**: azul para o estado, amarelo para o progresso das dobras da VC, vermelho para erros e verde para resultados.
- Um botão **Abort** para encerrar o processo de ajuste, se necessário.
- Um botão **Close**, que aparece depois da conclusão.

Se `callr` não estiver disponível, o ajuste é executado de forma síncrona (o aplicativo congela até o ajuste terminar).

Após um ajuste bem-sucedido, aparece uma marca de verificação branca no botão Fit e uma linha de estado mostra: “R-sq = X, CV R-sq = X, Dev = X%, AIC = X, n = X.”

9 Abas de resultados

9.1 Data

Mostra os dados importados como uma DataTable interativa, com rolagem horizontal e 15 linhas por página. Nos modos de avaliação e de mercado, a aba é dividida em uma tabela **Subject Property** (linha 1) e uma tabela **Comparable Sales** (as demais linhas); no modo General, exibe-se uma única tabela.

Cada célula é mantida em uma única linha e truncada na largura da coluna, de modo que campos largos de texto livre (como observações sobre o imóvel) não estiquem uma linha página abaixo. Um clique simples seleciona a linha; **clique duas vezes em qualquer célula** para abrir uma janela suspensa com seu conteúdo completo, sem truncamento.

9.2 Equation

Exibe o modelo GAM ajustado em duas partes:

Equação do modelo (renderizada em MathJax):

- O lado esquerdo mostra a resposta com a transformação apropriada (log, log10) e a função de ligação.
- O lado direito mostra o valor do intercepto, os termos suaves como $f_i(\text{variável})$, os termos de fator agrupados por variável com a contagem de níveis e os termos lineares com os valores dos coeficientes.
- A família e o método são exibidos acima da equação.

Tabela **Smooth Function Definitions**:

- Texto da fórmula do termo (por exemplo, `s(living_sqft, bs="cr", k=8)`)
- Nome da variável
- Tipo (s, te, ti ou linear)
- Tipo de função de base
- Valor de k
- Posições dos nós do earth (se importados)

9.3 Summary

Seis cartões de métricas exibidos na parte superior:

- **R-squared** — proporção da desviância explicada
- **CV R-squared** — R quadrado de validação cruzada (se a VC estiver habilitada)
- **Dev. Explained %** — percentual de desviância explicada
- **AIC** — critério de informação de Akaike
- **n** — número de observações utilizadas
- **Smooths** — número de termos suaves

Abaixo dos cartões:

- **Tabela Smooth Terms** — termo, EDF (graus de liberdade efetivos), Ref.df, estatística F, valor-p
- **Tabela Parametric Terms** — termo, estimativa, erro padrão, valor t, valor-p

9.4 Variable Importance

Um gráfico de barras horizontais mostrando a importância relativa de cada termo do modelo:

- **Termos suaves**: ordenados pela estatística F (barras azuis)
- **Termos paramétricos**: ordenados pelo valor $|t|$ (barras verdes)

Abaixo do gráfico, uma DataTable mostra termo, tipo, EDF, estatística e valor-p, ordenados pela estatística em ordem decrescente.

9.5 Gráficos de contribuição

Todo termo do modelo que contribui para o valor predito tem um gráfico de contribuição interativo em **plotly**, exibido em um cartão com borda. A aba mostra gráficos para todos os tipos de termo:

9.5.1 Termos suaves univariados — $s(x)$

- **Linha azul** com faixa de confiança de 95%
- **Pontos cinza de dispersão** mostrando os resíduos parciais (para modelos sem transformação logarítmica)
- **Linhas verticais vermelhas tracejadas** nas posições dos nós do earth (se o earth tiver sido importado)
- **Dica de tela ao passar o cursor:** valor da variável, contribuição (\$), IC de 95%, taxa (inclinação por unidade)

Nos **modelos com transformação logarítmica**, o eixo y é retrotransformado para contribuições em dólares pela fórmula $\bar{y} \times (e^{f(x)} - 1)$, o que torna as curvas diretamente interpretáveis em termos monetários.

Para as variáveis de **latitude/longitude**, a inclinação é mostrada por 0,001 grau, e não por grau, já que esse é um incremento geográfico mais significativo.

9.5.2 Termos fator por suave — $s(x, \text{by} = \text{factor})$

Uma linha colorida por nível do fator, sobrepostas em um único gráfico. Útil para ver como o efeito de um preditor varia entre grupos (por exemplo, entre bairros diferentes).

9.5.3 Termos de interação — $ti(x_1, x_2)$ e $te(x_1, x_2)$

Mapa de calor mostrando a superfície bidimensional de contribuição. Escala de cores: vermelho (negativo) → branco (zero) → azul (positivo). Ao passar o cursor, exibem-se os valores das duas variáveis e o montante da contribuição.

9.5.4 Termos paramétricos — lineares e de fator

- **Termos lineares numéricos:** gráfico de dispersão dos pontos de dados com a reta ajustada sobreposta
- **Termos de fator:** gráfico de barras mostrando a contribuição média por nível do fator

9.6 Correlation

Uma matriz em mapa de calor mostrando as correlações de Pearson entre todos os preditores numéricos. Disponível imediatamente após a importação dos dados — não é preciso ajustar o modelo.

- Gradiente de cores: azul (-1) → branco (0) → vermelho $(+1)$
- Valores de correlação impressos em cada célula
- Útil para identificar concurvidade (o análogo suave da multicolinearidade) antes do ajuste

9.7 Diagnostics

Dois conjuntos de visualizações de diagnóstico:

Painel de diagnóstico (via `gratia::appraise`): quatro gráficos — resíduos versus ajustados, gráfico Q-Q, histograma dos resíduos e resposta versus valores ajustados.

Gráfico de dispersão **Actual vs Predicted**: valores observados versus preditos, com uma reta de referência tracejada a 45 graus. Nos modelos com transformação logarítmica, os valores são retrotransformados para a escala original.

9.8 RCA Adjustments

Três histogramas exibidos após o cálculo do RCA (somente no modo de avaliação/mercado):

- **Residual Adjustment %** — distribuição dos ajustes de resíduo como percentual do preço de venda
- **Net Adjustment %** — distribuição dos ajustes líquidos
- **Gross Adjustment %** — distribuição dos ajustes brutos

Cada histograma mostra média, mediana e desvio padrão no subtítulo, com linhas de referência tracejadas para a média e a mediana.

9.9 Anova

Uma tabela de ANOVA combinada, reunindo os termos paramétricos e suaves do resumo do GAM. Mostra o termo, o tipo (paramétrico ou suave) e todas as estatísticas associadas.

9.10 Mgcvc Output

Saída em texto bruto mostrando:

- Tempos de execução do ajuste
- A fórmula do GAM
- A saída de impressão do modelo
- A saída completa de `summary(model)`
- A família e a função de ligação
- As medidas globais de concurvidade

9.11 Sign Check

Quando se importam nós do earthUI, esta aba compara a direção de cada variável no modelo earth com sua direção no GAM:

- **earth_direction**: crescente, decrescente ou mista (com base nos sinais das funções de dobradiça)
- **gam_direction**: crescente, decrescente ou mista (estimada a partir da derivada do termo suave em uma grade de 100 pontos)
- **consistent**: TRUE/FALSE
- **warning**: descrição de eventual divergência

Uma linha de estado mostra “All directions consistent” ou o número de variáveis com inconsistências.

9.12 Concurvity

A concurvidade é o análogo suave da multicolinearidade. Esta aba mostra:

- **Tabela Overall Concurvity:** linhas para as medidas worst/observed/estimate e colunas para cada termo suave. Com código de cores: branco ($<0,5$), amarelo ($0,5-0,8$), vermelho ($>0,8$).
- **Tabela Pairwise Concurvity:** matriz aos pares no pior caso, com o mesmo código de cores.

Valores acima de $\sim 0,8$ na linha “worst” indicam que um termo suave pode ser bem aproximado pelos demais termos suaves, sugerindo redundância.

9.13 Basis Check

Saída de `mgcv::gam.check()`: testes de adequação da dimensão da base para cada termo suave. Se o k de um termo suave for pequeno demais, este teste o sinalizará com um valor-p baixo, sugerindo que você aumente k .

10 Download dos dados

Após o ajuste, o botão **Download Output (Excel)** (seção 7 da barra lateral) exporta um arquivo Excel com predições e diagnósticos para a pasta de saída do mgcv do projeto ativo.

10.1 Colunas de saída

Coluna	Descrição
basis	Valor do intercepto retrotransformado (o mesmo para todas as linhas)
<var>_contribution	Contribuição de cada termo suave para a predição. Nos modelos logarítmicos, alocada proporcionalmente a valores em dólares.
est_<target>	Predição do modelo (por exemplo, est_sale_price), retrotransformada da escala logarítmica quando aplicável
residual	Observado – predito
calc_residual	Verificação: observado – soma das contribuições (retrotransformada)
cqa	Escore de Condition-Quality-Appeal (CQA), em escala de 0 a 10
residual_sf	Resíduo / área habitável (se living_area estiver designada)
cqa_sf	CQA calculado pela ordenação segundo residual_sf

No modo de avaliação, a linha 1 (avaliando) tem residual/cqa/cqa_sf definidos como NA. As linhas são ordenadas por residual_sf em ordem decrescente quando há uma coluna living_area designada.

Uma marca de verificação branca aparece no botão de download após a conclusão bem-sucedida.

10.2 Escores de CQA

O CQA ordena o resíduo de cada linha em relação a todos os demais, em uma escala de 0 a 10:

- **CQA alto (~9–10)**: vendeu por muito mais do que o predito — provavelmente condição/qualidade/atratividade superiores
- **CQA baixo (~0–1)**: vendeu por muito menos do que o predito — provavelmente condição inferior ou venda em situação de desconforto
- **CQA ~5**: próximo da mediana

11 Cálculos de RCA e download

O fluxo de trabalho do RCA (Residual Constraint Approach) está disponível nos modos **Appraisal** e **Market Area Analysis**, depois do ajuste do modelo. O imóvel avaliando deve estar na linha 1.

11.1 Abertura da caixa de diálogo do RCA

Clique no botão **Calculate RCA Adjustments & Download** na seção 8 da barra lateral. Aparece uma caixa de diálogo modal com:

- **Tipo de escore:** CQA ou CQA por SF (se `living_area` estiver designada). A opção por SF escala o resíduo do avaliando pela área habitável.
- **Subject CQA Score:** aceita valores de 0,00 a 10,00. O campo começa em branco; quando um modelo do earthUI tiver sido importado, o tipo e o valor de escore que você usou no earthUI são trazidos das configurações compartilhadas do projeto.

Clique em **Generate** para calcular e baixar. O arquivo Excel é gravado na pasta de saída do mgcv do projeto ativo (`<os>_out_mgcv`).

11.2 Interpolação do escore de CQA

1. Os escores de CQA e os resíduos dos comparáveis são ordenados
2. A interpolação linear (`approx()`) mapeia seu valor de CQA para um resíduo
3. Valor do avaliando = predição do modelo + resíduo interpolado
4. As linhas com peso 0 são tratadas como se tivessem o mesmo resíduo do avaliando

11.3 Colunas de saída

Coluna	Descrição
<code>subject_value</code>	Predição do modelo + resíduo interpolado
<code>subject_cqa</code>	Escore de CQA informado pelo usuário
<code><var>_adjustment</code>	Contribuição do avaliando – contribuição do comparável (em dólares)
<code>residual_adjustment</code>	Resíduo do avaliando – resíduo do comparável
<code>net_adjustments</code>	Soma de todos os ajustes
<code>gross_adjustments</code>	Soma dos ajustes em valor absoluto
<code>residual_adj_pct</code>	Ajuste de resíduo como % do preço de venda
<code>net_adj_pct</code>	Ajustes líquidos como % do preço de venda
<code>gross_adj_pct</code>	Ajustes brutos como % do preço de venda
<code>adjusted_sale_price</code>	Preço de venda observado + ajustes líquidos

Uma marca de verificação branca aparece no botão após o cálculo bem-sucedido.

12 Grade de comparação de vendas

A grade de comparação de vendas está disponível somente no modo **Appraisal**, depois do cálculo dos ajustes de RCA. Ela gera uma pasta de trabalho Excel formatada, adequada para inclusão em laudos de avaliação.

12.1 Caixa de diálogo de seleção de comparáveis

Clicar no botão **Generate Sales Grid & Download** abre uma caixa de diálogo modal listando todas as vendas comparáveis. Os comparáveis são divididos em dois grupos:

- **Comparáveis recomendados** (previamente marcados): ajuste bruto $< 25\%$ do preço de venda, ordenados pela idade da venda (o mais recente primeiro). No máximo 30 comparáveis.
- **Outros comparáveis** (desmarcados por padrão): os comparáveis restantes, ordenados pelo percentual de ajuste bruto.

Selecione seus comparáveis e clique em **Generate Sales Grid** para criar a pasta de trabalho.

Tipos especiais necessários: `contract_date` e `living_area` devem estar designados.

Tipos especiais recomendados: `latitude`, `longitude` e `lot_size` — a grade funciona sem eles, mas alguns campos ficarão em branco.

12.2 Estrutura da pasta de trabalho

A pasta de trabalho gerada contém várias planilhas, cada uma com 3 comparáveis lado a lado:

- **Linhas de cabeçalho:** endereços do avaliando e dos comparáveis, preços de venda, datas de venda e dias no mercado
- **Linhas agrupadas:** localização (latitude, longitude, região), tamanho do terreno (lot size, site dimensions) e idade (idade real, idade efetiva), com cabeçalhos de grupo e subitens
- **Linhas das variáveis do modelo:** uma linha por termo suave, mostrando a contribuição do avaliando, as contribuições dos comparáveis e os montantes de ajuste
- **Linhas de características residuais:** células vazias e desbloqueadas para preenchimento pelo avaliador quanto a características não captadas pelo modelo (por exemplo, condição, qualidade, vista)
- **Linhas de resumo:** % de ajuste líquido, % de ajuste bruto e preço de venda ajustado, com fórmulas ativas do Excel

12.3 Proteção da planilha e células residuais

Cada planilha é protegida para evitar edições acidentais, mas as células de valor das características residuais ficam explicitamente desbloqueadas. Isso permite que os avaliadores informem valores para características ausentes do modelo, preservando os cálculos de ajuste orientados por fórmulas.

13 Download de relatórios

A geração de relatórios é um fluxo de trabalho Quarto em duas etapas, que substitui o relatório de etapa única das versões anteriores.

13.1 Etapa 1: gerar o relatório Quarto

O botão **Generate Quarto Report** (disponível depois que o modelo é ajustado) grava um pacote `.qmd` autocontido na pasta de saída do `mgecv` do projeto ativo. O pacote contém o código-fonte Quarto, arquivos de gráficos pré-renderizados (PNG e PDF), uma carga de dados `report_data.rds` com todas as tabelas e os metadados do modelo, e um modelo de estilo `reference.docx` para a saída em Word. Como o pacote é autocontido, você pode editar o código `.qmd` ou combiná-lo com outros documentos Quarto antes de renderizar.

Os gráficos são pré-renderizados quando o pacote é gerado; se um gráfico individual falhar, grava-se uma imagem de espaço reservado e o restante do pacote é gerado normalmente.

13.2 Etapa 2: converter o relatório Quarto

A seção **Convert Quarto Report** renderiza um arquivo `.qmd` em um ou mais formatos de saída, via `quarto::quarto_render()`. Por padrão, usa o pacote gerado mais recentemente, mas o botão **Browse** permite selecionar qualquer arquivo `.qmd`. Há três formatos disponíveis:

- **HTML** — documento HTML autocontido. Funciona em todas as plataformas, sem software adicional. Padrão recomendado.
- **Word (.docx)** — renderizado com o modelo de estilo `reference.docx` incluído no pacote. Adequado para edição e distribuição.
- **PDF** — **exige uma instalação de LaTeX** (veja os requisitos de sistema adiante). Se nenhum LaTeX for detectado, a opção de PDF é automaticamente ocultada nas caixas de seleção de formato.

Os renderizadores legados de etapa única, `render_gam_report()` e `export_gam_docx()`, continuam disponíveis para uso programático.

13.3 Conteúdo do relatório

O relatório gerado inclui:

- **Conjunto de dados e modelo:** arquivo de dados, tamanho da amostra, família, método e informações sobre o ajuste
- **Resumo dos resultados:** R^2 , R^2 de VC, desviância explicada, AIC e métricas de ajuste relacionadas
- **Equação do modelo:** equação do modelo com a notação f_i para os termos suaves e coeficientes explícitos para os termos lineares, além das definições das funções suaves. As equações longas se quebram em várias linhas no PDF.
- **Tabelas de resumo:** termos suaves (EDF, F, valor-p) e termos paramétricos (estimativa, erro padrão, t, valor-p)
- **Interações habilitadas:** uma matriz dos termos de interação tensorial e de fator por suave incluídos no modelo (exibida somente quando há interações)
- **Importância das variáveis:** gráfico de barras ordenado pela estatística F (termos suaves) ou por $|t|$ (paramétricos)
- **Efeitos suaves** (todos os tipos de termo):
 - Gráficos de termos suaves univariados (versões estáticas em `ggplot2` dos gráficos interativos)

- Gráficos multilinhas de fator por suave
- Mapas de calor de interação (termos ti e te)
- Gráficos de termos paramétricos (dispersão para numéricos, barras para fatores)
- Mapa de calor de **correlações entre preditores**
- **Diagnósticos:** diagnósticos de resíduos e dispersão de observados versus preditos
- **Resumo do mgcv:** a saída de texto completa do modelo ajustado
- **Apêndice de notação:** uma seção final não numerada explicando a notação das funções de contribuição de valor usada no relatório

Uma marca de verificação branca aparece no botão Generate Quarto Report após a geração bem-sucedida.

14 Exportação das funções do modelo

O `mgcvUI` pode exportar as funções suaves do GAM ajustado como código autônomo em várias linguagens. Cada termo suave é avaliado em uma grade de 200 pontos e exportado como uma tabela de consulta com interpolação linear, o que permite usar o modelo fora do R sem qualquer dependência de `mgcv`.

Na versão atual, a exportação de funções é programática. Os auxiliares exportados recebem um objeto de resultado ajustado (produzido por `fit_gam()`) e gravam os arquivos descritos a seguir:

- `generate_r_code()`, `generate_python_code()` e `generate_cpp_code()` — arquivos de código-fonte por linguagem
- `export_functions_sqlite()` — banco de dados SQLite com as tabelas de consulta
- `export_functions_zip()` — reúne qualquer combinação dos quatro formatos em um arquivo zip

14.1 Funções em R

Arquivo: `gam_functions.R`

- Comentários de cabeçalho com a variável resposta, R^2 , R^2 de VC e a fórmula.
- Uma constante `INTERCEPT`.
- Um `approxfun()` por variável suave, com os vetores `x/y` pré-calculados.
- Os coeficientes dos termos lineares como constantes nomeadas.
- Uso: `some` o intercepto, cada `g_var(value)` e cada `coef_var * value`.

14.2 Funções em Python

Arquivo: `gam_functions.py`

- Importação de `numpy`.
- Constante `INTERCEPT`.
- Por variável: arranjos `_x_var` e `_fx_var` e função `g_var(x)` usando `np.interp()`.
- Os coeficientes dos termos lineares como constantes `COEF_*`.

14.3 Cabeçalho C++

Arquivo: `gam_functions.hpp`

- Somente de cabeçalho, com `#pragma once`.
- Espaço de nomes `gam_functions`.
- Auxiliar `interp()` usando `std::lower_bound`.
- Por variável: função inline com arranjos `static const`.
- Os coeficientes dos termos lineares como `constexpr double`.

14.4 Banco de dados SQLite

Arquivo: `gam_functions.sqlite`

- Tabela `smooth_grids`: colunas (`variable`, `x`, `fx`) — as tabelas de consulta de 200 pontos.
- Tabela `model_info`: `response`, `r_squared`, `cv_rsqr`, `dev_explained`, `aic`, `n_obs`, `family`, `method`, `formula`.
- Tabela `linear_terms`: `variable`, `coefficient`.
- Tabela `intercept`: valor do intercepto.

15 Comparação com earthUI e glmnetUI

O aplicativo earth e seus módulos de pós-processamento formam um conjunto de ferramentas para modelagem por regressão. Eles compartilham o mesmo formato de dados, os mesmos tipos especiais de coluna, o mesmo fluxo de trabalho de RCA e os mesmos conjuntos de dados de demonstração, mas usam motores de modelagem diferentes.

15.1 Principais diferenças

Característica	mgcvUI	earthUI	glmnetUI
Método	GAM (mgcv)	MARS/Earth (earth)	Elastic net (glmnet)
Relações	Não lineares suaves	Lineares por partes	Lineares
Seleção de variáveis	Encolhimento com <code>select=TRUE</code>	Poda por passos	Penalização Lasso
Interações	Produtos tensoriais (te/ti)	Produtos de dobradiças (grau 2-3)	Produtos aos pares
Efeitos parciais	Curvas suaves com IC	Funções g (por partes)	Inclinações lineares
Fluxo com o earth	Importa nós do earth	Fonte dos nós	Importa a base de dobradiças
Exportação de código	R, Python, C++, SQLite	R, Python, C++, SQLite	—
Formatos de relatório	Word, PDF, HTML	Word, PDF, HTML	Word, PDF, HTML

Vale registrar um diferencial exclusivo do earthUI: ele inclui um recurso experimental em Prolog/DCG que extrai atributos estruturados de imóveis a partir das observações em texto livre do MLS.

15.2 Recursos compartilhados

As três ferramentas oferecem:

- Uma árvore de projetos `regProj` compartilhada (por padrão `~/regProj`) — projetos, arquivos de entrada e saídas por método são compartilhados entre os aplicativos irmãos
- Importação de CSV e Excel com conversão dos nomes de coluna para `snake_case`
- Suporte a localidades de mais de 30 países, com separador de CSV, marcador decimal e tamanho de papel
- Designações especiais de colunas (`contract_date`, `living_area`, `area`, `latitude`, `longitude` etc.)
- Cálculo automático de `sale_age` a partir da data de referência
- Alternância entre modo claro e escuro (tema Nord)
- Persistência das configurações (SQLite e/ou `localStorage`)
- Mapa de calor de correlações, importância das variáveis, diagnósticos
- Escores de CQA e fluxo de trabalho de ajustes de RCA
- Grade de comparação de vendas com seleção de comparáveis e saída formatada em Excel
- Compatibilidade com os mesmos conjuntos de dados de demonstração (`Appraisal_1.csv`)

15.3 Fluxo de trabalho recomendado

- **Comece pelo earthUI** para descobrir as variáveis importantes, as relações não lineares e as interações. A construção automática de modelos do earth oferece o melhor ponto de partida.

- **Refine com o mgcvUI** se quiser efeitos parciais mais suaves. Importe os nós do earthUI para uma transição contínua.
- **Use o glmnetUI** quando precisar de um modelo puramente linear com seleção agressiva de variáveis ou restrições de sinal dos coeficientes.
- **Compare os três** sobre os mesmos dados para avaliar se a complexidade adicional dos modelos não lineares se justifica.

16 Conjunto de dados de demonstração: Appraisal_1.csv

16.1 Descrição

O conjunto de dados de demonstração é compartilhado com o earthUI e o glmnetUI. Ele está na pasta `demo_mls/` do código-fonte do mgcvUI ou, se o earthUI estiver instalado:

```
demo_file <- system.file("extdata", "Appraisal_1.csv", package = "earthUI")
```

O arquivo contém cerca de 1.500 registros de vendas residenciais de uma exportação simulada de MLS, com o imóvel avaliando na linha 1 e as vendas comparáveis nas demais linhas. Os dados representam vendas de residências unifamiliares em um mercado com várias regiões, com uma variedade de tamanhos, idades e localizações de imóveis.

Não se trata de dados reais, mas eles se baseiam em um bairro realista do norte da Califórnia. Todas as informações de identificação foram alteradas ou removidas.

16.2 Colunas

Coluna	Tipo	Tipo especial	Descrição
<code>sale_type</code>	character	<code>sale_type</code>	Código de tipo de registro que distingue o avaliando das vendas comparáveis
<code>weight</code>	numeric	<code>weight</code>	Peso da observação (0 = excluir do ajuste)
<code>id</code>	numeric	<code>display_only</code>	Identificador interno do registro
<code>property_id</code>	numeric	<code>display_only</code>	Identificador do imóvel no MLS
<code>listing_id</code>	character	<code>display_only</code>	Número do anúncio no MLS
<code>parcel_number</code>	character	<code>display_only</code>	Número da matrícula no cadastro do condado (APN)
<code>street_address</code>	character	<code>display_only</code>	Endereço do imóvel
<code>city_name</code>	character	<code>display_only</code>	Cidade
<code>postal_code</code>	character	<code>display_only</code>	Código postal (ZIP)
<code>county_name</code>	character	<code>display_only</code>	Condado
<code>contract_date</code>	Date	<code>contract_date</code>	Data do contrato de venda (calcula <code>sale_age</code>)
<code>sale_age</code>	numeric	—	Dias entre a data do contrato e a data de referência (pré-calculado)
<code>coe_date</code>	Date	<code>display_only</code>	Data de encerramento do escrow
<code>listing_status</code>	character	<code>display_only</code>	Situação do anúncio (por exemplo, "Sold")
<code>sale_price</code>	numeric	(alvo)	Preço de venda — variável resposta
<code>rent</code>	numeric	—	Aluguel mensal (para modelos com vários alvos)
<code>list_price</code>	numeric	<code>display_only</code>	Preço de anúncio
<code>original_list_price</code>	numeric	<code>display_only</code>	Preço de anúncio original
<code>living_sqft</code>	numeric	<code>living_area</code>	Área bruta habitável em pés quadrados
<code>ppsf</code>	numeric	<code>display_only</code>	Preço por pé quadrado
<code>beds_total</code>	integer	—	Número de dormitórios
<code>baths_total</code>	numeric	—	Número total de banheiros (por exemplo, 2,5 = 2 completos + 1 lavabo)
<code>lot_size</code>	numeric	<code>lot_size</code>	Tamanho do terreno em pés quadrados
<code>area_id</code>	integer	<code>area</code>	Identificador de região do MLS
<code>area_text</code>	character	<code>display_only</code>	Nome da região
<code>age</code>	numeric	<code>actual_age</code>	Idade do imóvel em anos
<code>year_built</code>	integer	<code>display_only</code>	Ano de construção
<code>latitude</code>	numeric	<code>latitude</code>	Latitude (arredondada para 3 casas decimais no modelo)
<code>longitude</code>	numeric	<code>longitude</code>	Longitude (arredondada para 3 casas decimais no modelo)
<code>latitude6</code>	numeric	<code>display_only</code>	Latitude com precisão total
<code>longitude6</code>	numeric	<code>display_only</code>	Longitude com precisão total
<code>garage_spaces</code>	integer	—	Número de vagas de garagem

Coluna	Tipo	Tipo especial	Descrição
<code>fp_count</code>	integer	—	Número de lareiras
<code>no_of_stories</code>	numeric	—	Número de pavimentos
<code>style</code>	character	—	Estilo arquitetônico
<code>view</code>	character	—	Tipo de vista (por exemplo, “Neighborhood”, “Hills”)
<code>days_on_market</code>	integer	<code>dom</code>	Dias no mercado
<code>listing_date</code>	Date	<code>listing_date</code>	Data do anúncio
<code>sale_concessions</code>	numeric	<code>concessions</code>	Concessões do vendedor

16.3 Roteiro rápido sugerido

1. Inicie a rotina mgcv: `earthUI::launch_mgcv()`
2. Crie um novo projeto (Seção 1) com o propósito **For Appraisal**
3. Coloque `Appraisal_1.csv` na pasta de entrada do projeto e selecione-o na Seção 2 (clique em **Refresh** se necessário)
4. Selecione `sale_price` como alvo
5. Opcionalmente, defina Response Transform como **Log (natural)** para obter ajustes proporcionais
6. Atribua os tipos especiais conforme a tabela acima
7. Inclua os preditores: `sale_age`, `living_sqft`, `baths_total`, `lot_size`, `area_id` (como fator), `age`, `latitude`, `longitude`, `garage_spaces`
8. Mantenha os padrões: REML, $\gamma = 1.2$, base tp, $k = \text{auto}$, `select = TRUE`
9. Clique em **Fit Mgcv GAM Model**
10. Examine as abas Summary, Contribution (gráficos suaves) e Diagnostics
11. Baixe a saída (etapa 7) e examine a ordenação por CQA
12. Calcule os ajustes de RCA (etapa 8) com um escore de CQA de ~5,00
13. Gere uma grade de comparação de vendas (etapa 9) com os comparáveis recomendados

17 Requisitos de sistema e solução de problemas

17.1 Plataformas suportadas

O mgcvUI roda em **macOS**, **Windows** e **Linux**. O aplicativo é desenvolvido e testado principalmente em macOS com o RStudio. As observações específicas de cada plataforma vêm a seguir.

17.2 Versão do R

É necessário $R \geq 4.1.0$. Recomenda-se fortemente o RStudio Desktop (2023.06 ou posterior) — ele já inclui o **Quarto** e o **pandoc** (necessários para renderizar relatórios) e oferece um ambiente conveniente para iniciar o aplicativo.

17.3 Pacotes de R necessários

Os pacotes a seguir são instalados automaticamente como dependências:

Pacote	Finalidade
shiny	Arcabouço de aplicações web
mgcv	Ajuste dos modelos GAM
valengrCore	Núcleo compartilhado de avaliação (papéis especiais, transferência a partir do earthUI)
ggplot2, plotly	Gráficos de contribuição estáticos e interativos
gratia	Extração de termos suaves e diagnósticos
DT	Tabelas de dados interativas
quarto, rmarkdown	Geração e renderização de relatórios Quarto
readxl	Importação de arquivos Excel
shinyFiles	Navegadores de arquivos e pastas
RSQLite, DBI	Bancos de dados de configurações e de projetos (SQLite)
jsonlite	Serialização do estado das variáveis

17.4 Dependências opcionais

Componente	Quando é necessário
LaTeX (TinyTeX, MiKTeX ou MacTeX)	Somente para gerar relatórios em PDF. Se nenhuma instalação de LaTeX for detectada, a opção de PDF é automaticamente ocultada no menu de formatos. Instale com: <code>tinytex::install_tinytex()</code>
sysfonts + showtext	Fonte Roboto Condensed para os gráficos em ggplot2. Se estiver indisponível ou o computador estiver off-line, o aplicativo recorre automaticamente à fonte sem serifa do sistema.
earth ($\geq 5.3.0$)	Necessário apenas para o fluxo de importação do earthUI. Não é exigido no uso autônomo.
callr	Ajuste do modelo em segundo plano (mantém o aplicativo responsivo). Sem ele, o ajuste é executado de forma síncrona.
writexl, openxlsx	Downloads em Excel (saída do modelo, RCA, grade de vendas). O aplicativo avisa se estiverem ausentes.
bslib	Tematização Bootstrap 5 (Nord Light/Dark).
officer	Exportação programática legada para Word (<code>export_gam_docx()</code>).

17.5 Observações por plataforma

17.5.1 macOS

É onde há menos problemas. Recomenda-se o Homebrew para as bibliotecas de sistema. Se forem necessários relatórios em PDF:

```
# In R:
install.packages("tinytex")
tinytex::install_tinytex()
```

17.5.2 Windows

Funciona bem com o RStudio Desktop. Observações principais:

- **Máquinas corporativas ou restritas:** se o diretório temporário ou o diretório de dados do usuário estiver restrito, a persistência das configurações e a geração de relatórios podem falhar. O aplicativo exibirá um aviso no console do R, mas continuará funcionando sem persistência.
- **LaTeX:** instale o TinyTeX a partir do R (mesmo comando do macOS acima) ou instale o MiKTeX em miktex.org.
- **Caminhos de arquivo longos:** o Windows tem um limite de 260 caracteres para caminhos. Mantenha curto o caminho do seu diretório de trabalho.

17.5.3 Linux

É a plataforma mais variável entre distribuições. Usuários de Ubuntu/Debian podem precisar de bibliotecas de sistema antes que os pacotes de R sejam compilados:

```
sudo apt install libcurl4-openssl-dev libssl-dev \
  libxml2-dev libsqlite3-dev libfontconfig1-dev
```

Para relatórios em PDF: `tinytex::install_tinytex()` ou instale `texlive-full` pelo gerenciador de pacotes do sistema.

Em servidores headless (sem monitor), os gráficos interativos em plotly são renderizados como alternativas estáticas nos relatórios. O próprio aplicativo Shiny exige uma conexão por navegador web.

17.6 Degradação graciosa

O mgcvUI foi concebido para funcionar mesmo quando componentes opcionais estão ausentes:

Componente ausente	Comportamento
Sem LaTeX	A opção de PDF fica oculta no menu de formatos de relatório. HTML e Word continuam disponíveis.
Sem internet / falha nas fontes	A fonte Roboto Condensed é substituída pela fonte sem serifa do sistema. Registra-se uma mensagem no console.
Sem RSQLite / sistema de arquivos somente leitura	As configurações não persistem entre sessões. Exibe-se um aviso no console. No mais, o aplicativo funciona normalmente.
Falha de <code>gratia::appraise</code>	A aba Diagnostics e os relatórios exibem uma mensagem de espaço reservado. Use <code>mgcv::gam.check()</code> no console como alternativa.
Diretório temporário sem permissão de escrita	A geração de relatórios falha com uma mensagem de erro clara. Resolva definindo a variável de ambiente <code>TMPDIR</code> para um local com permissão de escrita.

17.7 Solução de problemas

17.7.1 “PDF option is not available”

Nenhuma instalação de LaTeX foi detectada. Execute no R:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

Depois reinicie o aplicativo. A opção de PDF aparecerá.

17.7.2 “Settings will not persist”

O diretório do banco de dados SQLite não pode ser criado ou gravado. Verifique as permissões em:

- **macOS/Linux:** `~/Library/Application Support/R/mgcvUI/` ou `~/.local/share/R/mgcvUI/`
- **Windows:** `%APPDATA%/R/data/mgcvUI/`

17.7.3 “Diagnostics plot unavailable”

A função `gratia::appraise()` ocasionalmente falha em modelos complexos (muitas interações tensoriais, famílias não padronizadas). É um problema conhecido a montante. Solução alternativa: execute `mgcv::gam.check(model)` no console do R para ver diretamente os gráficos de diagnóstico.

17.7.4 “Report error: number of items to replace...”

Este foi um problema conhecido em versões anteriores, relacionado ao pacote `gratia`. Certifique-se de ter a versão mais recente do `mgcvUI` instalada. A versão atual captura esse erro e produz o relatório com um espaço reservado na seção de diagnósticos.

17.7.5 Porta 7880 já em uso

Se o aplicativo não iniciar porque a porta 7880 está ocupada (por uma sessão anterior), execute:

```
# macOS/Linux:
```

```
lsof -ti:7880 | xargs kill
```

```
# Windows (PowerShell):
```

```
Stop-Process -Id (Get-NetTCPConnection -LocalPort 7880).OwningProcess
```

17.7.6 Falha na instalação de pacotes no Linux

Faltam bibliotecas de sistema. Instale os cabeçalhos de desenvolvimento listados na seção sobre Linux acima e repita `install.packages()`.

Referências

- Wood, S. N. (2003). “Thin-plate regression splines”. Em: *Journal of the Royal Statistical Society (B)* 65.1, pp. 95–114. DOI: [10.1111/1467-9868.00374](https://doi.org/10.1111/1467-9868.00374).
- (2004). “Stable and efficient multiple smoothing parameter estimation for generalized additive models”. Em: *Journal of the American Statistical Association* 99.467, pp. 673–686. DOI: [10.1198/016214504000000980](https://doi.org/10.1198/016214504000000980).
- (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. Em: *Journal of the Royal Statistical Society (B)* 73.1, pp. 3–36. DOI: [10.1111/j.1467-9868.2010.00749.x](https://doi.org/10.1111/j.1467-9868.2010.00749.x).
- (2017). *Generalized Additive Models: An Introduction with R*. 2ª ed. Chapman e Hall/CRC.
- Wood, S. N., N. Pya e B. Säfken (2016). “Smoothing parameter and model selection for general smooth models (with discussion)”. Em: *Journal of the American Statistical Association* 111, pp. 1548–1575. DOI: [10.1080/01621459.2016.1180986](https://doi.org/10.1080/01621459.2016.1180986).

Fundamentos

Visão geral

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor, editor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/82vssv81](https://doi.org/10.67330/82vssv81)

A seção *Foundations* publica tratamentos precisos e citáveis dos conceitos que fundamentam a metodologia moderna de avaliação. Os verbetes tomam termos que os avaliadores em atividade encontram — às vezes com outros nomes, às vezes sem reconhecê-los — e os situam em seu contexto formal: de onde vêm, o que significam tecnicamente, onde aparecem na prática e por que importam para a defensabilidade.

A edição inaugural estabelece os seis conceitos centrais do arcabouço hedônico, dispostos de modo que cada verbete se apoie no vocabulário dos anteriores. A sequência vai de *Bem heterogêneo* (a propriedade dos imóveis que torna necessária a própria metodologia de avaliação), passando por *Espaço de características* e pela *Função de preços hedônicos* (o aparato formal que organiza o conjunto de atributos que um imóvel representa), por *Preço implícito* e *Regressão hedônica* (como a contribuição marginal de cada atributo é recuperada a partir da evidência de mercado), até chegar a *Variável latente* (o que de fato representa o sinal residual não recuperado em qualquer modelo hedônico).

Um único exemplo contínuo — oito elementos comparativos de um bairro de Pacifica, Califórnia — percorre os seis verbetes e constitui o fio pedagógico da edição.

Por que estes seis, nesta ordem

A escolha dos seis verbetes não é arbitrária. Cada conceito é estrutural para a metodologia hedônica moderna, cada um é genuinamente encontrado na prática cotidiana da avaliação de imóveis, e cada um é mal compreendido ou permanece sem nome com frequência suficiente para que colocá-lo em forma citável tenha valor prático.

A ordenação é ditada pela dependência conceitual. *Bem heterogêneo* é a raiz: nomeia a propriedade dos imóveis (e dos automóveis usados, dos mercados de trabalho e de alguns outros bens) que torna necessário todo o arcabouço hedônico. *Espaço de características* é o aparato formal para representar bens heterogêneos como vetores; nada mais pode ser dito com precisão sem ele. A *função de preços hedônicos* é o objeto de equilíbrio que associa preços aos pontos do espaço de características. Os *preços implícitos* são suas derivadas parciais — as contribuições marginais que cada ajuste em uma planilha de homogeneização do método comparativo direto de dados de mercado pode ser lido como afirmando, quer o avaliador os nomeie assim ou não. A *regressão hedônica* é o estimador empírico que recupera os preços implícitos a partir das transações observadas; é o verbete em que a teoria se converte em análise de dados. As *variáveis latentes* começam onde começam os resíduos da regressão: os atributos não observados cujos efeitos sobrevivem no resíduo como sinal estruturado, e não como ruído descartável.

Um leitor que avance nesta ordem deve terminar com um fio contínuo que vai de “por que a metodologia de avaliação existe, afinal?” até “o que devemos fazer com os resíduos de um modelo ajustado?” — um percurso que, na experiência do autor, a maioria dos avaliadores completa apenas aos pedaços, ao longo de muitos anos de prática.

O exemplo contínuo

Um único conjunto de dados de oito elementos comparativos residenciais de Pacifica, Califórnia, percorre os seis verbetes:

Elem.	GLA (sf)	Terreno (sf)	Vista	Conserv.	Preço de venda
A	1.650	9.000	sim	bom	\$1.425.000
B	1.650	7.200	não	bom	\$1.280.000
C	1.850	7.500	não	bom	\$1.407.500
D	1.450	7.000	não	bom	\$1.155.000
E	1.700	8.000	sim	excelente	\$1.470.000
F	1.550	7.100	não	médio	\$1.177.500
G	1.700	10.500	não	bom	\$1.392.500
H	1.600	7.300	não	bom	\$1.312.500

O exemplo é deliberadamente pequeno o bastante para ser raciocinado à mão e rico o bastante para exercitar cada conceito. O verbete 001 examina a diferença de \$145.000 entre os elementos A e B como a contribuição *conjunta* de duas características (vista e terreno) emaranhadas em uma única observação de pares de vendas. O verbete 002 situa os oito elementos como pontos em um espaço de características de quatro dimensões e mostra que métricas de distância ingênuas os ordenam mal. O verbete 003 propõe uma forma funcional para a função de preços hedônicos deste bairro. O verbete 004 recupera analiticamente os preços implícitos e decompõe a diferença original entre A e B em uma contribuição de \$127.039 vinda da vista e do terreno, restando \$18.000 sem explicação. O verbete 005 ajusta o modelo formalmente em R e reporta a saída completa de `lm()`, incluindo erros padrão e resíduos. O verbete 006 examina o único resíduo que o modelo não consegue explicar — um desvio de +\$46.455 no elemento H — e o conecta ao enquadramento de variáveis latentes que fundamenta a metodologia do Residual Constraint Approach, desenvolvida pelo autor em outros trabalhos.

O conjunto de dados está disponível no repositório da edição em `shared/data/pacifica_comps.csv` para os leitores que desejarem reproduzir qualquer um dos cálculos.

Como ler esta edição

Um leitor novo na metodologia hedônica deve ler os seis verbetes em ordem. As dependências são reais: cada verbete pressupõe o vocabulário estabelecido nos anteriores, e saltar adiante convida a confusões que os próprios verbetes não têm como antecipar.

Um leitor com exposição prévia à regressão hedônica — o que inclui a maioria dos avaliadores que cursaram estatística avançada, trabalharam com um AVM ou leram a literatura de avaliação em massa — pode usar a edição como revisão e como obra de referência. Os verbetes são escritos para serem citáveis atômicamente; um artigo publicado em outra parte da revista que faça referência a *preço implícito* pode remeter ao verbete 004 pelo DOI em vez de derivar o conceito novamente.

Um leitor interessado especificamente na metodologia do Residual Constraint Approach deve ler primeiro o verbete 006, para compreender o enquadramento de variáveis latentes, voltar em seguida aos verbetes 003-005, para o aparato hedônico que o RCA estende, e depois aos verbetes 001-002, para o vocabulário fundacional sobre o qual tudo repousa.

Um leitor interessado na questão de *o que torna defensável um ajuste de avaliação* descobrirá que a seção “Por que importa para a defensabilidade” de cada verbete, lida em sequência, equivale

a um breve ensaio sobre o estatuto metodológico da avaliação baseada em ajustes. Esse ensaio não está reunido explicitamente em nenhuma parte da edição, mas é o tema implícito de cada verbete.

Postura editorial

A seção *Foundations* não é uma seção de livro-texto. Cada verbete é breve o bastante para ser lido de uma só vez, escrito para um público já fluente na prática da avaliação de imóveis, e concebido para dar uma formulação precisa de um conceito em vez de uma exposição exaustiva. O objetivo é oferecer um referente estável que outros artigos da revista possam citar.

Os verbetes passam por revisão editorial e podem trazer comentários convidados de um especialista na matéria. São publicados sob licença Creative Commons Atribuição e recebem DOI tanto pelo Crossref (como citação canônica) quanto pelo Zenodo (como arquivo versionado). Cada verbete pode ser aprimorado ao longo do tempo — revisões menores atualizam a versão no Zenodo, revisões maiores atualizam ambos os DOIs com uma nota editorial que descreve a mudança. O DOI do Crossref sempre resolve para a versão mais recente; versões históricas específicas são endereçáveis pelo Zenodo.

Edições futuras estenderão o vocabulário de *Foundations* a áreas adjacentes: dependência espacial, esperança condicional, endogeneidade, viés de variável omitida, identificação e o aparato formal que distingue variáveis *proxy* de variáveis latentes. Sugestões de conceitos que mereçam tratamento são bem-vindas e podem ser enviadas ao editor.

Os seis verbetes desta edição:

1. **Bem heterogêneo** — por que a metodologia de avaliação precisa existir.
2. **Espaço de características** — a representação vetorial dos atributos da cesta.
3. **Função de preços hedônicos** — o mapeamento de equilíbrio das características para os preços.
4. **Preço implícito** — o preço marginal de mercado de cada característica.
5. **Regressão hedônica** — a estimação da função a partir de transações observadas, com um exemplo completo resolvido em R.
6. **Variável latente** — a recuperação de sinal estruturado a partir dos resíduos.

Referências.

- Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Craytor, B. (2025). Residual Constraint Approach: Framework & Protocol. Zenodo. [doi:10.5281/zenodo.14787917](https://doi.org/10.5281/zenodo.14787917)

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (`000-overview.pl`) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```

%% 000-overview.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Overview
%% Title: Overview
%% Author: Bert Craytor, Editor
%% Piece ID: foundations_overview
%% Kind: editorial_piece (not a concept; sits outside the dep graph)
%%
%% The Overview is an editorial piece, not a defined concept. Its .pl
%% companion records the issue's structure (which entries appear in
%% what reading order), the running example, and the reader pathways
%% the editor proposes for readers with different backgrounds.

% =====
% DICTIONARY IMPORT
% =====

:- use_module('.././../././tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(foundations_overview, "Overview", foundations_editorial,
%      "Editorial overview of a Foundations issue. Frames the concepts
%      covered, situates them within appraisal practice, and is not
%      itself a defined concept in the dependency graph.",
%      published, 'dictionary-2026.1').
%
% Plus the six concept terms covered by this issue (heterogeneous_good,
% characteristics_space, hedonic_price_function, implicit_price,
% hedonic_regression, latent_variable).
% -----

:- use_terms([ foundations_overview,
                heterogeneous_good, characteristics_space,
                hedonic_price_function, implicit_price,
                hedonic_regression, latent_variable,
                residual_constraint_approach ]).
:- use_relations([depends_on/2]).

% =====
% PIECE METADATA
% =====

article_piece(foundations_overview).
article_kind(overview).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.overview').

% =====
% ISSUE STRUCTURE: the six concept entries in dependency-ordered sequence
% =====
%
% The published reading order. Each entry assumes vocabulary established
% in the previous ones. This is the same as queries:reading_order/1
% derived from the depends_on graph; recording it here lets readers and
% downstream tools query the issue's structure without invoking the
% topological-sort machinery.

issue_entry(1, heterogeneous_good).

```

```

issue_entry(2, characteristics_space).
issue_entry(3, hedonic_price_function).
issue_entry(4, implicit_price).
issue_entry(5, hedonic_regression).
issue_entry(6, latent_variable).

% Each entry's one-line theme as stated in the Overview's closing list.
entry_theme(heterogeneous_good,
  "Why valuation methodology must exist at all.").
entry_theme(characteristics_space,
  "The vector representation of bundle attributes.").
entry_theme(hedonic_price_function,
  "The equilibrium mapping from characteristics to prices.").
entry_theme(implicit_price,
  "The marginal market price of each characteristic.").
entry_theme(hedonic_regression,
  "Estimating the function from observed transactions, with a full R worked example.").
entry_theme(latent_variable,
  "Recovering structured signal from residuals.").

% =====
% THE RUNNING EXAMPLE
% =====
%
% A single dataset of 8 Pacifica residential comparables runs through
% all six entries. The data lives in kb/pacifica_comps.pl and is
% consulted by each entry that uses it. The Overview's role here is
% only to point at it.

running_example_dataset(pacifica_comps).
running_example_n_comps(8).
running_example_location(pacifica_california).

% Which entries exercise the dataset.
:- consult('../..kb/pacifica_comps').

% =====
% READER PATHWAYS
% =====
%
% The editor proposes four pathways through the issue, calibrated to
% the reader's background. These are advisory orderings; the canonical
% reading order is issue_entry/2 above.

reader_pathway(new_to_hedonic_methodology,
  "Read entries 1-6 in order. Dependencies are real and skipping invites confusion.",
  [heterogeneous_good, characteristics_space, hedonic_price_function,
   implicit_price, hedonic_regression, latent_variable]).

reader_pathway(prior_exposure_to_hedonic_regression,
  "Use as a refresher and atomic reference. Read in any order.",
  [hedonic_regression, hedonic_price_function, implicit_price,
   characteristics_space, heterogeneous_good, latent_variable]).

reader_pathway(interested_in_rca_methodology,
  "Read 006 first for the framing, then 003-005, then 001-002.",
  [latent_variable, hedonic_price_function, implicit_price,
   hedonic_regression, heterogeneous_good, characteristics_space]).

```

```

reader_pathway(interested_in_adjustment_defensibility,
    "Read the 'Why it matters for defensibility' sections in entry order. Forms an implicit short
    ↪ essay.",
    [heterogeneous_good, characteristics_space, hedonic_price_function,
    implicit_price, hedonic_regression, latent_variable]).

% =====
% EDITORIAL POSTURE
% =====

editorial_posture(
    "Foundations is not a textbook department. Each entry is short, written for an audience already
    ↪ fluent in appraisal practice, and designed to give a precise formulation rather than a
    ↪ comprehensive exposition.").

publication_model(
    [ license(cc_by),
      crossref_doi(canonical_citation),
      zenodo_doi(versioned_archive),
      revision_policy("Minor revisions bump Zenodo version; major revisions bump both DOIs with
    ↪ editorial note.") ] ).

% =====
% FUTURE DIRECTIONS
% =====
%
% Concepts the editor flags as deserving treatment in future Foundations
% issues. Recorded here so the journal's planning is machine-queryable.

future_topic(spatial_dependence).
future_topic(conditional_expectation).
future_topic(endogeneity).
future_topic(omitted_variable_bias_extended).
future_topic(identification).
future_topic(proxy_vs_latent_variable).

```

Bem heterogêneo

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/slscsv65](https://doi.org/10.67330/slscsv65)

Definição formal. Bens heterogêneos são bens que não são intercambiáveis porque cada um é uma cesta distinta de características que os compradores valoram separadamente, e cujas características não são negociadas separadamente.

Enquadramento intuitivo. Um bem homogêneo é aquele em que qualquer unidade serve: um barril de petróleo WTI, uma saca de milho amarelo n.º 2, um quilowatt-hora da rede, uma ação de uma determinada empresa. Os padrões de classificação ou as especificações por trás desses bens existem precisamente para tornar as unidades individuais intercambiáveis, de modo que os preços se apliquem à categoria e não à unidade específica. Um bem heterogêneo é aquele em que a unidade específica importa — *esta* casa, não *aquela* casa; *este* carro usado, não o que está ao lado dele.

Duas características tornam um bem heterogêneo no sentido técnico. Primeiro, os compradores se importam com múltiplos atributos distinguíveis que variam entre as unidades do mercado. Segundo, esses atributos são agrupados pela tecnologia de produção do bem; não são compráveis separadamente. Não se pode comprar 200 pés quadrados de área construída isoladamente. Não se pode adquirir um distrito escolar independentemente de um lote de terreno. As características chegam agrupadas, e o mercado fixa preço sobre cestas inteiras, e não sobre as características em si.

O imóvel é o bem heterogêneo canônico e exibe essa propriedade de forma excepcionalmente forte. As características locais são perfeitamente não replicáveis: dois lotes não podem ocupar a mesma localização, de modo que todo imóvel tem ao menos um atributo (suas coordenadas) que o distingue de qualquer outro imóvel do mercado. Atributos físicos, jurídicos e ambientais acrescentam dimensões adicionais de variação a um vetor que, tomado como um todo, é único para cada imóvel.

Onde os avaliadores o encontram. Todo ajuste no método comparativo direto de dados de mercado é um reconhecimento implícito de que o imóvel avaliando e os elementos comparativos são bens heterogêneos. Se os imóveis fossem homogêneos, os preços de venda já seriam diretamente comparáveis; nenhuma planilha de homogeneização seria necessária. A planilha é uma tentativa de conciliar os preços de cestas distintas ajustando as diferenças de composição da cesta. Vale notar o que a planilha *não* faz: ela nunca decompõe um preço de venda em contribuições de valor por característica. A planilha tradicional não tem coluna para contribuições de valor — apenas ajustes, estimados diretamente, característica por característica.

O método da quantificação do custo lida com a heterogeneidade reconstruindo a cesta a partir de seus componentes — fundação, estrutura, acabamentos, depreciação — e precificando cada um separadamente. O método da capitalização da renda lida com ela abstraindo a cesta em um único perfil de fluxo de caixa, com taxas de capitalização e de retorno absorvendo a heterogeneidade do nível da cesta em um multiplicador derivado do mercado. Os três métodos de avaliação são respostas metodológicas ao fato subjacente de que o imóvel é um bem heterogêneo. Eles diferem em como navegam esse fato, não em se o enfrentam.

Os avaliadores também encontram a heterogeneidade informalmente sempre que raciocinam sobre seleção de dados de mercado, análise de pares de vendas ou comparações “entre iguais”. O instinto de encontrar elementos comparativos semelhantes no maior número possível de atributos é a resposta do profissional à heterogeneidade: minimizar a distância entre as cestas do avaliando e do elemento comparativo faz com que a componente não observada da cesta contribua menos para a diferença dos preços de venda.

Por que importa para a defensabilidade. A heterogeneidade dos imóveis é a razão pela qual a metodologia de avaliação precisa existir como disciplina. Se os imóveis fossem intercambiáveis, o valor de mercado seria diretamente observável a partir dos preços anunciados; avaliar se reduziria a consultar uma tabela. Como não são, o valor precisa ser inferido dos preços de imóveis relacionados-mas-não-idênticos, e cada passo de inferência é uma escolha metodológica que precisa ser justificada.

Reconhecer formalmente a heterogeneidade esclarece várias questões práticas das quais depende a defensabilidade sob o USPAP e suas *Standards Rules*. A seleção de dados de mercado é a escolha de unidades no espaço de características suficientemente próximas do avaliando para que os preços implícitos delas recuperados se generalizem. A razoabilidade de um ajuste depende de o preço implícito subjacente ser sustentável a partir da evidência de mercado. A reconciliação entre métodos reflete o julgamento do avaliador sobre qual método navega a heterogeneidade *deste* imóvel de maneira mais crível. Cada uma dessas questões de defensabilidade fica mais nítida quando “bem heterogêneo” é nomeado como a condição subjacente.

Exemplo de avaliação resolvido. Duas vendas recentes no mesmo bairro de Pacifica, Califórnia:

Atributo	Elem. A	Elem. B
GLA (sf)	1.650	1.650
Dormitórios / Banh.	3 / 2	3 / 2
Terreno (sf)	9.000	7.200
Vista	Sim	Não
Preço de venda	\$1.425.000	\$1.280.000

Sob um enquadramento de bens homogêneos, trata-se do mesmo imóvel: GLA idêntica, mesmo número de cômodos, mesmo bairro. Sob o enquadramento de bens heterogêneos, são cestas distintas, que diferem em duas características (terreno e vista) que os compradores valoram separadamente. A diferença de preço de \$145.000 é a contribuição *conjunta* desses dois atributos — dois preços implícitos emaranhados em um único par.

Recuperar os preços implícitos individualmente exige mais pares, idealmente variando um atributo de cada vez. Três ou quatro vendas adicionais que cubram as combinações com vista/sem vista e terreno grande/terreno pequeno começam a identificar os preços implícitos separadamente. Com variação suficiente em elementos comparativos suficientes, a regressão hedônica realiza o desemaranhamento formalmente; com variação limitada, o avaliador fica restrito ao raciocínio por pares de vendas com balizamento, cuja confiabilidade depende explicitamente de quão próximas as cestas estão umas das outras no espaço de características.

O exemplo ilustra por que a heterogeneidade é a condição-raiz: mesmo duas casas “idênticas” não são idênticas, e o aparato metodológico da avaliação existe precisamente para lidar com a distância entre elas.

Referências.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Referências cruzadas: espaço de características; função de preços hedônicos; preço implícito.

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (`001-heterogeneous-good.pl`) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```
%% 001-heterogeneous-good.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 001
%% Title: Heterogeneous Good
%% Author: Bert Craytor
%% Concept: heterogeneous_good
%% Version: 0.1.0-draft
%%
%% This file is the Prolog galley for the article: it encodes the
%% article's main claims, worked example, and inference rules in a form
%% that other Prolog code in the journal can consult, query, and verify
%% against. Read alongside the entry.tex / entry.pdf prose.

% =====
% DICTIONARY IMPORT
% =====
%
% The journal-wide dictionary is the source of truth for term definitions
% and relation schemas. The use_module below pulls those facts in at
% runtime; the comments that follow reproduce the relevant entries so a
% reader skimming this file in isolation can see what's been imported.

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(heterogeneous_good, "Heterogeneous Good", economic_theory,
%     "Heterogeneous goods are goods that are not interchangeable
%     because each is a distinct bundle of characteristics that buyers
%     value separately, and whose characteristics are not separately
%     traded.", published, 'dictionary-2026.1').
%
% term(characteristics_space, "Characteristics Space", economic_theory,
%     "The characteristics space of a class of heterogeneous goods is
%     the vector space whose coordinates index the attributes that
%     buyers value, and in which each good is represented by the
%     vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(implicit_price, "Implicit Price", economic_theory,
%     "The implicit price of a characteristic is the partial derivative
%     of the hedonic price function with respect to that characteristic,
%     evaluated at a point in the characteristics space; it represents
%     the marginal market price of an incremental unit of the
%     characteristic, holding all other characteristics fixed.",
%     published, 'dictionary-2026.1').
```

```

%
% term(hedonic_price_function, ...).
%
% relation(depends_on, 2, "depends on", [foundations_concept, foundations_concept],
%         foundations_methodology, "...", accepted, 'dictionary-2026.1').
% relation(related, 2, ...).
%
% These facts are NOT declared in this file; the use_module above is
% what makes them available at runtime.
% -----

:- use_terms([ % Foundations concepts (terms/foundations.pl)
               heterogeneous_good, characteristics_space, implicit_price,
               hedonic_price_function,
               % Practice terms (terms/valuation-practice.pl)
               real_estate, homogeneous_good,
               sales_comparison_approach, cost_approach, income_approach,
               comp_selection, paired_sales_analysis, adjustment,
               reconciliation, uspap, standards_rule ]).

:- use_relations([related/2, depends_on/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(heterogeneous_good).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.001').

% =====
% ARTICLE-LOCAL VOCABULARY
% =====
%
% Terms the article needs that are NOT yet in the dictionary, and that
% the editor decided are too narrow to promote. These are field names
% for the worked-example property records | not concepts in the
% journal's ontology.

local_term(gla_sqft).
local_term(bedrooms).
local_term(baths).
local_term(lot_sqft).
local_term(view).

% =====
% FORMAL DEFINITION (restated as Prolog rule)
% =====
%
% From the formal definition: a good is heterogeneous IF buyers care
% about multiple distinguishable attributes that vary across items in
% the market, AND those attributes are bundled by production rather
% than separately traded.

%! heterogeneous(?Good) is nondet.
%
% True when Good satisfies both bundling criteria. The two predicates
% buyers_value_multiple_attributes/1 and attributes_bundled_by_production/1
% must be asserted for specific goods elsewhere; here they characterize
% the formal definition itself.

```

```

heterogeneous(Good) :-
    buyers_value_multiple_attributes(Good),
    attributes_bundled_by_production(Good).

% =====
% INTUITIVE FRAMING | CONTRASTS WITH HOMOGENEOUS GOODS
% =====

% Canonical examples of homogeneous goods cited in the article.
canonical_homogeneous_example(wti_crude).
canonical_homogeneous_example(no_2_yellow_corn).
canonical_homogeneous_example(kilowatt_hour).
canonical_homogeneous_example(common_stock_share).

% Their distinguishing feature: grading or specification standards make
% individual items interchangeable.
interchangeable_via_grading(wti_crude).
interchangeable_via_grading(no_2_yellow_corn).
interchangeable_via_specification(kilowatt_hour).
interchangeable_via_specification(common_stock_share).

% Canonical heterogeneous goods.
canonical_heterogeneous_example(real_estate).
canonical_heterogeneous_example(used_car).
canonical_heterogeneous_example(unique_machinery).

% Why real estate exhibits heterogeneity unusually strongly:
% locational characteristics are perfectly non-replicable.
buyers_value_multiple_attributes(real_estate).
attributes_bundled_by_production(real_estate).
non_replicable_attribute(real_estate, location).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% The three classical valuation approaches are all methodological
% responses to heterogeneity. They differ in HOW they navigate the
% bundle, not in WHETHER they confront it.

valuation_approach(sales_comparison_approach).
valuation_approach(cost_approach).
valuation_approach(income_approach).

responds_to_heterogeneity(sales_comparison_approach,
    "Reconciles the prices of distinct bundles by adjusting for composition differences; the grid
    ↪ records only adjustments, estimated directly, and never decomposes a price into feature value
    ↪ contributions.").
responds_to_heterogeneity(cost_approach,
    "Reconstructs the bundle from priced components: foundation, framing, finishes, depreciation.").
responds_to_heterogeneity(income_approach,
    "Abstracts the bundle into a single cash-flow profile; cap/yield rates absorb heterogeneity.").

% Practitioner-level encounters with heterogeneity.
practitioner_response_to_heterogeneity(comp_selection,
    "Minimize bundle distance between subject and comp in characteristics space.").
practitioner_response_to_heterogeneity(paired_sales_analysis,

```

```

    "Isolate an attribute's contribution by holding other attributes constant.").
practitioner_response_to_heterogeneity(adjustment,
    "Quantify the price impact of one differing attribute at a time.").

% =====
% DEFENSIBILITY
% =====
%
% Why naming "heterogeneous good" sharpens USPAP defensibility questions.

defensibility_consequence(comp_selection,
    "Comps must be close enough in characteristics space that recovered implicit prices generalize
    ↪.").
defensibility_consequence(adjustment,
    "Each adjustment must be supportable from market evidence of the implicit price.").
defensibility_consequence(reconciliation,
    "Choice among approaches reflects which best navigates THIS property's heterogeneity.").

% Root claim: valuation methodology exists as a discipline BECAUSE
% real estate is heterogeneous. If it were homogeneous, value would be
% directly observable from posted prices.
root_claim(
    "Valuation methodology exists as a discipline because real estate is heterogeneous; in the
    ↪homogeneous case, value would be directly observable.").

% =====
% WORKED EXAMPLE | PACIFICA COMPS
% =====
%
% Two recent sales in the same Pacifica, California neighborhood. The
% example demonstrates that even "identical" houses are not identical:
% the $145,000 price spread is the joint contribution of differences in
% lot size and view, two implicit prices entangled in a single pair.

% The 8-comp Pacifica dataset is shared across entries 001-006 and
% lives in kb/pacifica_comps.pl. Consulting it here brings property/1,
% in_market/2, attribute_value/3, and sale_price/2 into this article's
% namespace.
:- consult('.../kb/pacifica_comps').

% --- Derived facts from the example -----

% The attributes that differ between the two comps.
differing_attribute(comp_a, comp_b, lot_sqft).
differing_attribute(comp_a, comp_b, view).

% The price differential and what it's attributable to.
price_differential(comp_a, comp_b, 145000).
attributed_to(price_differential(comp_a, comp_b, 145000),
    [implicit_price_of(lot_sqft), implicit_price_of(view)]).

% --- Article's argument about identification -----

% A single pair cannot separately identify the two implicit prices |
% they are entangled in the single $145K observation.
joint_identification_only(price_differential(comp_a, comp_b, 145000),
    [lot_sqft, view]).

% Separate identification requires more pairs varying one attribute at

```

```

% a time. With sufficient variation, hedonic regression performs the
% disentanglement formally; with limited variation, the appraiser uses
% bracketed paired-sales reasoning, whose reliability depends on how
% close the bundles are in characteristics space.
identification_strategy(few_pairs,
    "Bracketed paired-sales reasoning. Reliability depends on bundle proximity in characteristics
    ↪ space.").
identification_strategy(many_pairs_with_variation,
    "Hedonic regression (see hedonic_regression).").

% =====
% INFERENCE RULES
% =====

%! comparable(?A, ?B) is nondet.
%
% A and B are comparable when they are distinct properties in the same
% market. This is the operational definition of "comp" used implicitly
% in the article.

comparable(A, B) :-
    property(A),
    property(B),
    A \== B,
    in_market(A, M),
    in_market(B, M).

%! attribute_difference(?A, ?B, ?Attr, ?DA, ?DB) is nondet.
%
% A and B differ on Attr, with DA the value at A and DB the value at B.
% Use this to enumerate the dimensions on which two comps actually vary.

attribute_difference(A, B, Attr, DA, DB) :-
    comparable(A, B),
    attribute_value(A, Attr, DA),
    attribute_value(B, Attr, DB),
    DA \== DB.

%! identifiable_implicit_price(?Attr) is nondet.
%
% An attribute's implicit price is identifiable from the observed comps
% if there exists at least one pair of comparables differing ONLY in
% that attribute. With our two-comp example this fails for lot_sqft AND
% view (both differ in the same pair), which is exactly the article's
% point.

identifiable_implicit_price(Attr) :-
    comparable(A, B),
    findall(X, attribute_difference(A, B, X, _, _), Diffs),
    Diffs == [Attr].

% =====
% CROSS-REFERENCES
% =====
%
% Concepts the article cross-references (corresponding to the
% \foundationsref{} macros in entry.tex).

cross_reference(characteristics_space).
```

```
cross_reference(hedonic_price_function).  
cross_reference(implicit_price).
```

Espaço de características

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/3yymh010](https://doi.org/10.67330/3yymh010)

Definição formal. O espaço de características de uma classe de bens heterogêneos é o espaço vetorial cujas coordenadas indexam os atributos que os compradores valoram, e no qual cada bem é representado pelo vetor de seus níveis de atributo.

Enquadramento intuitivo. O verbete anterior estabeleceu que um bem heterogêneo é uma cesta distinta de características. A questão seguinte é natural: como descrever e comparar formalmente tais cestas? A resposta, devida a Lancaster (1966), é tratar cada bem como um ponto de um espaço vetorial cujos eixos são as características que interessam aos compradores.

Nessa representação, um imóvel residencial de Pacífica deixa de ser “uma casa em Park Pacífica” e passa a ser um vetor com coordenadas para área construída, área do terreno, vista, estado de conservação, distrito escolar, distância até o oceano e assim por diante. Dois imóveis que compartilhassem todos esses valores de atributo ocupariam o mesmo ponto no espaço de características; imóveis que diferem em qualquer atributo ocupam pontos distintos. A geometria do espaço codifica quão próximos ou quão diferentes são dois imóveis, de um modo que capta todas as dimensões sobre as quais os compradores fazem distinções.

A dimensionalidade do espaço depende de quais atributos são relevantes para o mercado em estudo. Para uma avaliação residencial típica, o espaço relevante pode ter uma dúzia de dimensões: físicas (GLA, terreno, dormitórios, banheiros, idade, padrão construtivo, estado de conservação), locais (distrito escolar, vista, proximidade de amenidades ou desamenidades) e jurídico-econômicas (zoneamento, servidões). Para uma avaliação de imóvel comercial, as dimensões relevantes se deslocam para atributos de fluxo de caixa (estrutura de locação, qualidade do locatário, ocupação), mas a ideia geométrica subjacente é a mesma.

O que torna o espaço de características útil é que ele converte uma intuição qualitativa do profissional (“estes elementos comparativos são parecidos com o avaliando”) em uma afirmação quantitativa (“estes elementos comparativos estão próximos do avaliando no espaço de características”). Uma vez definida a distância, toda decisão do método comparativo direto de dados de mercado pode ser examinada como uma afirmação sobre localização no espaço.

Onde os avaliadores o encontram. Os avaliadores trabalham no espaço de características o tempo todo, geralmente sem nomeá-lo. Cada linha de uma planilha de homogeneização é um eixo coordenado; cada elemento comparativo é um vetor-linha de valores de coordenadas; cada ajuste é uma tentativa de mover-se ao longo de um eixo mantendo os demais fixos.

Os oito elementos comparativos de Pacífica que usaremos ao longo desta edição ilustram a construção diretamente. Cada elemento é um ponto em um espaço de características de quatro dimensões, gerado por GLA, área do terreno, vista e estado de conservação:

Elem.	GLA (sf)	Terreno (sf)	Vista	Conserv.	Preço de venda
A	1.650	9.000	sim	bom	\$1.425.000
B	1.650	7.200	não	bom	\$1.280.000
C	1.850	7.500	não	bom	\$1.407.500
D	1.450	7.000	não	bom	\$1.155.000
E	1.700	8.000	sim	excelente	\$1.470.000
F	1.550	7.100	não	médio	\$1.177.500
G	1.700	10.500	não	bom	\$1.392.500
H	1.600	7.300	não	bom	\$1.312.500

Os elementos A e B do verbete anterior são as duas primeiras linhas; os seis restantes ampliam o exemplo de maneiras que exploraremos no restante da edição.

Cada elemento é um ponto em $\mathbb{R}^4$:

- O elemento A está em (1650, 9000, 1, 2).
- O elemento B está em (1650, 7200, 0, 2).
- O elemento G está em (1700, 10500, 0, 2).

As coordenadas de vista e de estado de conservação são codificadas numericamente (vista como $\{0, 1\}$, conservação como $\{1, 2, 3\}$); o preço *não* é uma coordenada do espaço de características, mas uma função da posição dentro dele.

Por que importa para a defensabilidade. O arcabouço do espaço de características aguça o que a seleção de dados de mercado de fato é. Um elemento comparativo não é apenas “um imóvel parecido”, mas um ponto da região relevante do espaço de características próximo do avaliando. Quanto mais próximo do avaliando estiver o elemento no espaço, menores os ajustes exigidos e mais confiável se torna o valor indicado.

Esse reenquadramento expõe várias questões práticas que a planilha informal obscurece:

- *Quais dimensões importam?* A escolha dos eixos é uma decisão de modelagem. Incluir “vista” como eixo só faz sentido se o mercado precifica a vista de forma distinta; incluir “orientação segundo o Feng Shui” só faz sentido se algum segmento de mercado a precifica. A avaliação defensável exige defender os eixos, e não apenas os valores ao longo deles.
- *Qual métrica de distância se aplica?* Dois elementos podem diferir em \$200.000 de preço porque estão distantes um do outro no espaço de características, mas o cálculo da distância exige um escalonamento para cada eixo. A distância euclidiana não transformada trata um pé quadrado de GLA como equivalente a um pé quadrado de terreno, coisa que o mercado não faz.
- *Os elementos estão em uma região em que a função de preços é bem comportada?* Se o avaliando se situa em um canto pouco amostrado do espaço de características, nenhum elemento comparativo está próximo dele e a indicação de valor é necessariamente extrapolativa.

Exemplo de avaliação resolvido. Os oito elementos comparativos de Pacífica ocupam uma região relativamente compacta do espaço de características de quatro dimensões. Considere a questão de qual elemento é o “mais próximo” de um imóvel avaliando hipotético em (1,700, 7,500, 1, 2) — 1.700 sf, terreno de 7.500 sf, com vista, em bom estado de conservação.

Se usarmos ingenuamente a distância euclidiana com eixos não escalonados, a dimensão do terreno domina, porque seus valores (7.000–10.500) são numericamente muito maiores do que os de vista (0–1) ou de conservação (1–3). Distâncias até o avaliando:

Elem.	Distância (euclidiana não escalonada)	Observações
A	1.501	vista coincide; terreno domina
B	308	vista difere; terreno próximo
C	230	vista difere; terreno próximo
D	559	vista difere; GLA fora
E	502	vista coincide; conservação difere
F	422	vista e conservação diferem
G	3.000	vista difere; terreno distante
H	224	vista difere; o mais próximo de todos

Pela distância bruta, o elemento H é o vizinho mais próximo. Mas o elemento H não tem vista, e o avaliando tem. A métrica ingênua nos induziu ao erro: a escala numérica da dimensão do terreno sufocou a dimensão categórica da vista, que na verdade importa mais para o preço.

Uma métrica de distância defensável escalonaria cada eixo por seu preço implícito — grosso modo, por quantos dólares vale uma unidade de variação naquele eixo. Uma vez que os eixos sejam escalonados pelo preço implícito, a vista (que vale ~\$100.000) deixa de ser dominada pelo terreno (~\$25 por pé quadrado). O elemento A, que coincide com o avaliando quanto à vista, emerge então como o elemento comparativo primário apropriado.

É exatamente isso que o maquinário de preços implícitos desenvolvido nos verbetes seguintes fornece: um escalonamento fundamentado do espaço de características, que transforma a “proximidade” de um palpite geométrico em um cálculo derivado do mercado.

Referências.

- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.

Referências cruzadas: bem heterogêneo; função de preços hedônicos; preço implícito.

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (002-characteristics-space.pl) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```
%% 002-characteristics-space.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 002
%% Title: Characteristics Space
%% Author: Bert Craytor
%% Concept: characteristics_space
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').
```

```

% --- From the dictionary (loaded via :- use_module above) -----
% term(characteristics_space, "Characteristics Space", economic_theory,
%      "The characteristics space of a class of heterogeneous goods is
%      the vector space whose coordinates index the attributes that
%      buyers value, and in which each good is represented by the
%      vector of its attribute levels.", published, 'dictionary-2026.1').
%
% term(heterogeneous_good, ...).
% term(hedonic_price_function, ...).
% term(implicit_price, ...).
%
% generalized_by(characteristics_space, lancaster_1966).
% depends_on(characteristics_space, heterogeneous_good).
% -----

:- use_terms([ characteristics_space, heterogeneous_good,
                hedonic_price_function, implicit_price,
                comp_selection, sales_comparison_approach,
                real_estate ]).
:- use_relations([related/2, depends_on/2, generalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(characteristics_space).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.002').

% =====
% LOCAL VOCABULARY
% =====
%
% Narrow article-local field names + the dimensions of the worked
% example's characteristics space.

local_term(dimension).
local_term(coordinate).
local_term(euclidean_distance).
local_term(distance_metric).
local_term(nearest_neighbor).

% =====
% FORMAL ATTRIBUTION
% =====

% The construction is due to Lancaster (1966), as noted in citations.
attributed_construction(characteristics_space, lancaster_1966).

% =====
% THE ARTICLE'S CLAIM ABOUT REPRESENTATION
% =====
%
% A heterogeneous good becomes a point in a vector space. Each axis is
% a characteristic buyers value; each good is the vector of its
% attribute levels.

representation_of(heterogeneous_good, point_in_vector_space).

```

```

% Examples of dimensions relevant to residential appraisal.
residential_dimension(gla_sqft).
residential_dimension(lot_sqft).
residential_dimension(view).
residential_dimension(cond).
residential_dimension(bedrooms).
residential_dimension(baths).
residential_dimension(age).
residential_dimension(quality).
residential_dimension(school_district).
residential_dimension(distance_to_amenity).
residential_dimension(zoning).

% Examples of dimensions relevant to commercial appraisal | included
% to make the article's point that the dimensions shift with market.
commercial_dimension(lease_structure).
commercial_dimension(tenant_quality).
commercial_dimension(occupancy).
commercial_dimension(cash_flow_profile).

% =====
% APPRAISER ENCOUNTERS
% =====
%
% Mapping appraisal practice onto the characteristics-space framework.

practice_to_space_correspondence(
    sales_comparison_grid_line, coordinate_axis,
    "A grid line is an axis.").
practice_to_space_correspondence(
    comp_row, row_vector,
    "A comp is a vector of coordinate values.").
practice_to_space_correspondence(
    adjustment, movement_along_axis,
    "An adjustment moves along one axis holding others fixed.").

% =====
% WORKED EXAMPLE
% =====
%
% The 8 Pacifica comps are points in a 4D characteristics space spanned
% by GLA, lot_sqft, view, cond_numeric. Bring in the shared dataset.

:- consult('../..kb/pacifica_comps').

% Axes of the chosen space for this article's analysis.
axis(gla_sqft).
axis(lot_sqft).
axis(view).
axis(cond_numeric).

% Numeric encoding for non-numeric axes.
encode(view, no, 0).
encode(view, yes, 1).
encode(cond_numeric, 1, 1).
encode(cond_numeric, 2, 2).
encode(cond_numeric, 3, 3).

```

```

%! coord(?Property, ?Axis, ?Numeric) is nondet.
%
% Numeric coordinate of Property along Axis. Handles symbolic attribute
% values by encoding them.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V ; encode(A, V, N) ).

%! point(?Property, -Vector) is det.
%
% Vector representation of Property in the 4D characteristics space.

point(P, [GLA, Lot, View, Cond]) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond).

% =====
% THE HYPOTHETICAL SUBJECT
% =====
%
% The article tests distance metrics against this hypothetical property.

hypothetical_subject(subject).
attribute_value(subject, gla_sqft, 1700).
attribute_value(subject, lot_sqft, 7500).
attribute_value(subject, view, yes).
attribute_value(subject, cond_numeric, 2).

% =====
% NAIVE EUCLIDEAN DISTANCE (the article's straw distance metric)
% =====
%
% Computes the unscaled Euclidean distance between Property and the
% hypothetical subject. Used to demonstrate that unscaled distances
% mislead: the lot dimension dominates because its numeric range is
% much larger than view (0..1) or cond (1..3).

naive_distance_to_subject(P, D) :-
    point(P, V_P),
    point(subject, V_S),
    sum_of_squares(V_P, V_S, S),
    D is sqrt(S).

sum_of_squares([], [], 0).
sum_of_squares([X|Xs], [Y|Ys], S) :-
    sum_of_squares(Xs, Ys, S0),
    D is X - Y,
    S is S0 + D * D.

% =====
% THE ARTICLE'S CENTRAL OBSERVATION
% =====
%
% Under naive Euclidean distance, Comp H is the nearest neighbor. But
% Comp H has no view and the subject does; view matters more to price
% than the naive metric admits. The fix (developed in entries 003-005)

```

```

% is to scale each axis by its implicit price.

claim(naive_metric_misleads,
    "Unscaled Euclidean distance crowds out small-range axes (like view) under large-range axes (
    ↪ like lot_sqft). A market-defensible distance scales each axis by its implicit price.").

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====
%
% Three questions the characteristics-space framing makes explicit.

defensibility_question(which_dimensions,
    "The choice of axes is a modeling decision. Defending an appraisal requires defending the axes,
    ↪ not just the values on them.").
defensibility_question(distance_metric,
    "Untransformed Euclidean distance treats one sqft of GLA as equivalent to one sqft of lot,
    ↪ which the market does not. Distance requires a scaling per axis.").
defensibility_question(extrapolation_region,
    "If the subject sits in a sparsely-sampled corner of the space, no comp is truly close and the
    ↪ indication is extrapolative.").

% =====
% INFERENCE RULES
% =====

%! same_point(?A, ?B) is nondet.
%
% Two properties occupy the same point in the characteristics space.
% (Doesn't occur in the running example, but the article notes that it
% would if all coordinates matched.)

same_point(A, B) :-
    point(A, V),
    point(B, V),
    A \== B.

%! nearest_neighbor_by_naive_distance(-P) is semidet.
%
% The comp closest to the subject under the naive Euclidean metric.
% Used to illustrate the article's point that the naive metric chooses
% Comp H, not Comp A.

nearest_neighbor_by_naive_distance(P):-
    findall(D-Q,
        ( property(Q), Q \== subject, naive_distance_to_subject(Q, D) ),
        Pairs),
    keysort(Pairs, [_-P|_]).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).

```

Função de preços hedônicos

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/9x1a0489](https://doi.org/10.67330/9x1a0489)

Definição formal. A função de preços hedônicos é o mapeamento de equilíbrio dos pontos do espaço de características de um bem heterogêneo para os preços de transação, formado como a envoltória das funções de lance dos compradores e das funções de oferta dos vendedores em um mercado competitivo.

Enquadramento intuitivo. Temos agora dois ingredientes: um bem heterogêneo (uma cesta de características) e um espaço de características (um espaço vetorial em que cada cesta é um ponto). A função de preços hedônicos combina os dois. É a função $p(\mathbf{z})$ que atribui um preço a cada ponto $\mathbf{z}$ do espaço de características, dizendo-nos quanto o mercado paga por uma cesta com aquela combinação particular de atributos.

A percepção central de Rosen (1974) é que $p(\mathbf{z})$ não é escolhida por nenhum comprador ou vendedor isolado. Ela emerge do equilíbrio de mercado. Cada comprador tem uma *função de lance* que expressa o máximo que pagaria por qualquer cesta, dadas suas preferências e sua renda. Cada vendedor tem uma *função de oferta* que expressa o mínimo que aceitaria, dados seus custos e suas alternativas. A função de preços hedônicos é a envoltória desses lances e ofertas — a fronteira ao longo da qual compradores e vendedores efetivamente transacionam.

Para imóveis, $p(\mathbf{z})$ é a resposta à pergunta: se você especificasse uma cesta completa de características, que preço o mercado produziria para ela? A resposta não é arbitrária. Ela reflete as preferências agregadas de todos os compradores do mercado e os custos e alternativas agregados de todos os vendedores, mediados pelo pareamento de equilíbrio entre compradores e cestas.

A função de preços hedônicos é o objeto teórico central da análise hedônica. Os preços implícitos são derivadas parciais dela. A regressão hedônica é uma estimativa empírica dela. A defensabilidade de qualquer ajuste do método comparativo direto de dados de mercado reduz-se, em última instância, a uma afirmação sobre essa função.

Onde os avaliadores o encontram. Os avaliadores raciocinam sobre $p(\mathbf{z})$ continuamente, de novo em geral sem nomeá-la. Várias atividades da prática são manipulações implícitas da função de preços hedônicos:

Ajustes no método comparativo direto de dados de mercado. Quando um avaliador ajusta o elemento B para cima em \$100.000 para levar em conta o fato de que o avaliando tem vista e o elemento B não, o avaliador está afirmando algo sobre $p(\mathbf{z})$: a saber, que, mantidas fixas todas as demais características, a função aumenta \$100.000 quando a característica vista passa de 0 para 1. Trata-se de uma afirmação sobre a inclinação local de p na dimensão da vista.

Reconciliação entre métodos. O método da quantificação do custo constrói uma indicação de valor de baixo para cima, somando os custos dos componentes de uma cesta replicada. O método da capitalização da renda deriva o valor dos fluxos de caixa, que em equilíbrio refletem o preço de cestas que produzem fluxos semelhantes. O método comparativo direto de dados de mercado lê o preço diretamente de transações observadas de cestas semelhantes. Em um mercado que funciona

bem, os três métodos devem convergir para o mesmo ponto de $p(\mathbf{z})$. Quando divergem, o avaliador precisa julgar qual método mediu a função de forma mais confiável para a localização do avaliando no espaço de características.

Análise de mercado e ajustes temporais. A função de preços hedônicos se desloca ao longo do tempo, à medida que oferta, demanda, financiamento e expectativas mudam. Uma taxa de valorização anual de 6 por cento é uma afirmação sobre como $p(\mathbf{z})$ se deslocou como um todo; a constatação de que imóveis com prêmio de vista se valorizaram mais rápido do que imóveis sem vista é uma afirmação de que a função se deslocou de forma desigual pelo espaço de características.

Por que importa para a defensabilidade. Nomear explicitamente a função de preços hedônicos esclarece várias questões fundacionais de defensabilidade:

A função existe? O arcabouço de Rosen pressupõe um mercado competitivo com transações suficientes para que uma função de equilíbrio seja significativa. Em mercados rarefeitos, em mercados que mudam rapidamente ou em submercados com conjuntos idiossincráticos de compradores, a função pode ser mal definida ou instável. Um avaliador que trabalhe em tal mercado deve reconhecer que o maquinário inferencial está operando sobre uma fundação teórica frágil.

A função é a mesma em toda parte? O arcabouço de Rosen produz uma função por segmento de mercado, e não uma função global. Pacífica e São Francisco são mercados distintos, com funções hedônicas distintas; casas unifamiliares e apartamentos em condomínio são submercados distintos mesmo dentro de uma única cidade. Um avaliador que confunde segmentos está implicitamente supondo que uma única função abrange ambos, o que em geral é falso.

Que forma funcional p assume? A função não precisa ser linear, log-linear nem de qualquer outra forma específica. Rosen mostrou apenas que p é a envoltória de lances e ofertas; ele não especificou seu formato. Empiricamente, p exibe não linearidades (o valor marginal de um dormitório adicional muitas vezes não é constante em relação à GLA), interações (o prêmio de vista depende da orientação do terreno) e limiares (uma cozinha abaixo de certo patamar de padrão pode atrair um deságio desproporcional à sua nota de estado de conservação). A estimação defensável exige ou uma forma funcional sustentável a partir do mercado, ou um estimador flexível capaz de recuperar o formato a partir dos dados.

Exemplo de avaliação resolvido. Para o bairro de Pacífica em estudo, trabalharemos com uma forma funcional aditiva simples, que o próximo verbete exercitará mais plenamente:

$$p(\mathbf{z}) = \beta_0 + \beta_{\text{GLA}} z_{\text{GLA}} + \beta_{\text{lot}} z_{\text{lot}} + \beta_{\text{view}} z_{\text{view}} + \beta_{\text{cond}} z_{\text{cond}}.$$

Essa forma embute duas suposições fortes: que a contribuição de cada característica é linear em seu nível e que as características não interagem. Ambas as suposições são testáveis e frequentemente violadas; adotamo-las aqui como ponto de partida.

Para os oito elementos comparativos do exemplo contínuo desta edição, nossa análise nos verbetes seguintes estimará valores de β que produzem uma função mais ou menos assim:

Parâmetro	Valor estimado
β_0 (intercepto)	\$80.777
β_{GLA} (por sf)	\$575
β_{lot} (por sf)	\$22,70
β_{view} (sim/não)	\$86.179
β_{cond} (por passo)	\$49.824

Inserida na forma aditiva, essa função prevê um preço para qualquer ponto do espaço de características de quatro dimensões. Substituindo as coordenadas do elemento A, obtém-se um preço previsto próximo dos \$1.425.000 observados; substituindo o avaliando hipotético em (1700, 7500, 1, 2), obtém-se uma indicação de valor para aquele imóvel.

O objetivo do verbete não são esses números específicos, mas o movimento conceitual: *a função existe, em princípio, antes de a estimarmos*. A função é uma propriedade do mercado. Nossa estimação é uma tentativa de recuperá-la. A recuperação pode ser parcial, viesada ou ruidosa, e os três verbetes seguintes tratam de como fazer essa recuperação de forma responsável.

Referências.

- Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Triplett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

Referências cruzadas: bem heterogêneo; espaço de características; preço implícito; regressão hedônica.

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (003-hedonic-price-function.pl) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```
%% 003-hedonic-price-function.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 003
%% Title: Hedonic Price Function
%% Author: Bert Craytor
%% Concept: hedonic_price_function
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_price_function, "Hedonic Price Function", economic_theory,
%      "The hedonic price function is the equilibrium mapping from
%      points in the characteristics space of a heterogeneous good to
%      transaction prices, formed as the envelope of buyer bid
%      functions and seller offer functions in a competitive market.",
%      published, 'dictionary-2026.1').
%
% term(bid_function, "Bid Function", economic_theory, ...).
% term(offer_function, "Offer Function", economic_theory, ...).
% term(sales_comparison_approach, ...).
% term(cost_approach, ...).
% term(income_approach, ...).
```

```

%
% formalized_by(hedonic_price_function, rosen_1974).
% depends_on(hedonic_price_function, heterogeneous_good).
% depends_on(hedonic_price_function, characteristics_space).
% -----

:- use_terms([ hedonic_price_function, heterogeneous_good,
               characteristics_space, implicit_price, hedonic_regression,
               bid_function, offer_function,
               sales_comparison_approach, cost_approach, income_approach,
               adjustment, reconciliation ]).
:- use_relations([related/2, depends_on/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_price_function).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.003').

% =====
% LOCAL VOCABULARY
% =====

local_term(equilibrium).
local_term(envelope).
local_term(functional_form).
local_term(market_segment).
local_term(linearity).
local_term(separability).
local_term(time_adjustment).
local_term(coefficient).
local_term(intercept).

% =====
% FORMAL ATTRIBUTION
% =====

attributed_formalization(hedonic_price_function, rosen_1974).

% =====
% CONSTRUCTION: ENVELOPE OF BID AND OFFER FUNCTIONS
% =====
%
% The Rosen (1974) construction:  $p(z)$  is the envelope of buyer bid
% functions and seller offer functions.

primitive_of_construction(bid_function, buyer).
primitive_of_construction(offer_function, seller).

constructed_as_envelope_of(hedonic_price_function,
                           [bid_function, offer_function]).

emerges_from(hedonic_price_function, market_equilibrium).

% Property that flows from the construction: the function is not
% chosen by any single market participant.
property(hedonic_price_function, not_set_by_individual_participant).

```

```

property(hedonic_price_function, equilibrium_object).
property(hedonic_price_function, defined_per_market_segment).

% =====
% APPRAISER ENCOUNTERS | IMPLICIT MANIPULATIONS OF p(z)
% =====

implicit_manipulation_of_hedonic(
    adjustment,
    sales_comparison_approach,
    "Asserts a local slope of p in the relevant characteristic dimension.").
implicit_manipulation_of_hedonic(
    reconciliation,
    multi_approach_valuation,
    "Three approaches should converge on the same point of p; divergence requires editorial
    ↪ judgment.").
implicit_manipulation_of_hedonic(
    time_adjustment,
    market_analysis,
    "p shifts over time; appreciation rates are statements about that shift.").

% =====
% DEFENSIBILITY QUESTIONS
% =====
%
% The article names three questions the hedonic-price-function framing
% sharpens.

defensibility_question(function_existence,
    "In thin or rapidly-changing markets, the equilibrium function may be poorly defined; the
    ↪ inferential machinery operates on a shaky foundation.").
defensibility_question(segment_specificity,
    "One function per market segment, not one function globally. Conflating segments is usually
    ↪ wrong.").
defensibility_question(functional_form,
    "Rosen showed only that p is an envelope, not its shape. Empirically nonlinear with
    ↪ interactions and thresholds; estimation requires defending the form.").

% =====
% WORKED EXAMPLE: ADDITIVE LINEAR SPECIFICATION
% =====
%
% For the Pacifica neighborhood, the article adopts the additive linear
% specification:
%  $p(z) = \beta_0 + \beta_{\text{gla}}z_{\text{gla}} + \beta_{\text{lot}}z_{\text{lot}} + \beta_{\text{view}}z_{\text{view}} + \beta_{\text{cond}}z_{\text{cond}}$ 
%
% The next entry (004) takes the coefficients as implicit prices; entry
% 005 estimates them by hedonic regression.

:- consult('../..kb/pacifica_comps').

specification_form(additive_linear).
specification_assumption(linearity_in_levels,
    "Contribution of each characteristic is linear in its level.").
specification_assumption(no_interactions,
    "Characteristics do not interact in the price function.").
specification_caveat(
    "Both assumptions are testable and frequently violated; adopted here as a starting point.").

```

```

% Coefficient values reported in the article (estimated in entry 005,
% asserted here so this entry's worked example can be queried).
coefficient(intercept,      80777.07).
coefficient(beta_gla,       574.93).
coefficient(beta_lot,       22.70).
coefficient(beta_view,     86179.03).
coefficient(beta_cond,     49824.33).

%! predicted_price(?Property, -Predicted) is det.
%
% The fitted additive-linear hedonic price function applied to a
% property's characteristics. Symbolic view/cond are mapped to {0,1}
% and {1,2,3} respectively.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),
    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    coefficient(intercept, B0),
    coefficient(beta_gla, Bg),
    coefficient(beta_lot, B1),
    coefficient(beta_view, Bv),
    coefficient(beta_cond, Bc),
    Predicted is B0 + Bg * GLA + B1 * Lot + Bv * View + Bc * Cond.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V
    ; A == view, V == yes -> N = 1
    ; A == view, V == no  -> N = 0
    ; V = N
    ).

% =====
% THE CENTRAL CONCEPTUAL MOVE
% =====
%
% The function exists, in principle, before estimation. The function is
% a property of the market; estimation is recovery.

claim(function_precedes_estimation,
    "The hedonic price function is a property of the market and exists conceptually prior to any
    ↪ specific fitted model. Estimation is recovery, and recovery may be partial, biased, or noisy
    ↪ .").

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(implicit_price).
cross_reference(hedonic_regression).

```

Preço implícito

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/md6wke17](https://doi.org/10.67330/md6wke17)

Definição formal. O preço implícito de uma característica é a derivada parcial da função de preços hedônicos em relação a essa característica, avaliada em um ponto do espaço de características; ele representa o preço marginal de mercado de uma unidade incremental da característica, mantidas fixas todas as demais características.

Enquadramento intuitivo. A função de preços hedônicos $p(\mathbf{z})$ atribui um preço a cada cesta. A maior parte das questões práticas da avaliação, porém, não trata do preço de uma cesta completa, mas do preço de um *componente* específico de uma cesta. Quanto vale a vista? Quanto acrescentam mais cem pés quadrados? Quanto contribui a diferença entre estado de conservação “bom” e “excelente”?

O preço implícito responde a essas perguntas. O termo tem origem em Rosen (1974), que mostrou que os preços observados de produtos diferenciados, tomados junto com suas características, “definem um conjunto de preços implícitos ou ‘hedônicos’”; Lancaster (1966) forneceu a visão subjacente dos bens como cestas de características. Para cada característica z_i , o preço implícito é a taxa à qual $p(\mathbf{z})$ varia quando z_i varia, mantendo-se fixas todas as demais características:

$$\pi_i(\mathbf{z}) = \frac{\partial p(\mathbf{z})}{\partial z_i}.$$

O preço implícito tem unidades de dólares por unidade da característica — dólares por pé quadrado de GLA, dólares por pé quadrado de terreno, dólares por passo de estado de conservação. Para uma característica binária como a vista, o preço implícito é o prêmio integral pago para passar de sem vista (0) para com vista (1).

A palavra “implícito” capta o fato central: esses preços *não* são anunciados. Ninguém se oferece para vender “um pé quadrado de GLA” por \$575. O preço está implícito nos preços de equilíbrio das cestas que *são* anunciados, e o aparato hedônico existe para extraí-lo.

Os preços implícitos são o produto do arcabouço hedônico que interessa à avaliação. O ajuste em uma planilha de homogeneização do método comparativo direto de dados de mercado é, em princípio, uma estimativa de um preço implícito multiplicada por uma diferença de característica. Todo ajuste é uma aposta sobre o valor de π_i no ponto relevante do espaço de características.

Onde os avaliadores o encontram. A planilha de homogeneização do método comparativo é uma tabela de aplicações de preços implícitos, mesmo quando os avaliadores não os nomeiam assim. Uma célula da planilha que ajusta um elemento comparativo em “+\$10.000 por 200 sf de GLA a menos que o avaliando” está afirmando que o preço implícito da GLA, na localização relevante do espaço de características, é de \$50 por pé quadrado. Uma célula que ajusta “\$100.000 por vista inferior” está afirmando um preço implícito de \$100.000 para a característica binária vista. Toda a defensabilidade do método comparativo direto de dados de mercado repousa sobre a possibilidade de sustentar esses preços implícitos a partir do mercado. Na prática, contudo, os avaliadores estimam

os ajustes diretamente — a partir de pares de vendas, da experiência ou da prática dos pares — sem jamais calcular as contribuições de valor subjacentes; a planilha tradicional registra apenas os ajustes em si. A leitura em termos de preço implícito é uma interpretação do que esses ajustes afirmam, não uma descrição de como são obtidos.

Essa nomeação também esclarece por que a análise de pares de vendas é a técnica fundamental para a obtenção de ajustes. Dois elementos comparativos que diferem exatamente uma característica fornecem uma estimativa direta do preço implícito dessa característica: a diferença de preço dividida pela diferença da característica. A confiabilidade da análise de pares de vendas depende de os dois elementos estarem realmente próximos o bastante no espaço de características para que o preço implícito deles recuperado se generalize ao avaliando.

Três propriedades dos preços implícitos que importam para a prática:

São locais. O preço implícito da GLA no ponto de 1.650 sf não é necessariamente o mesmo que no ponto de 2.500 sf. A função de preços hedônicos é geralmente não linear, de modo que suas derivadas parciais variam com a localização no espaço de características. Um ajuste obtido de uma análise de pares de vendas em 1.500 sf não se aplica necessariamente a um avaliando de 3.000 sf.

São condicionais. O preço implícito da vista em um submercado de frente para o mar não é o mesmo de um submercado do interior. A derivada parcial é tomada mantendo fixas as demais características, mas o valor dessa derivada depende de *onde* as demais características são mantidas fixas.

São dependentes do mercado. Preços implícitos são preços de equilíbrio e, portanto, refletem as preferências do conjunto de compradores à época das transações. Uma mudança de mercado pode alterar os preços implícitos mesmo que as próprias características permaneçam inalteradas. Os ajustes temporais são, implicitamente, admissões de que os preços implícitos de ontem podem não ser os de hoje.

Por que importa para a defensabilidade. O enquadramento por preços implícitos expõe questões de defensabilidade dos ajustes que a planilha informal pode obscurecer:

Onde você está no espaço de características? Um preço implícito só é defensável se for estimado nas proximidades da localização do avaliando no espaço de características. Um prêmio de vista obtido de elementos comparativos do interior e aplicado a um avaliando de frente para o mar é uma extrapolação por uma região do espaço de características em que o comportamento da função é desconhecido.

Qual é a geometria local? Se duas características interagem — por exemplo, se a vista vale mais em um terreno maior — então tratá-las separadamente como aditivas é uma especificação incorreta. O preço implícito da vista depende da área do terreno, e um ajuste que ignora a interação subestima a estrutura efetiva da função.

Como o preço implícito foi recuperado? Uma análise de pares de vendas, um coeficiente de regressão, uma regra prática do setor e um “julgamento profissional” apontam todos para a mesma grandeza — o preço implícito — mas com estatutos epistêmicos muito diferentes. A defensabilidade do avaliador depende de dizer qual fonte produziu o preço implícito usado em cada ajuste.

Exemplo de avaliação resolvido. Para o exemplo contínuo de Pacífica, com a especificação aditiva adotada no verbete anterior, os preços implícitos são simplesmente os coeficientes da função linear:

$$\pi_{\text{GLA}} = \beta_{\text{GLA}}, \quad \pi_{\text{lot}} = \beta_{\text{lot}}, \quad \pi_{\text{view}} = \beta_{\text{view}}, \quad \pi_{\text{cond}} = \beta_{\text{cond}}.$$

Com os coeficientes estimados do verbete 003:

Característica	Preço implícito	Unidades
GLA	\$575	por sf
Terreno	\$22,70	por sf
Vista	\$86.179	por sim/não
Conservação	\$49.824	por passo

Esses preços implícitos implicam uma análise de pares de vendas. Retomando a diferença entre o elemento A e o elemento B do verbete 001:

- O elemento A tem vista ($\Delta z_{\text{view}} = +1$); contribuição +\$86.179.
- O elemento A tem um terreno 1.800 sf maior ($\Delta z_{\text{lot}} = +1,800$); contribuição $1,800 \times \$22.70 = +\40.860 .
- Todas as demais características idênticas.
- Diferença de preço prevista: $\$86.179 + \$40.860 = \$127.039$.

A diferença de preço observada foi de \$145.000. A decomposição por preços implícitos explica cerca de \$127.000 dela; os \$18.000 restantes refletem a variabilidade amostral do pequeno conjunto de dados (e, de modo mais geral, o tipo de fatores não observados de que tratará o verbete sobre variável latente).

Dois pontos operacionais merecem nota. Primeiro, a suposição de linearidade está fazendo trabalho real: sob um $p(\mathbf{z})$ não linear, a mesma diferença física de terreno poderia valer mais em um terreno pequeno do que em um grande, e a decomposição aditiva acima seria viesada. Segundo, o preço implícito de \$22,70 por pé quadrado de terreno é o valor marginal de um pé quadrado incremental, e não o valor total do terreno. A contribuição total de um terreno de 9.000 sf para o preço da cesta é $9,000 \times \$22.70 = \204.300 , mas acrescentar mais um pé quadrado eleva o preço da cesta em apenas \$22,70.

Essa distinção — marginal *versus* total — é fonte de considerável confusão na linguagem cotidiana da avaliação e é um dos ganhos conceituais mais importantes de nomear explicitamente o preço implícito.

Referências.

- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2), 132–157.
- Freeman, A. M., III (1979). Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues. *Scandinavian Journal of Economics*, 81(2), 154–173.
- Taylor, L. O. (2003). The Hedonic Method. In *A Primer on Nonmarket Valuation* (pp. 331–393). Kluwer Academic Publishers.
- Palmquist, R. B. (2005). Property Value Models. In *Handbook of Environmental Economics* (Vol. 2, pp. 763–819). North-Holland.
- Tripllett, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

Referências cruzadas: bem heterogêneo; espaço de características; função de preços hedônicos; regressão hedônica.

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (004-implicit-price.pl) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```
%% 004-implicit-price.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 004
%% Title: Implicit Price
%% Author: Bert Craytor
%% Concept: implicit_price
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(implicit_price, "Implicit Price", economic_theory,
%      "The implicit price of a characteristic is the partial
%      derivative of the hedonic price function with respect to that
%      characteristic, evaluated at a point in the characteristics
%      space; it represents the marginal market price of an
%      incremental unit of the characteristic, holding all other
%      characteristics fixed.", published, 'dictionary-2026.1').
%
% term(hedonic_price_function, ...).
% term(paired_sales_analysis, ...).
% term(adjustment, ...).
%
% formalized_by(implicit_price, rosen_1974).
% depends_on(implicit_price, hedonic_price_function).
% -----

:- use_terms([ implicit_price, hedonic_price_function, characteristics_space,
               heterogeneous_good, hedonic_regression,
               sales_comparison_approach, paired_sales_analysis, adjustment,
               comp_selection ]).
:- use_relations([related/2, depends_on/2, formalized_by/2]).

% Provenance: the term originates with Rosen (1974).
formalized_by(implicit_price, rosen_1974).

% =====
% ARTICLE SUBJECT
% =====

article_concept(implicit_price).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.004').

% =====
% LOCAL VOCABULARY
% =====
```

```

local_term(partial_derivative).
local_term(marginal).
local_term(total).
local_term(locality).
local_term(conditionality).
local_term(market_dependence).

% =====
% IMPLICIT PRICE AS PARTIAL DERIVATIVE
% =====
%
%  $\pi_i(z) = \partial p(z) / \partial z_i$ 

implicit_price_definition(
    "Implicit price of characteristic i at point z = partial derivative of p with respect to z_i,
    ↪ evaluated at z.").

unit_of_implicit_price(continuous, "dollars per unit of the characteristic").
unit_of_implicit_price(binary, "the full premium between the two states").

% =====
% THREE PROPERTIES OF IMPLICIT PRICES (per the article)
% =====

property_of_implicit_prices(local,
    "The implicit price of GLA at 1,650 sf is not necessarily the same as at 2,500 sf. p is
    ↪ generally nonlinear; partial derivatives vary with location.").
property_of_implicit_prices(conditional,
    "Implicit price depends on where the OTHER characteristics are held fixed. View premium on a
    ↪ beachfront submarket differs from inland.").
property_of_implicit_prices(market_dependent,
    "Implicit prices are equilibrium prices reflecting the current buyer pool. Market shifts change
    ↪ implicit prices even when characteristics do not.").

% =====
% IMPLICIT-PRICE / ADJUSTMENT EQUIVALENCE
% =====
%
% Every adjustment in a sales-comparison grid can be READ as asserting an
% implicit price applied to a characteristic difference. This is an
% interpretation of what the adjustment asserts, not a description of how
% traditional adjustments are derived in practice.

adjustment_as_implicit_price_application(
    "Interpretation: adjustment magnitude = (implicit price of attribute at relevant point) * (
    ↪ characteristic difference between comp and subject) --- what the adjustment asserts, not how
    ↪ it is derived.").

adjustment_derivation_in_practice(
    "Traditional adjustments are estimated directly --- from matched pairs, experience, or peer
    ↪ practice --- without computing the underlying value contributions; the traditional grid
    ↪ records only the adjustments themselves.").

% =====
% LINEAR-FORM IMPLICIT PRICES (from entry 003's adopted specification)
% =====
%
% Under the additive linear specification, implicit prices ARE the
% regression coefficients. The article reproduces them:

```

```

:- consult('../..//kb/pacifica_comps').

% Implicit price per characteristic, as inherited from the additive
% form of p adopted in entry 003.
implicit_price_of(gla_sqft, 575.00, "per square foot").
implicit_price_of(lot_sqft, 22.70, "per square foot").
implicit_price_of(view, 86179.03, "per yes/no").
implicit_price_of(cond, 49824.33, "per step on the ordinal scale").

% =====
% PAIRED-SALES DECOMPOSITION: COMP A vs COMP B
% =====
%
% Returns to the article-001 paired comparison and decomposes it using
% implicit prices.
%
% Comp A and Comp B share GLA, bedrooms, baths, condition.
% They differ in: view (+1) and lot_sqft (+1800).
% Predicted differential = 86179 + 40860 = 127039.
% Observed differential = 145000.
% Unexplained = 17961 (~$18,000).

paired_sales_decomposition(comp_a, comp_b,
  [ contribution(view, 86179),
    contribution(lot_sqft, 40860) ],
  127039,
  145000,
  17961).

% Derived from the data + coefficients (for the linter to type-check
% against the dataset rather than the hand-computed values).
contribution_for_pair(A, B, Attr, Contribution) :-
  coord(A, Attr, VA),
  coord(B, Attr, VB),
  Diff is VA - VB,
  Diff == 0,
  implicit_price_of(Attr, Pi, _),
  Contribution is Diff * Pi.

% Tiny coord helper for symbolic view/cond values.
coord(P, A, N) :-
  attribute_value(P, A, V),
  ( number(V) -> N = V
  ; A == view, V == yes -> N = 1
  ; A == view, V == no -> N = 0
  ; V = N
  ).

% =====
% MARGINAL vs TOTAL CONTRIBUTION
% =====
%
% The article emphasizes the distinction. The implicit price of lot is
% $22.70 per square foot (marginal); the TOTAL contribution of Comp A's
% 9,000-sf lot is 9,000 * 22.70 = $204,300.

distinction(marginal_vs_total,
  "Implicit price is MARGINAL ($22.70 per additional sqft of lot). Total contribution of a 9,000-
```

```

    ↪sf lot is 9,000 * $22.70 = $204,300. Confusing them is a common source of error in everyday
    ↪appraisal language.")).

%! total_contribution(?Property, ?Attribute, -Total) is det.
%
% Total dollar contribution of a characteristic to a property's bundle
% price under the linear specification.

total_contribution(P, A, Total) :-
    coord(P, A, V),
    implicit_price_of(A, Pi, _),
    Total is V * Pi.

% =====
% DEFENSIBILITY QUESTIONS RAISED
% =====

defensibility_question(location_in_space,
    "An implicit price is defensible only when estimated near the subject's location in
    ↪characteristics space; a view premium from inland comps applied to a beachfront subject
    ↪extrapolates into unknown territory.")).
defensibility_question(local_geometry,
    "If view interacts with lot size, treating them as additively separable misspecifies the
    ↪function. The implicit price of view depends on lot.")).
defensibility_question(epistemic_provenance,
    "Paired-sales analysis, regression coefficient, industry rule, and professional judgment all
    ↪point to the same quantity but have different epistemic standing. Defensibility requires
    ↪declaring the source.")).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(hedonic_regression).

```

Regressão hedônica

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/ams1t139](https://doi.org/10.67330/ams1t139)

Definição formal. A regressão hedônica é a estimação da função de preços hedônicos a partir de transações observadas em um mercado de bens heterogêneos, que produz estimativas empíricas dos preços implícitos das características que entram no modelo.

Enquadramento intuitivo. Os verbetes anteriores descreveram a função de preços hedônicos e seus preços implícitos como objetos teóricos. A função existe em princípio, mas na prática observamos apenas um conjunto finito de transações: as cestas que foram negociadas e os preços pelos quais foram negociadas. A regressão hedônica é a ponte empírica entre as duas coisas.

A montagem é direta. Cada transação fornece um par $(\mathbf{z}_i, p_i)$: o vetor de características da cesta transacionada e seu preço de venda. A coleção desses pares é uma amostra da função de preços hedônicos (não observada). A regressão ajusta uma forma funcional a essa amostra, estimando os parâmetros que indexam a função. Uma vez estimados os parâmetros, as derivadas parciais da função ajustada produzem estimativas dos preços implícitos.

Três decisões de modelagem precisam ser tomadas antes que a regressão possa ser rodada:

1. *Quais características entram no modelo?* A escolha determina as dimensões da função estimada. Omitir uma característica relevante faz com que seu efeito vaze para os coeficientes das características incluídas que sejam correlacionadas (viés de variável omitida).
2. *Qual forma funcional?* Linear, log-linear, log-log, polinomial, com expansão de bases, semi-paramétrica, de aprendizado de máquina. A escolha restringe que formato a função estimada pode assumir e quais padrões de preços implícitos ela pode recuperar.
3. *Qual estimador?* MQO, mínimos quadrados generalizados, M-estimadores robustos, regressão quantílica, estimadores de econometria espacial, regressão penalizada. A escolha determina o que o estimador está otimizando e quais suposições sustentam sua inferência.

Cada escolha é metodologicamente substantiva e cada uma acarreta consequências de defensabilidade.

Onde os avaliadores o encontram. A regressão hedônica é o fundamento da avaliação em massa e o motor da maioria dos modelos de avaliação automatizada (AVM). Quando o órgão fiscal de um condado reavalia 100.000 imóveis por ano, é uma regressão hedônica sobre vendas recentes que está fazendo o trabalho. Quando o Zillow ou o Redfin produz uma estimativa automatizada, quem a produz é um modelo de estilo hedônico (muitas vezes um *ensemble* de árvores em vez de MQO, mas conceitualmente ainda um estimador hedônico). Quando o AVM de mercado secundário de um credor aprova ou rejeita um imóvel para refinanciamento, o mesmo maquinário está envolvido.

A avaliação de imóveis individuais também recorre cada vez mais à regressão hedônica, mesmo quando a indicação final de valor vem do método comparativo direto de dados de mercado. Preços implícitos estimados a partir de uma regressão sobre uma amostra do bairro podem sustentar as magnitudes dos ajustes de forma muito mais defensável do que a intuição do avaliador ou percentuais

de regra prática. Os coeficientes da regressão tornam-se evidência, e não opinião, para os ajustes aplicados na planilha.

Por que importa para a defensabilidade. Estimativas baseadas em regressão têm um perfil de defensabilidade distinto do da análise de pares de vendas. As vantagens: uma regressão usa todas as vendas disponíveis, em vez de duas por vez; consegue isolar o efeito de uma característica controlando simultaneamente muitas outras; e produz erros padrão que quantificam a incerteza de cada preço implícito. As contrapartidas: uma regressão impõe uma forma funcional que pode não corresponder à verdadeira função de preços; exige mais dados do que uma análise de pares de vendas; e seus resultados podem ser sensíveis à escolha da amostra e da especificação.

Várias questões de defensabilidade recorrem na avaliação baseada em regressão:

A amostra é apropriada? A regressão estima a função hedônica tal como ela opera no segmento de mercado e no período da amostra. Uma regressão sobre vendas do interior aplicada a um avaliando de frente para o mar é uma falha de defensabilidade, por mais limpas que pareçam as estatísticas.

A especificação é apropriada? Uma especificação linear é uma suposição forte. Se a função verdadeira exibe retornos decrescentes da GLA acima de algum limiar, o ajuste linear estimará mal os preços implícitos nessa faixa. Diagnósticos, gráficos de resíduos e especificações alternativas fazem parte do fluxo de trabalho defensável.

O que significa o erro padrão? Um erro padrão capta a variabilidade amostral sob a especificação suposta. Ele não capta especificação incorreta do modelo, viés de variável omitida nem seleção amostral. Um erro padrão pequeno sobre uma especificação errada é mais perigoso do que um erro padrão grande sobre uma correta.

O resultado triangula com os demais métodos? Um preço implícito defensável baseado em regressão deve ser coerente com o que sugere a análise de pares de vendas, com o que implicam os custos dos componentes no método da quantificação do custo e com o que acreditam os avaliadores que conhecem o mercado. Um coeficiente de regressão em desacordo com os três é um convite a investigar, não a publicar.

Exemplo de avaliação resolvido. Podemos agora exercitar todo o maquinário da regressão hedônica sobre os oito elementos comparativos de Pacifica. O conjunto de dados está disponível em `shared/data/pacifica_comps.csv`.

Código em R:

```
comps <- read.csv("shared/data/pacifica_comps.csv")
fit <- lm(price ~ gla + lot + view + cond, data = comps)
summary(fit)
```

Saída:

```
Call:
lm(formula = price ~ gla + lot + view + cond, data = comps)

Residuals:
    Min       1Q   Median       3Q      Max
-17995   -8243   -5433   -1409    46455

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  80777.07   164509.35    0.491   0.6571
gla           574.93     109.76    5.238   0.0135 *
lot           22.70      10.38    2.187   0.1165
view        86179.03   31342.97    2.750   0.0708 .
cond        49824.33   27761.27    1.795   0.1706
```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30481 on 3 degrees of freedom
Multiple R-squared:  0.9707, Adjusted R-squared:  0.9317
F-statistic: 24.86 on 4 and 3 DF, p-value: 0.01231

```

Interpretação. O modelo ajustado estima a função de preços hedônicos como:

$$\hat{p}(z) = 80,777 + 575 z_{\text{GLA}} + 22,70 z_{\text{lot}} + 86,179 z_{\text{view}} + 49,824 z_{\text{cond}}.$$

Cada coeficiente é o preço implícito estimado da característica correspondente. A GLA contribui com \$575 por pé quadrado (com o erro padrão mais estreito e a significância mais forte); a vista contribui com cerca de \$86.000 para um imóvel de Pacifica; uma unidade adicional de estado de conservação (por exemplo, passar de “bom” para “excelente”) vale aproximadamente \$50.000.

Várias ressalvas são imediatamente visíveis:

A amostra é minúscula. Oito observações e quatro preditores deixam três graus de liberdade residuais. Os erros padrão são correspondentemente amplos, e os valores p de terreno, vista e estado de conservação não alcançam o limiar convencional de 0,05, apesar de sua plausibilidade econômica. Trata-se de um artefato de amostra pequena, e não de evidência de que essas características não importam. Com 80 elementos comparativos em vez de 8, os valores p encolheriam drasticamente.

O ajuste global é forte. Um R^2 ajustado = 0,93 e um valor p do modelo de 0,012 indicam que as características incluídas explicam conjuntamente a maior parte da variação de preços na amostra. A combinação de um ajuste conjunto forte com significância individual fraca é um sinal de manual de multicolinearidade, que aqui é estrutural: GLA, terreno e estado de conservação covariam na amostra.

Um resíduo se destaca. O resíduo máximo é de +\$46.455 e corresponde ao elemento H. Os outros sete resíduos caem todos dentro de uma faixa de $\pm \$18.000$. Algo do elemento H não é captado pela função ajustada — o modelo está subprevendo sistematicamente seu preço por um valor quase três vezes maior que o segundo maior resíduo.

Esse resíduo atípico não é ruído. É o sinal de que uma característica não observada do elemento H está contribuindo para seu preço de uma forma que as variáveis incluídas não conseguem recuperar. O verbete sobre variável latente trata exatamente dessa questão.

Validação frente ao raciocínio por pares de vendas. A análise de pares de vendas do verbete 004 previu uma diferença de \$127.039 entre os elementos A e B a partir da vista e do terreno apenas. Os preços implícitos da regressão são idênticos (uma vez que são a fonte daquele cálculo). A regressão generaliza a lógica dos pares de vendas aos oito elementos comparativos simultaneamente, mas não produz nada fundamentalmente novo com este tamanho de amostra; ela confirma a análise por pares ao mesmo tempo em que reporta erros padrão que o raciocínio por pares não é capaz de oferecer.

Referências.

- Court, A. T. (1939). Hedonic Price Indexes with Automotive Examples. In *The Dynamics of Automobile Demand*. General Motors Corporation.
- Griliches, Z. (1961). Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. In *The Price Statistics of the Federal Government*. National Bureau of Economic Research.
- Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), 34–55.
- Triplet, J. E. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*. OECD Publishing.

Referências cruzadas: bem heterogêneo; espaço de características; função de preços hedônicos; preço implícito; variável latente.

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (005-hedonic-regression.pl) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```
%% 005-hedonic-regression.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 005
%% Title: Hedonic Regression
%% Author: Bert Craytor
%% Concept: hedonic_regression
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(hedonic_regression, "Hedonic Regression", econometric_method,
% "Hedonic regression is the estimation of the hedonic price
% function from observed transactions in a heterogeneous-goods
% market, yielding empirical estimates of the implicit prices of
% the characteristics that enter the model.",
% published, 'dictionary-2026.1').
%
% term(mass_appraisal, ...).
% term(automated_valuation_model, ...). % synonym: avm
% term(omitted_variable_bias, ...).
% term(multicollinearity, ...).
% term(residual, ...).
% term(ordinary_least_squares, ...).
%
% introduced_by(hedonic_regression, court_1939).
% introduced_by(hedonic_regression, griliches_1961).
% formalized_by(hedonic_regression, rosen_1974).
% depends_on(hedonic_regression, hedonic_price_function).
% depends_on(hedonic_regression, implicit_price).
% -----

:- use_terms([ hedonic_regression, hedonic_price_function, implicit_price,
characteristics_space, heterogeneous_good, latent_variable,
mass_appraisal, automated_valuation_model,
omitted_variable_bias, multicollinearity, residual,
ordinary_least_squares,
sales_comparison_approach, paired_sales_analysis, adjustment ]).
:- use_relations([related/2, depends_on/2,
introduced_by/2, formalized_by/2]).
```

```

% =====
% ARTICLE SUBJECT
% =====

article_concept(hedonic_regression).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.005').

% =====
% LOCAL VOCABULARY
% =====

local_term(specification).
local_term(coefficient).
local_term(standard_error).
local_term(p_value).
local_term(r_squared).
local_term(adjusted_r_squared).
local_term(residual_standard_error).
local_term(degrees_of_freedom).
local_term(f_statistic).
local_term(significance_level).
local_term(diagnostic).

% =====
% THE EMPIRICAL BRIDGE
% =====
%
% Hedonic regression is the bridge from theoretical  $p(z)$  to a sample of
% observed  $(z_i, p_i)$  pairs.

setup_of_regression(
    each_transaction_supplies_pair("(z_i, p_i): characteristics vector and sale price"),
    sample_from("the unobserved hedonic price function"),
    fits("a functional form to that sample, estimating parameters")).

% =====
% THREE MODELING CHOICES (per the article)
% =====

modeling_choice(which_characteristics,
    "Determines the dimensions of the estimated function. Omission causes omitted_variable_bias.").
modeling_choice(functional_form,
    "Constrains what shape the estimated function can take. Linear, log-linear, polynomial, basis-
    ↪ expanded, semiparametric, ML.").
modeling_choice(estimator,
    "OLS, GLS, robust, quantile, spatial, penalized. Determines optimization criterion and
    ↪ inference assumptions.").

% =====
% PRACTICE CONTEXTS WHERE REGRESSION IS THE ENGINE
% =====

regression_engine_context(mass_appraisal,
    "Assessor revaluation of 100,000 parcels annually.").
regression_engine_context(automated_valuation_model,
    "Zillow/Redfin/lender AVMs; tree ensembles are conceptually hedonic estimators.").
regression_engine_context(single_property_appraisal,
    "Regression coefficients support adjustments more defensibly than appraiser intuition.").

```

```

% =====
% DEFENSIBILITY PROFILE: REGRESSION vs PAIRED-SALES
% =====

advantage_over_paired_sales(uses_all_data,
  "Regression uses all available sales rather than two at a time.").
advantage_over_paired_sales(simultaneous_control,
  "Isolates one characteristic while controlling for many others.").
advantage_over_paired_sales(quantified_uncertainty,
  "Produces standard errors quantifying uncertainty in each implicit price.").

tradeoff(functional_form_constraint,
  "Regression imposes a form that may not match the true price function.").
tradeoff(data_requirement,
  "Requires more observations than paired-sales analysis.").
tradeoff(sample_specification_sensitivity,
  "Results can be sensitive to the choice of sample and specification.").

% =====
% RECURRING DEFENSIBILITY QUESTIONS
% =====

defensibility_question(sample_appropriateness,
  "The regression estimates the function for the sample's market segment and time period;
  ↪ mismatch is a defensibility failure regardless of statistical fit.").
defensibility_question(specification_appropriateness,
  "Linear is a strong assumption; diagnostics, residual plots, and alternative specifications
  ↪ belong in the defensible workflow.").
defensibility_question(standard_error_interpretation,
  "Standard errors capture sampling variability under the assumed specification, NOT
  ↪ misspecification, omitted-variable bias, or sample selection.").
defensibility_question(triangulation,
  "Regression coefficients should be consistent with paired-sales, cost components, and local
  ↪ appraiser opinion; isolated coefficients are an invitation to investigate.").

% =====
% WORKED EXAMPLE: lm(price ~ gla + lot + view + cond)
% =====

:- consult('../..kb/pacifica_comps').

% The fitted model.
fit_call("lm(formula = price ~ gla + lot + view + cond, data = comps)").
fit_estimator(ordinary_least_squares).
fit_software(r_lm).

% Fitted coefficients, with standard errors and significance from
% the R output reproduced in the article.
fitted_coefficient(intercept, 80777.07, 164509.35, 0.491, 0.6571, ' ').
fitted_coefficient(gla, 574.93, 109.76, 5.238, 0.0135, '*').
fitted_coefficient(lot, 22.70, 10.38, 2.187, 0.1165, ' ').
fitted_coefficient(view, 86179.03, 31342.97, 2.750, 0.0708, '.').
fitted_coefficient(cond, 49824.33, 27761.27, 1.795, 0.1706, ' ').

% Diagnostics from the R output.
residual_summary(min, -17995).
residual_summary(q1, -8243).
residual_summary(median, -5433).

```

```

residual_summary(q3, -1409).
residual_summary(max, 46455).

model_summary(residual_standard_error, 30481).
model_summary(degrees_of_freedom, 3).
model_summary(r_squared, 0.9707).
model_summary(adjusted_r_squared, 0.9317).
model_summary(f_statistic, 24.86).
model_summary(model_p_value, 0.01231).

% =====
% INTERPRETIVE CAVEATS FROM THE ARTICLE
% =====

caveat(tiny_sample,
    "n=8, four predictors, three residual df. Standard errors are correspondingly wide; lot/view/
    ↪ cond fail to reach the 0.05 threshold despite economic plausibility. Small-sample artifact.").
    ↪

caveat(structural_multicollinearity,
    "Adj R^2 = 0.93 with model p = 0.012 indicates strong joint fit; weak individual significance
    ↪ is a textbook multicollinearity signature. Here GLA, lot, and condition covary in the sample
    ↪ .").

caveat(comp_h_residual,
    "Comp H's residual is +$46,455; the other seven fall within +/- $18,000. The model
    ↪ systematically under-predicts Comp H by ~3x the next-largest residual. NOT noise. Signal of
    ↪ an unobserved characteristic.").

% =====
% RELATING TO PAIRED-SALES (entry 004)
% =====
%
% The regression's implicit prices ARE the values used in entry 004's
% paired-sales decomposition of A vs B. The regression generalizes that
% logic to all 8 comps simultaneously while reporting standard errors
% that paired-sales reasoning cannot produce.

confirms_paired_sales_from(entry_004,
    "A vs B differential of $127,039 from view and lot alone | coefficients are the same;
    ↪ regression confirms the paired analysis while adding standard errors.").

% =====
% RESIDUAL OF COMP H POINTS FORWARD TO ENTRY 006
% =====

points_forward_to(latent_variable,
    "Comp H's anomalous +$46,455 residual is the signal that an unobserved characteristic is
    ↪ contributing to price. Entry 006 takes up the question.").

% =====
% PREDICTED PRICES FROM THE FITTED MODEL
% =====
%
% Computes predicted price from the fitted coefficients. The residual
% for each comp is sale_price - predicted_price.

predicted_price(P, Predicted) :-
    coord(P, gla_sqft, GLA),

```

```

    coord(P, lot_sqft, Lot),
    coord(P, view, View),
    coord(P, cond_numeric, Cond),
    fitted_coefficient(intercept, B0, _, _, _, _),
    fitted_coefficient(gla,      Bg, _, _, _, _),
    fitted_coefficient(lot,      Bl, _, _, _, _),
    fitted_coefficient(view,     Bv, _, _, _, _),
    fitted_coefficient(cond,     Bc, _, _, _, _),
    Predicted is B0 + Bg * GLA + Bl * Lot + Bv * View + Bc * Cond.

residual_of(P, R) :-
    sale_price(P, Observed),
    predicted_price(P, Predicted),
    R is Observed - Predicted.

coord(P, A, N) :-
    attribute_value(P, A, V),
    ( number(V) -> N = V
    ; A == view, V == yes -> N = 1
    ; A == view, V == no  -> N = 0
    ; V = N
    ).

% =====
% CROSS-REFERENCES
% =====

cross_reference(heterogeneous_good).
cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(latent_variable).

```

Variável latente

Tradução para o português do verbete publicado originalmente em inglês. Tradução elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Foundations • The Valuation Engineer

Bert Craytor • 3 de agosto de 2026

Versão 1.0.0

DOI: [10.67330/11n3g168](https://doi.org/10.67330/11n3g168)

Definição formal. Uma variável latente é um atributo não observado que influencia resultados observados, cujo valor precisa ser inferido de seus efeitos sobre variáveis observáveis em vez de medido diretamente.

Enquadramento intuitivo. Os cinco primeiros verbetes trataram o vetor de características $\mathbf{z}$ como se ele pudesse ser plenamente medido: GLA, terreno, vista, estado de conservação e assim por diante, todos disponíveis como colunas de dados. Na prática, nenhum vetor de características é algum dia completo. Alguns atributos não estão nos dados porque são difíceis de medir (padrão de acabamento interno, fluidez do layout, microorientação microlocacional), porque estão ausentes das bases de dados padronizadas (os registros do MLS não captam qualidade de iluminação, isolamento acústico ou benfeitorias recentes sem alvará) ou porque ainda não foram reconhecidos como relevantes (características em nível de bairro que acabam se revelando preditoras de preço apenas depois que a análise as revela).

Esses atributos não observados, porém relevantes, são *variáveis latentes*. Elas existem, influenciam o preço, mas não estão nas colunas da nossa regressão. Seus efeitos não desaparecem por não serem medidos; acumulam-se nos resíduos do modelo, onde aparecem como “ruído” sob a perspectiva das variáveis incluídas, mas são de fato sinal estruturado sob a perspectiva do processo gerador de dados subjacente.

As variáveis latentes são a ponte entre o arcabouço hedônico idealizado e a realidade desordenada dos dados de avaliação. Reconhecê-las explicitamente é pré-condição para raciocinar honestamente sobre incerteza do modelo, estrutura dos resíduos e os limites do que as características observáveis podem explicar.

Onde os avaliadores o encontram. Os avaliadores encontram variáveis latentes em três contextos recorrentes.

Os ajustes de padrão construtivo e de estado de conservação. “Padrão construtivo” e “estado de conservação”, como categorias de avaliação, são eles próprios tentativas de colapsar um construto latente multidimensional em uma única escala ordinal. Uma nota de conservação C3 no UAD 3.6 pretende resumir a manutenção adiada de um imóvel, o desgaste dos acabamentos, a idade dos sistemas prediais e a apresentação — de quatro a dez atributos subjacentes condensados em um único número. O resumo em um só número é uma *proxy* de um construto latente; o ofício do avaliador inclui o julgamento sobre quão bem a *proxy* capta a verdade subjacente em cada caso.

O resíduo inexplicado da planilha. Quando três elementos comparativos ajustados indicam valor mas discordam mais do que o esperado, a discordância é informativa. Raramente é toda ruído amostral; mais frequentemente reflete um atributo que o avaliador ainda não identificou nem ajustou. Investigar a discordância residual — “por que este elemento indica \$50.000 acima dos demais?” — é a versão informal, feita pelo avaliador, da inferência de variáveis latentes.

A benfeitoria recente que ninguém consegue ver. Um imóvel pode ter passado por uma reforma de cozinha, uma troca de telhado ou a construção licenciada de uma unidade acessória

(ADU) que ainda não se reflete nos dados do MLS ou do cadastro fiscal. Sob a perspectiva do avaliador, a benfeitoria é observável presencialmente, mas não nos dados. Sob a perspectiva da regressão, a benfeitoria é latente. O avaliador precisa conciliar a observação presencial com a indicação orientada por dados, e ambas devem convergir para a verdadeira contribuição do atributo latente.

Por que importa para a defensabilidade. As variáveis latentes introduzem várias considerações de defensabilidade que o modelo linear aditivo do verbete 005 não trata diretamente.

Viés de variável omitida. Se uma variável latente é correlacionada com uma variável observada, seu efeito vaza para o coeficiente da variável observada. Se as casas de Pacífica vendidas recentemente tendem a ter cozinhas reformadas (um atributo latente) e a data do contrato da venda está nos dados, então o coeficiente da idade da venda (isto é, `sale_age`, ou “dias desde a venda”) absorverá parte do efeito da reforma de cozinha. A contribuição estimada do preço implícito de `sale_age` torna-se viesada — ela reflete `sale_age` mais uma parcela da condição não medida de reforma da cozinha. A avaliação defensável reconhece que os coeficientes das variáveis observadas são condicionais àquilo que falta no modelo.

Interpretação dos resíduos. A teoria padrão da regressão trata os resíduos como ruído gaussiano de média zero. A visão de variáveis latentes trata os resíduos como uma mistura de ruído e sinal: parte é variação amostral, parte é o efeito cumulativo de atributos não observados. Identificar o que é o quê é metodologicamente substantivo. Um resíduo correlacionado com a localização, com a data da venda, com a identidade do avaliador ou com qualquer outro observável sugere que a variável latente correspondente pode ser parcialmente recuperada.

Seleção de modelo. Acrescentar mais variáveis observadas é o remédio óbvio para o viés de variável omitida, mas as variáveis latentes que mais importam costumam ser justamente as mais difíceis de acrescentar ao modelo (granularidade fina do estado de conservação, qualidade de benfeitorias recentes, microefeitos de bairro). A resposta defensável às vezes é reconhecer que o modelo nunca as captará e alocar a variância inexplicada ao resíduo, em vez de forçar o encaixe de variáveis adicionais.

Exemplo de avaliação resolvido. A regressão do verbete 005 produziu um resíduo incomum para o elemento H: o modelo subestimou o preço do elemento H em \$46.455, enquanto os outros sete resíduos ficaram dentro de uma faixa de $\pm \$18.000$. As características observadas do elemento H — GLA 1.600, terreno 7.300, sem vista, conservação “bom” — situam-no firmemente no meio da amostra; nada em seu perfil observável prevê um preço excepcionalmente alto. E, ainda assim, ele foi vendido por \$46.455 a mais do que o modelo diz que deveria.

A hipótese natural é que o elemento H tenha um atributo latente que contribui para seu preço mas não aparece nos dados. Na construção do conjunto de dados desta edição, é exatamente esse o caso: o elemento H passou por uma reforma de cozinha recente que acrescenta cerca de \$60.000 ao valor de sua cesta. A reforma não constava da lista de variáveis da regressão, de modo que a contribuição se acumulou no resíduo em vez de em um coeficiente. O resíduo estimado (\$46.455) é ligeiramente menor do que a verdadeira contribuição latente (\$60.000) porque a regressão espalha pequenas parcelas do efeito latente pelos demais coeficientes, viesando-os para cima nas dimensões correlacionadas.

Seguem-se três consequências operacionais:

Trate o resíduo atípico como evidência, não como erro. A primeira resposta apropriada ao resíduo do elemento H não é descartá-lo como *outlier*, mas investigar o que, no imóvel, explica a subprevisão. Neste caso, uma inspeção presencial ou uma pesquisa nos registros de alvarás revelaria a reforma. A falha do modelo tem valor diagnóstico.

Reconheça que os demais coeficientes estão ligeiramente viesados. Os coeficientes de

vista, estado de conservação e terreno absorveram pequenas frações do efeito da reforma por meio de correlações com as características observadas do elemento H. O viés é pequeno em $n = 8$ e encolheria ainda mais com mais elementos comparativos e mais covariáveis, mas não é nulo.

Os resíduos como um todo carregam estrutura recuperável. Esta é a afirmação substantiva da metodologia do Residual Constraint Approach (RCA), desenvolvida pelo autor (Craytor, 2025): em dados de avaliação residencial, os resíduos sistematicamente não são ruído gaussiano iid. Eles refletem a contribuição estruturada de atributos não observados — granularidade fina do estado de conservação, granularidade fina do padrão construtivo, utilidade funcional e outros construtos latentes — que pode ser parcialmente recuperada tratando o próprio resíduo como um sinal de variável latente a ser modelado, e não descartado.

O tratamento que o RCA dá ao estado de conservação e ao padrão construtivo torna essa estrutura de dois níveis operacional. As notas ordinais do avaliador — as *proxies* de granularidade grossa discutidas acima, que podem ser rotuladas como conservação-1 (conservação de granularidade grossa) e padrão-1 (padrão de granularidade grossa) — são rudimentares o bastante para que pudessem muito bem ser tratadas como variáveis latentes elas próprias, mas ainda assim podem entrar como regressores para se obterem contribuições aproximadas de preço implícito. O protocolo então desloca essas contribuições estimadas para fora dos atributos medidos na seção de regressão da planilha de homogeneização e de volta para o resíduo, de onde são reatribuídas às variáveis latentes de granularidade fina conservação-2 (conservação de granularidade fina) e padrão-2 (padrão de granularidade fina) e refinadas frente a evidências adicionais, como os comentários públicos do MLS. A nota de granularidade grossa fornece uma atribuição inicial de valor; a variável latente de granularidade fina carrega a estrutura recuperável. Ao final, a contribuição de valor de granularidade fina substitui a contribuição de valor de granularidade grossa.

A ilustração com um único elemento comparativo apresentada aqui é o caso mais simples da intuição do RCA. Com mais dados e modelagem mais rica, a estrutura presente no resíduo pode ser particionada em fatores latentes reconhecíveis, cada um contribuindo para o preço de mercado da cesta de maneiras que a especificação hedônica padrão não alcança. As variáveis latentes não são um incômodo a ser controlado e eliminado; são a próxima fronteira da modelagem de avaliação defensável.

Referências.

- Spearman, C. (1904). “General Intelligence,” Objectively Determined and Measured. *American Journal of Psychology*, 15(2), 201–292.
- Bollen, K. A. (2002). Latent Variables in Psychology and the Social Sciences. *Annual Review of Psychology*, 53, 605–634.
- Craytor, B. (2025). Residual Constraint Approach: Framework & Protocol. Zenodo. [doi:10.5281/zenodo.14787917](https://doi.org/10.5281/zenodo.14787917)

Referências cruzadas: espaço de características; função de preços hedônicos; preço implícito; regressão hedônica.

Adendo: formalização em Prolog

O material complementar legível por máquina apresentado a seguir (006-latent-variable.pl) formaliza este verbete na base de conhecimento Prolog da revista. Ele também é publicado como composição final navegável junto com este artigo. O código não é traduzido.

```

%% 006-latent-variable.pl
%%
%% Prolog companion to Foundations Vol. 1, Issue 1, Article 006
%% Title: Latent Variable
%% Author: Bert Craytor
%% Concept: latent_variable
%% Version: 0.1.0-draft

% =====
% DICTIONARY IMPORT
% =====

:- use_module('../tools/dictionary/dictionary').
:- dictionary_release('dictionary-2026.1').

% --- From the dictionary (loaded via :- use_module above) -----
% term(latent_variable, "Latent Variable", measurement_theory,
%      "A latent variable is an unobserved attribute that influences
%      observed outcomes, whose value must be inferred from its
%      effects on observable variables rather than measured directly.",
%      published, 'dictionary-2026.1').
%
% term(residual, ...).
% term(omitted_variable_bias, ...).
% term(proxy_variable, ...).
% term(residual_constraint_approach, ...). % synonym: rca
%
% introduced_by(latent_variable, spearman_1904).
% formalized_by(latent_variable, bollen_2002).
% introduced_by(residual_constraint_approach, craytor_2025).
% depends_on(latent_variable, characteristics_space).
% -----

:- use_terms([ latent_variable, residual, omitted_variable_bias,
               proxy_variable, residual_constraint_approach,
               characteristics_space, hedonic_price_function,
               implicit_price, hedonic_regression ]).
:- use_relations([related/2, depends_on/2,
                 introduced_by/2, formalized_by/2]).

% =====
% ARTICLE SUBJECT
% =====

article_concept(latent_variable).
article_issue(volume(1), number(1), year(2026)).
article_id('foundations.2026.006').

% =====
% LOCAL VOCABULARY
% =====

local_term(unobserved).
local_term(noise).
local_term(signal).
local_term(structured_signal).
local_term(uad_3_6).
local_term(c3_condition_rating).
local_term(mls).

```

```

local_term(permit_records).
local_term(neighborhood_micro_effect).

% =====
% THE CENTRAL CLAIM
% =====

claim(latent_variables_accumulate_in_residuals,
      "Unobserved-but-relevant attributes do not vanish when omitted from the model; their effects
      ↪ accumulate in the residual, appearing as noise from the model's perspective but as structured
      ↪ signal from the data-generating process's perspective.").

claim(residuals_are_mixture,
      "Standard regression theory treats residuals as iid Gaussian noise; the latent-variable view
      ↪ treats them as a mixture of sampling noise AND signal from unobserved attributes.").

% =====
% THREE CONTEXTS WHERE APPRAISERS ENCOUNTER LATENT VARIABLES
% =====

appraiser_encounter(quality_and_condition_proxies,
      "UAD 3.6 'C3 condition' is a single ordinal scale collapsing a multi-dimensional construct (
      ↪ deferred maintenance, finish wear, mechanical-systems age, presentation). The rating is a
      ↪ proxy_variable for a latent construct.").

appraiser_encounter(unexplained_grid_disagreement,
      "When three adjusted comps disagree by more than expected, the disagreement is rarely all
      ↪ sampling noise; usually reflects an attribute not yet identified or adjusted for.").

appraiser_encounter(unobserved_recent_improvement,
      "A recent remodel, roof replacement, or permitted ADU buildout may be observable in person but
      ↪ not in MLS/assessor data | observable to the appraiser but latent to the regression.").

% =====
% DEFENSIBILITY ISSUES
% =====

defensibility_issue(omitted_variable_bias,
      "If a latent variable is correlated with an observed variable, its effect leaks into the
      ↪ observed variable's coefficient. Observed-variable coefficients are conditional on what is
      ↪ missing from the model.").

defensibility_issue(residual_interpretation,
      "Treat residuals as a mixture, not pure noise. A residual that correlates with location, sale
      ↪ date, appraiser identity, or any other observable suggests the latent variable can be
      ↪ partially recovered.").

defensibility_issue(model_selection_limits,
      "The latent variables that matter most are often hardest to add (condition fine-grain, quality,
      ↪ micro-locational effects). Sometimes the defensible response is to allocate the unexplained
      ↪ variance to the residual rather than force-fit additional variables.").

% =====
% WORKED EXAMPLE: COMP H'S ANOMALOUS RESIDUAL
% =====

:- consult('../..//kb/pacifica_comps').

% Comp H's regression residual from entry 005.

```

```

comp_h_residual_dollars(46455).

% The latent attribute revealed by investigation.
comp_h_latent_attribute(kitchen_remodel).
comp_h_latent_attribute_value_estimate(60000).

% Why the estimated residual is smaller than the latent attribute's
% true contribution: regression spreads small portions of the latent
% effect across correlated coefficients, biasing them upward.
residual_smaller_than_truth_explanation(
    "Regression spreads small portions of the latent effect across the other coefficients via
    ↪ correlations, biasing them slightly upward in correlated dimensions.>").

% =====
% THREE OPERATIONAL CONSEQUENCES
% =====

operational_consequence(outlier_residual_is_evidence,
    "The first response to Comp H's residual is to investigate, not to drop. An in-person
    ↪ inspection or permit search reveals the remodel; the model's failure has diagnostic value.>").

operational_consequence(other_coefficients_slightly_biased,
    "View, condition, and lot coefficients absorbed small fractions of the remodel effect via
    ↪ correlations with Comp H's observed characteristics. Bias is small at n=8 but not zero.>").

operational_consequence(residuals_carry_recoverable_structure,
    "Residuals in residential appraisal data are systematically not iid Gaussian. They reflect
    ↪ structured contributions of unobserved attributes that can be partially recovered: this is
    ↪ the Residual Constraint Approach intuition.>").

% =====
% CONNECTION TO RESIDUAL CONSTRAINT APPROACH (RCA)
% =====
%
% The single-comp illustration here is the simplest case of the RCA
% intuition. With richer modeling, residual structure is partitioned
% into recognizable latent factors.

framework_reference(residual_constraint_approach, craytor_2025).

rca_intuition(
    "Treat the residual itself as a latent-variable signal to be modeled rather than discarded.
    ↪ Latent variables are not a nuisance to be controlled away; they are the next frontier of
    ↪ defensible valuation modeling.>").

rca_factor_examples(
    [condition_fine_grain, quality_fine_grain, functional_utility]).

% Two-level condition/quality structure (RCA protocol; mirrors the prose).
rating_levels(condition, condition_coarse_grain, condition_fine_grain).
rating_levels(quality, quality_coarse_grain, quality_fine_grain).

coarse_to_fine_grain_protocol(
    "Regress on the appraiser's coarse-grain C1-C6/Q1-Q6 ratings to obtain approximate value
    ↪ contributions; move those contributions out of the measured attributes back into the residual;
    ↪ reassign them to the fine-grain latent variables and refine against MLS public remarks. In
    ↪ the end, the fine-grain value contribution replaces the coarse-grain value contribution.>").

% =====

```

```
% INFERENCE RULE: anomalous-residual detection
% =====
%
% A property's residual is "anomalous" if its magnitude exceeds the
% typical residual band reported in entry 005 (+/- $18,000 for the other
% seven comps). Useful for flagging properties whose hedonic prediction
% suggests a latent attribute warrants investigation.

typical_residual_band(18000).

anomalous_residual(P, R) :-
    comp_h_residual_dollars(R0), % stand-in: this entry references the
                                % residual computed in entry 005; future
                                % integration would pull from there.

    P == comp_h,
    R = R0.

% =====
% CROSS-REFERENCES
% =====

cross_reference(characteristics_space).
cross_reference(hedonic_price_function).
cross_reference(implicit_price).
cross_reference(hedonic_regression).
```

Decisões judiciais recentes

Decisões judiciais recentes — Estados Unidos

Tradução para o português, elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Recent Court Decisions • The Valuation Engineer

William Bert Craytor • Vol. 1, Número 1 • Verão de 2026

DOI: 10.67330/fz34f711

Resumos de casos por jurisdição

Califórnia

The Retail Property Trust v. Orange County Assessment Appeals Board No. 1. California Court of Appeal (Tribunal de Apelação da Califórnia), julgado em abril de 2026.

O Court of Appeal examinou a possibilidade de concessão de desoneração do imposto predial prevista na seção 170 do California Revenue and Taxation Code (Código de Rendas e Tributos da Califórnia) em razão de restrições de acesso impostas por atos governamentais editados durante a pandemia de COVID-19. O centro comercial autor sustentou que o acesso restringido durante a pandemia configurava “dano” ao imóvel no sentido da seção 170, o que lhe daria direito à desoneração do lançamento tributário.

Implicação para os avaliadores. A interpretação que o acórdão dá ao termo “dano” na seção 170 moldará futuros pedidos de desoneração fundados em reduções de utilidade do imóvel causadas por fatores externos, sejam elas decorrentes de atos de saúde pública, de obras de infraestrutura ou de outras ações governamentais que afetem o acesso. Os avaliadores chamados a opinar sobre o impacto de restrições de acesso no valor de imóveis comerciais devem examinar a fundamentação do tribunal quanto à questão de saber se a utilidade restringida — por oposição à lesão física — constitui dano indenizável para fins de desoneração tributária.

Colorado

Cheroutes v. Rocket Mortgage & Solidifi. U.S. District Court, D. Colorado (Tribunal Distrital Federal do Distrito do Colorado); em curso, com atualizações processuais em 2026.

A petição inicial alega viés na avaliação, em violação ao Fair Housing Act (Lei da Moradia Justa), em conexão com uma avaliação de 2021 de um imóvel bifamiliar em Denver estimado em aproximadamente \$640,000 — mais de \$220,000 abaixo do valor avaliado no ano anterior. Um juiz federal já havia indeferido o pedido de extinção do processo formulado pelos réus, firmando o entendimento de que a Rocket Mortgage e a empresa de gestão de avaliações (*appraisal management company*) Solidifi devem defender-se das alegações, em vez de afastar sua responsabilidade sob a tese de que credores e AMCs não respondem pela conduta de avaliadores autônomos contratados. O caso encontra-se em fase instrutória em 2026.

Implicação para os avaliadores. O caso é o principal veículo ativo para testar a responsabilidade das AMCs e dos credores, sob o Fair Housing Act, pela conduta dos avaliadores contratados. Caso o

tribunal venha a reconhecer, no mérito, a responsabilidade da AMC e do credor, isso ampliaria substancialmente o universo de partes expostas a demandas de moradia justa decorrentes de avaliações e provavelmente se propagaria para os contratos de contratação entre AMCs e avaliadores.

Flórida

Ação coletiva na Flórida que impugna as taxas de AMC cobradas de mutuários residenciais. Ajuizada em 2 de janeiro de 2026.

A petição impugna as taxas cobradas por empresas de gestão de avaliações de mutuários de crédito imobiliário residencial, alegando que a margem entre o valor pago ao avaliador contratado e o valor cobrado do mutuário no fechamento excede os limites estabelecidos pela legislação federal e é divulgada de forma inadequada. O caso situa-se na interseção entre a exigência de honorários “customary and reasonable” (costumeiros e razoáveis) da Dodd-Frank (12 U.S.C. § 1639e) e os regimes de divulgação ao consumidor do Truth in Lending Act (Lei da Veracidade nas Operações de Crédito) e do Real Estate Settlement Procedures Act (Lei dos Procedimentos de Fechamento de Negócios Imobiliários).

Implicação para os avaliadores. O litígio testa a estrutura de remuneração das AMCs e, se houver certificação da ação como coletiva e decisão de mérito, poderá remodelar as práticas de divulgação em todo o setor. O desfecho pode afetar indiretamente o ambiente de formação de honorários dos avaliadores que atuam por meio de AMCs, ao submeter ao escrutínio judicial a diferença entre o valor pago pela AMC e o valor cobrado do mutuário.

Geórgia

Casos de servidões de conservação Green Rock. United States Tax Court (Tribunal Fiscal dos Estados Unidos), acórdãos comentados na *Tax Notes* em 2 de março de 2026.

O Tax Court percorreu cada um dos requisitos do teste do maior e melhor uso ao avaliar servidões de conservação instituídas sobre imóveis lindeiros (a Green Rock Property e a Green Rock South) na província do Piedmont, na Geórgia. Os peritos dos peticionários sustentaram que o maior e melhor uso dos imóveis antes da instituição da servidão era a exploração de uma pedreira de granito e utilizaram análise de fluxo de caixa descontado para estimar um valor justo de mercado anterior à servidão de aproximadamente \$13.75 milhões, contra um valor posterior à servidão de \$180,000.

O tribunal manifestou reservas quanto a cada requisito do teste. No que toca à admissibilidade jurídica, não se convenceu de que fosse razoavelmente provável a alteração do zoneamento para permitir a operação da pedreira, observando em particular que a mudança da classificação de zoneamento da Green Rock Property havia sido obtida pouco depois de solicitada nova avaliação da servidão de conservação — uma sequência que o tribunal considerou difícil de acolher. Quanto à possibilidade física, o tribunal entendeu que as provas geológicas dos peticionários eram insuficientes para demonstrar que uma pedreira viável pudesse ser desenvolvida no imóvel da Harman Road, não obstante a notória presença de granito lavrável na província do Piedmont em geral. O tribunal ainda firmou o entendimento de que, mesmo que se admitissem a admissibilidade jurídica e a possibilidade física, isso não seria determinante sem o atendimento dos demais requisitos da análise.

Implicação para os avaliadores. O acórdão é um exemplo trabalhado de aplicação sequencial disciplinada do teste de quatro requisitos do maior e melhor uso. Reforça a linha firmada em *Ranch Springs, LLC v. Commissioner*, 164 T.C. No. 6 (2025), e em *Beaverdam Creek Holdings, LLC v. Commissioner*, T.C. Memo. 2025-53, de que o Tax Court não acolherá alegações especulativas de

maior e melhor uso mesmo quando a viabilidade geral do uso proposto esteja documentada na região do entorno. Os avaliadores que preparam avaliações de servidões devem produzir prova específica de cada imóvel para cada requisito, em vez de se apoiar em viabilidade regional ou sustentada por analogia, e devem ser especialmente cautelosos com padrões de temporalidade — como mudanças de zoneamento obtidas simultaneamente à contratação da avaliação — que possam suscitar ceticismo judicial.

Maryland

Aquisição do Pimlico Race Course por Baltimore. Processos de desapropriação (*eminent domain*), em curso em 2025–2026.

A aquisição do Pimlico Race Course pela City of Baltimore mediante o poder de desapropriação suscita questões de justa indenização relativas à relação entre o valor do negócio em funcionamento (*going-concern value*) e o valor do bem imóvel em instalações de uso especializado. Como a operação empresarial do Pimlico é funcionalmente inseparável de sua planta física, a análise convencional de vendas comparáveis de bens imóveis tem utilidade limitada, e as avaliações pelo método da renda dependem fortemente da premissa de continuidade da operação de corridas.

Implicação para os avaliadores. Vale acompanhar o processo qualquer avaliador chamado a avaliar uma instalação de propósito único em que o negócio operacional e o imóvel sejam funcionalmente inseparáveis, como hipódromos, estúdios cinematográficos, parques temáticos, marinas e certos ativos de saúde ou hotelaria. A metodologia que emergir dos processos litigiosos provavelmente influenciará o modo como os cálculos de justa indenização tratarão os acréscimos de valor de negócio em funcionamento em futuras desapropriações.

Demandas federais contra avaliadores com base no Fair Housing Act

Braxton v. Bank of America. Ajuizada em abril de 2026.

Uma proprietária ajuizou ação contra o Bank of America e um avaliador sediado na Virgínia, alegando discriminação racial em violação ao Fair Housing Act. A autora afirma que ela e seus inquilinos estavam presentes durante a vistoria de avaliação e que a conduta do avaliador evidenciou animosidade racial. O caso dá continuidade ao padrão firmado em litígios anteriores sobre viés em avaliações, incluindo o caso *Connolly/Mott v. LoanDepot*, no Distrito de Maryland, e o caso *Austin v. Miller*, no condado de Marin, Califórnia, de demandas individuais fundadas no Fair Housing Act contra avaliadores, AMCs e credores.

Implicação para os avaliadores. O ritmo continuado de ajuizamentos de casos de viés em avaliações contra avaliadores individuais, somado ao recuo federal na aplicação da doutrina do impacto desproporcional (*disparate impact*), indica que o campo de exposição à responsabilidade se deslocou: autores privados e órgãos estaduais de fiscalização da moradia justa são hoje as principais fontes de responsabilização por viés em avaliações, e não as agências federais. Os avaliadores devem revisar suas práticas de documentação para assegurar que a seleção de comparáveis, a derivação de ajustes e a reconciliação do valor final sejam defensáveis de forma autônoma diante de escrutínio posterior, e devem estar cientes de que declarações ou condutas durante a própria vistoria são, com frequência, a base fática das alegações dos autores.

Decisões judiciais recentes — União Europeia

Tradução para o português, elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Recent Court Decisions • The Valuation Engineer

William Bert Craytor • Vol. 1, Número 1 • Verão de 2026

DOI: 10.67330/r1yn1y02

Alcance e fontes

Este levantamento é extraído de fontes primárias de acesso público quando disponíveis — notadamente a base oficial de jurisprudência do Tribunal de Justiça da União Europeia (curia.europa.eu), o HUDOC para o Tribunal Europeu dos Direitos Humanos, as decisões publicadas do Bundesfinanzhof (Tribunal Fiscal Federal alemão) (bundesfinanzhof.de), o Bundesverfassungsgericht (Tribunal Constitucional Federal alemão) (bverfg.de), o Légifrance para a jurisprudência francesa e o Upper Tribunal (Lands Chamber) (Tribunal Superior, Câmara de Terras) para o Reino Unido — e de comentários secundários em publicações profissionais e acadêmicas. As decisões em línguas nacionais foram lidas no original e são descritas em inglês; os títulos dos casos conservam a grafia e os diacríticos originais.

Bases de pesquisa jurídica exclusivas para assinantes (juris.de, beck-online, Westlaw UK, Lamy, Aranzadi) não são utilizadas. As decisões de tribunais nacionais abaixo do nível das cortes supremas são levantadas seletivamente, com prioridade para as decisões de clara relevância precedental ou metodológica. As citações de casos aqui constantes devem ser conferidas com o repositório oficial do tribunal antes de qualquer utilização na prática ou em nova publicação.

Esta edição de referência trata a Alemanha, a França e o Reino Unido como as principais jurisdições nacionais. Itália, Espanha, Países Baixos e países nórdicos não são levantados separadamente nesta edição, mas serão acrescentados nas edições trimestrais seguintes conforme o material justificar. A Suíça, embora fora da União Europeia, será objeto de levantamento em edições futuras em razão da forte influência metodológica das normas suíças de avaliação sobre a prática europeia em geral.

Contexto regulatório europeu

Dois desenvolvimentos regulatórios em nível da União Europeia afetam materialmente o ambiente de atuação dos avaliadores, embora nenhum deles seja em si uma decisão judicial:

EU AI Act e modelos automatizados de avaliação. O Regulamento (UE) 2024/1689 (o EU AI Act, Regulamento de Inteligência Artificial da União Europeia) entrou em vigor em 2024, com disposições aplicáveis aos sistemas de IA de alto risco produzindo efeitos de forma escalonada até agosto de 2026. Em 19 de maio de 2026, a Comissão Europeia publicou o projeto de diretrizes sobre a classificação dos sistemas de IA de alto risco para consulta às partes interessadas. O Anexo III do AI Act arrola a avaliação de capacidade creditícia de pessoas naturais entre as oito categorias de casos de uso presumidamente de alto risco, e o projeto de diretrizes da Comissão desenvolve essa categoria em detalhe. Como os modelos automatizados de avaliação (AVMs) usados

na originação de crédito imobiliário, no monitoramento de carteiras hipotecárias e na subscrição de seguros são funcionalmente inseparáveis da avaliação de capacidade creditícia dos empréstimos subjacentes, os AVMs operados por instituições de crédito ou seguradoras na União Europeia serão, na maior parte das configurações, classificados como sistemas de IA de alto risco. As consequências de conformidade incluem sistemas obrigatórios de gestão de riscos, controles de governança de dados, registro automático de *logs* com retenção mínima de seis meses, mecanismos de supervisão humana que permitam aos operadores compreender e sobrepor-se às decisões do sistema, e obrigações de monitoramento pós-comercialização.

Implicação para os avaliadores. Os avaliadores europeus devem esperar que os fluxos de trabalho apoiados em AVMs operados por seus clientes credores fiquem sujeitos a exigências substancialmente maiores de documentação e explicabilidade. Quando os avaliadores fornecerem insumos de modelo, dados de calibração ou revisão de supervisão humana dos resultados de AVMs, essas próprias atividades passam a integrar a postura de conformidade do credor quanto à IA de alto risco.

Diretiva relativa ao Desempenho Energético dos Edifícios (reformulação). A Diretiva (UE) 2024/1275, versão reformulada da Energy Performance of Buildings Directive (EPBD), foi publicada no Jornal Oficial em 8 de maio de 2024 e entrou em vigor em 28 de maio de 2024. O prazo de transposição pelos Estados-Membros é 29 de maio de 2026. Entre outras disposições, a EPBD reformulada exige que os Estados-Membros priorizem a renovação dos dezesseis por cento de pior desempenho do parque edificado (não residencial) e dos quarenta e três por cento de pior desempenho do parque residencial; estabelece normas mínimas de desempenho energético (MEPS); exige edifícios com emissões zero para as novas construções públicas a partir de 2028 e para todas as novas construções a partir de 2030; e torna obrigatória a avaliação de carbono ao longo de todo o ciclo de vida para novos edifícios a partir de 2030.

Implicação para os avaliadores. A EPBD reformulada afetará as avaliações de imóveis por diversos canais: os limiares mínimos de desempenho energético criam uma nova categoria de obsolescência movida por conformidade; as classificações do certificado de desempenho energético (EPC) tornam-se determinante material de preço tanto para o estoque residencial quanto para o comercial; e os custos de renovação para atingir a conformidade devem ser refletidos nos métodos do custo e na dedução da vida econômica remanescente. Os avaliadores devem também antecipar que os credores sujeitos à Mortgage Credit Directive (Diretiva do Crédito Hipotecário) exigirão, de forma crescente, tratamento explícito das classificações EPC e das necessidades de investimento de capital relacionadas às MEPS nas avaliações de garantias.

Notas para as edições seguintes

As edições trimestrais posteriores a esta referência concentrar-se-ão nas mudanças: novas decisões, desfechos das matérias pendentes aqui identificadas (notadamente o tratamento dado pelo Tribunal Constitucional Federal alemão ao *1 BvR 472/26*) e a legislação de implementação dos Estados-Membros que transponha a EPBD reformulada para o direito nacional. O arcabouço jurídico suíço, o tratamento das decisões das cortes supremas italiana e espanhola e os desenvolvimentos do *Conseil constitutionnel* francês serão acrescentados conforme o material justificar.

Resumos de casos por jurisdição

Tribunal de Justiça da União Europeia

GR Real (Processo C-351/23). Tribunal de Justiça da União Europeia, acórdão de 24 de junho de 2025.

O Tribunal de Justiça decidiu sobre reenvio prejudicial de um tribunal eslovaco originado de procedimento extrajudicial de execução hipotecária no qual consumidores perderam a residência familiar em leilão enquanto ainda pendia processo judicial que impugnava cláusulas contratuais potencialmente abusivas. O Tribunal firmou o entendimento de que as garantias de proteção do consumidor previstas na Diretiva 93/13/CEE exigem que as regras processuais nacionais permitam o controle judicial efetivo da abusividade contratual antes que uma residência hipotecada possa ser vendida por execução extrajudicial, e que os procedimentos que permitem à venda prosseguir de forma irreversível antes de concluído esse controle são incompatíveis com a Diretiva.

Implicação para os avaliadores. A decisão não é diretamente um caso de avaliação, mas remodela o ambiente processual em que as avaliações de garantias hipotecárias são contratadas e utilizadas. Em vários Estados-Membros com procedimentos extrajudiciais de execução (notadamente Eslováquia, República Tcheca, Hungria e países bálticos), o efeito prático é retardar o processo de leilão executivo e prolongar o período durante o qual a avaliação da garantia do credor é utilizada. Os avaliadores que preparam avaliações de garantias para contextos de execução nessas jurisdições devem antecipar que os laudos de avaliação poderão ficar em uso por períodos mais longos do que antes e poderão ser testados contra os movimentos de mercado ocorridos entre a data-base da avaliação original e a venda efetiva.

Tribunal Europeu dos Direitos Humanos

Demolição de edifício em Baku (Azerbaijão), acórdão de julho de 2025. Tribunal Europeu dos Direitos Humanos, artigo 1.º do Protocolo n.º 1.

O Tribunal Europeu dos Direitos Humanos constatou violação do direito ao respeito dos bens em conexão com a demolição e a desapropriação de um edifício em Baku determinadas após as autoridades invocarem risco sísmico. O tribunal de Estrasburgo entendeu que as autoridades nacionais, ao fixarem a indenização, não apresentaram razões suficientes para rejeitar o laudo de avaliação apresentado pelos requerentes e adotaram injustificadamente apenas o montante do laudo pericial determinado pelo próprio tribunal. O Tribunal arbitrou um total de €75,000 por dano material, €3,000 por dano moral e €1,055 a título de custas.

Implicação para os avaliadores. A decisão reforça uma linha de jurisprudência do artigo 1.º do Protocolo n.º 1 que exige que os tribunais nacionais deem consideração fundamentada às provas de avaliação concorrentes nos processos de desapropriação, e que a rejeição do laudo de avaliação do requerente em favor de laudo pericial determinado pelo tribunal seja ela própria fundamentada. Para os avaliadores europeus que atuam em contextos de desapropriação e de justa indenização, o caso sublinha que a defensabilidade documental de um laudo de avaliação — o raciocínio explícito que sustenta a seleção de comparáveis, a derivação de ajustes e a reconciliação do valor final — é ela própria matéria de direitos fundamentais quando o laudo é apresentado em apoio a pretensão contra o Estado.

França

Cour de cassation, Chambre commerciale, 19 de junho de 2024, nº 22-24.169. (Comentada na doutrina profissional ao longo de 2025.)

A Cour de cassation (Corte de Cassação francesa) firmou o entendimento, no contexto de controvérsias sobre a *valeur vénale* (valor de mercado) para fins de imposto sobre heranças, de que a venda do imóvel herdado ocorrida após o fato gerador (o falecimento do *de cujus*) não pode ser utilizada pelos herdeiros para revisar para baixo a *valeur vénale* originalmente declarada, mesmo quando o preço da venda posterior ao fato gerador for significativamente menor e a venda ocorrer poucos meses após o falecimento. A decisão aplica o princípio de que a *valeur vénale* para fins de imposto sobre heranças é apurada na data do fato gerador e não é corrigida por referência a transações posteriores com o mesmo imóvel.

Implicação para os avaliadores. A decisão confirma uma disciplina estrita quanto à data-base da avaliação na prática francesa do imposto sobre heranças e limita a relevância das vendas posteriores ao fato gerador como prova retrospectiva de valor. Os avaliadores que preparam avaliações francesas para fins de imposto sobre heranças devem produzir prova abrangente de vendas comparáveis com base em transações *na data-base da avaliação ou antes dela* e não devem apresentar vendas posteriores do próprio imóvel avaliando como prova confirmatória. O mesmo raciocínio provavelmente se estenderá ao imposto sobre doações (*droits de donation*) e aos contextos do imposto sobre o patrimônio (*IFI*).

Cour de cassation, 3^e Chambre civile, 8 de janeiro de 2026, nº 23-22.803. Laudos periciais amigáveis (não judiciais).

A Terceira Câmara Cível da Cour de cassation proferiu decisão que redefine o valor jurídico dos laudos periciais amigáveis (não judiciais) no processo civil francês. A decisão segue a reforma processual mais ampla operada pelo *décret* de 18 de julho de 2025, que tornou a mediação e a solução amigável de conflitos a regra padrão no processo civil. O Tribunal esclareceu que os laudos periciais amigáveis elaborados por contratação de uma única parte podem, ainda assim, ser recebidos como prova e ter peso probatório reconhecido, observados o princípio do contraditório e a avaliação, pelo juízo, da independência e da metodologia do perito.

Implicação para os avaliadores. A decisão fortalece a posição probatória das *expertises immobilières* (perícias imobiliárias) elaboradas fora do quadro dos peritos nomeados judicialmente (*experts judiciaires*), desde que o laudo observe os padrões metodológicos da profissão e a independência do perito esteja documentada. Para os avaliadores franceses que atuam em avaliações contestadas em disputas sucessórias, de divórcio ou comerciais, a decisão cria base mais sólida para que laudos contratados por clientes sejam citados e utilizados em litígios posteriores. Torna-se correspondentemente mais importante o cuidado particular em documentar a independência (ausência de subordinação, de interesse econômico ou de relação familiar com o cliente) e em aplicar os métodos prescritos pelos padrões profissionais.

Alemanha

Bundesfinanzhof, acórdãos de 12 de novembro de 2025 (II R 25/24, II R 31/24, II R 3/25). A trilogia do *Grundsteuer Bundesmodell* — constitucionalidade do sistema reformado de avaliação para o imposto predial. O Segundo Senado do Bundesfinanzhof (Tribunal Fiscal Federal alemão, BFH) decidiu, em três recursos de revisão paralelos, que as dis-

posições do método de avaliação por capitalização da renda (*Ertragswertverfahren*) aplicadas sob o chamado *Bundesmodell* para a avaliação de imóveis residenciais em onze estados federados alemães, para fins do imposto predial reformado (*Grundsteuer*) vigente a partir de 1.º de janeiro de 2025, são constitucionais. Os autores eram proprietários de apartamentos na Renânia do Norte-Vestfália, na Saxônia e em Berlim, que impugnaram seus lançamentos ao fundamento de que as regras regulamentadas de avaliação violam o princípio geral da igualdade (artigo 3.º, n.º 1, da Lei Fundamental) e produzem valores sistematicamente imprecisos.

O BFH firmou o entendimento de que o legislador criou um sistema de avaliação *conceitualmente orientado ao valor de mercado* (*konzeptionell einer Verkehrswertorientierung folgt*) e concebido para aproximar, em média sobre todos os objetos avaliados, o “valor imobiliário objetivamente realista” dentro de uma faixa do valor justo de mercado (*gemeiner Wert*). O Tribunal entendeu que o legislador não excedeu a margem que lhe é permitida ao ponderar os objetivos do esquema de avaliação em face das desigualdades necessariamente inerentes a tal esquema, e que o peso significativo atribuído a evitar novo “represamento de avaliações” mediante atualização periódica automatizada era constitucionalmente admissível. O BFH também enfatizou que os contribuintes conservam a opção, firmada em anteriores *Beschlüsse* do BFH de 27 de maio de 2024 (II B 78/23 e II B 79/23), de demonstrar valor justo de mercado inferior (*gemeiner Wert*) por meio de avaliação qualificada nos termos do § 198 BewG, com a limitação de que o valor inferior deve ser comprovado por laudo de avaliação elaborado por avaliador de imóveis publicamente nomeado que aplique métodos de avaliação reconhecidos.

Uma reclamação constitucional (*Verfassungsbeschwerde*) apoiada pelo *Bund der Steuerzahler Deutschland* e pela *Haus & Grund Deutschland* encontra-se agora pendente no Bundesverfassungsgericht (Tribunal Constitucional Federal alemão) sob o número de processo 1 BvR 472/26.

Implicação para os avaliadores. A trilogia é a mais significativa jurisprudência alemã de avaliação de imóveis do período levantado. Para os avaliadores, a principal consequência prática é a demanda elevada por avaliações de “valor inferior” do § 198 BewG (*Verkehrswertgutachten zum Nachweis des niedrigeren gemeinen Werts*) nos casos em que o *Grundsteuerwert* apurado sob o *Bundesmodell* pareça exceder o valor de mercado efetivo do imóvel avaliando. O endosso explícito do BFH a essa via, somado à sua deferência geral ao *Bundesmodell*, significa que o laudo de avaliação passa a ser o mecanismo operativo de alívio individual do contribuinte contra a supervalorização sistemática. Os avaliadores devem esperar aumento sustentado das contratações desse tipo de laudo e devem assegurar conformidade com a metodologia da ImmoWertV 2021 e com as exigências do § 198 BewG. A reclamação constitucional pendente no BVerfG (1 BvR 472/26) será material a acompanhar quanto a novas mudanças nesse arcabouço, e espera-se que seja um dos principais temas da próxima edição trimestral.

Bundesfinanzhof, acórdão de 7 de outubro de 2025, IX R 26/24. Rateio do preço de compra entre terreno e edificação (*Kaufpreisaufteilung*). Em caso de imposto de renda envolvendo a base de depreciação (*Absetzung für Abnutzung*, AfA) de um imóvel residencial para locação, o Nono Senado do Bundesfinanzhof examinou a metodologia de rateio do preço contratual de compra entre a parcela não depreciable do terreno (*Grund und Boden*) e a parcela depreciable da edificação (*Gebäude*). Os autores haviam impugnado a utilização, pelo juízo de origem, dos *Bodenrichtwerte* (valores de referência do terreno) ao fundamento de que a base de dados subjacente não estava confiavelmente documentada para a localização específica do imóvel avaliando.

Implicação para os avaliadores. A decisão dá continuidade a uma linha de jurisprudência do BFH

(notadamente BFH II R 25/24, de 12 de novembro de 2025, citado acima, e o anterior BFH IX R 12/21, de 20 de setembro de 2022) que confirma ser a ImmoWertV a fonte dos princípios de avaliação reconhecidos para o rateio do preço de compra. O caso também ilustra a importância da base de dados subjacente aos *Bodenrichtwerte*: quando os *Gutachterausschüsse* (comitês periciais de avaliação) não dispõem de dados suficientes de vendas comparáveis para a localização específica, o *Bodenrichtwert* pode ser impugnado por deficiência de qualidade dos dados. Para os avaliadores que preparam rateios de preço de compra ou pareceres sobre base de depreciação, isso significa que a documentação da base de dados do *Bodenrichtwert* — inclusive a *Kaufpreissammlung* subjacente — passa a integrar a análise defensável.

Reino Unido

Mehrunnisa Alikhan Ahmed Khan v Wolverhampton City Council. Upper Tribunal (Lands Chamber), 2025.

O Upper Tribunal (Lands Chamber) confirmou que a data-base correta de avaliação para apurar a indenização na aquisição compulsória (*compulsory purchase*) é a data em que o imóvel se transfere para a autoridade adquirente por meio de uma *general vesting declaration* (GVD, declaração geral de transferência de titularidade). O Tribunal também examinou a relação entre avaliações contemporâneas e retrospectivas: no caso concreto, a avaliação inicial da autoridade local, feita imediatamente após a GVD, foi de £150,000, enquanto uma segunda avaliação, elaborada em 2025 no curso do processo perante o tribunal e referida à data presumida de transferência três meses após a notificação do proprietário, foi de £120,000. Não obstante a confirmação da data-base legal de avaliação, o Tribunal preferiu o valor calculado em momento mais próximo da data em que a GVD foi feita, o que resultou em pagamento maior ao requerente.

Implicação para os avaliadores. A decisão é metodologicamente instrutiva para os avaliadores britânicos que atuam em contextos de aquisição compulsória. Embora a data-base legal de avaliação seja a data da transferência, a preferência do Tribunal pela avaliação contemporânea em detrimento de avaliação retrospectiva elaborada anos depois, sobre a mesma data-base legal, ilustra que a prova de mercado contemporânea tem peso prático mesmo quando não fosse formalmente admissível na própria data de transferência. Os avaliadores devem assegurar que os laudos elaborados em momento próximo a uma GVD estejam documentados com prova de apoio suficiente para servir de registro principal, antecipando que processos posteriores perante o tribunal poderão apoiar-se no laudo contemporâneo em preferência à reconstrução retrospectiva.

Decisões judiciais recentes — América Latina

Tradução para o português, elaborada com assistência de IA e revisada pela redação; a versão em inglês é a autorizada.

Recent Court Decisions • The Valuation Engineer

William Bert Craytor • Vol. 1, Número 1 • Verão de 2026

DOI: 10.67330/ywghzg47

Alcance e fontes

Este levantamento é extraído de fontes primárias de acesso público quando disponíveis — notadamente o repositório de casos da Corte Interamericana de Direitos Humanos (corteidh.or.cr), os portais do Supremo Tribunal Federal (STF) e do Superior Tribunal de Justiça (STJ) do Brasil (portal.stf.jus.br, stj.jus.br), o *Semanario Judicial de la Federación* mexicano (sjfsemanal.scjn.gob.mx), a *Corte Suprema de Justicia de la Nación* argentina (Corte Suprema de Justiça da Nação) (csjn.gov.ar), o portal do *Poder Judicial* chileno (poderjudicial.cl) e o *Diario Constitucional*, e a *Corte Constitucional* colombiana (Corte Constitucional) (corteconstitucional.gov.co) — e de comentários secundários em publicações profissionais e acadêmicas. As decisões em espanhol e em português foram lidas no original e são descritas em inglês; os títulos dos casos conservam a grafia e os diacríticos originais.

Bases de pesquisa jurídica exclusivas para assinantes (vLex, La Ley, Thomson Reuters Proview Latin America) não são utilizadas. As decisões de tribunais nacionais abaixo do nível das cortes supremas são levantadas seletivamente, com prioridade para as decisões de clara relevância precedental ou metodológica. As citações de casos aqui constantes devem ser conferidas com o repositório oficial do tribunal antes de qualquer utilização na prática ou em nova publicação.

Esta edição de referência trata Argentina, Brasil, Chile, Colômbia e México como as principais jurisdições nacionais. Peru, Venezuela, Equador, Bolívia, Paraguai, Uruguai e as jurisdições da América Central e do Caribe não são levantados separadamente nesta edição, mas serão acrescentados nas edições trimestrais seguintes conforme o material justificar. O trabalho de avaliação ligado à restituição de propriedade na Venezuela e às sanções relativas a Cuba será objeto de levantamento em edição futura em razão de sua distinção metodológica e de sua relevância para a prática transfronteiriça.

Contexto regional

Ao contrário do levantamento da União Europeia, este levantamento latino-americano não tem análogo direto à camada regulatória supranacional representada pelo EU AI Act e pela Diretiva relativa ao Desempenho Energético dos Edifícios, na versão reformulada. A regulação da avaliação de imóveis na América Latina é predominantemente nacional; as principais influências supranacionais são a jurisprudência de direitos fundamentais da Corte Interamericana de Direitos Humanos sob o artigo 21 da Convenção Americana (Direito à Propriedade) e a adoção das Normas Internacionais de Relatório Financeiro (em particular a IFRS 13, Mensuração do Valor Justo) pela maioria das jurisdições da região para fins de relatório financeiro. Dois temas metodológicos, ainda assim, percorrem a jurisprudência regional e merecem menção aqui:

Avaliação de propriedade comunal indígena. O desafio metodológico mais distintivo da prática avaliatória latino-americana é a avaliação da propriedade comunal indígena e afrodescendente em contextos de restituição, desapropriação e indenização. A jurisprudência da Corte Interamericana tem firmado de forma consistente o entendimento de que a propriedade comunal indígena é constitucionalmente protegida em pé de igualdade com a propriedade privada individual, mas os métodos pelos quais tal propriedade é avaliada — em que tipicamente não há transações comparáveis em condições de mercado, em que a propriedade é mantida de forma comunal e inalienável e em que o significado cultural e espiritual do território é parte integrante de seu valor — permanecem pouco desenvolvidos. Ao final de 2025, vinte e quatro decisões da Corte Interamericana favoráveis a povos indígenas permaneciam pendentes ou apenas parcialmente cumpridas, várias delas envolvendo cálculos de indenização em aberto que dependem de metodologia de avaliação ainda não resolvida.

Inflação, moeda e indexação na indenização por desapropriação. Várias jurisdições da região, notadamente a Argentina, exigem que as metodologias de avaliação e os cálculos de indenização lidem com inflação elevada, *tasaciones* (avaliações) em moeda estrangeira e regimes complexos de indexação. A Corte Suprema argentina firmou o entendimento de que a indenização por desapropriação originalmente expressa em moeda estrangeira não pode ser convertida em moeda local a uma paridade artificial, e de que a avaliação deve refletir o valor econômico real do imóvel no momento da indenização. Preocupações semelhantes aplicam-se na Venezuela (em conexão com avaliações de restituição e relativas a sanções) e periodicamente no Brasil em períodos de inflação elevada. Para os avaliadores que atuam em várias jurisdições da região, essas considerações de inflação e moeda integram a metodologia de avaliação de um modo que não tem paralelo europeu ou estadunidense.

Notas para as edições seguintes

As edições trimestrais posteriores a esta referência concentrar-se-ão nas mudanças: novas decisões da Corte Interamericana, desfechos das matérias pendentes aqui identificadas, avanços na implementação da Jurisdicción Agraria y Rural da Colômbia rumo ao seu início operacional em 2027, e desenvolvimentos nos Estados quanto à prática de mensuração do valor justo sob a IFRS 13. Peru, Venezuela, Equador e as jurisdições da América Central serão acrescentados conforme o material justificar. Atenção particular será dada ao desenvolvimento metodológico da avaliação de propriedade comunal indígena, campo em que os entendimentos substantivos da Corte Interamericana superaram a metodologia prática de avaliação disponível aos avaliadores e às autoridades de indenização.

Resumos de casos por jurisdição

Corte Interamericana de Derechos Humanos

Maya Q'eqchi' Agua Caliente vs. Guatemala e Comunidad Garífuna de San Juan y sus Miembros vs. Honduras. Corte Interamericana de Derechos Humanos, sentenças de 2024, comentadas na jurisprudência regional ao longo de 2025.

Em duas sentenças contemporâneas, a Corte Interamericana decidiu sobre as obrigações da Guatemala e de Honduras de demarcar, titular e impedir o esbulho de territórios ancestrais pertencentes a comunidades indígenas e afrodescendentes. A comunidade Maya Q'eqchi' Agua Caliente obteve título definitivo de propriedade sobre seu território somente após quarenta e cinco anos de atuação administrativa e judicial; a comunidade Garífuna de San Juan, em Honduras, enfrentava o esbulho

contínuo de terras costeiras ancestrais. Ambas as sentenças trataram das obrigações do Estado sob o artigo 21 (direito à propriedade) da Convenção Americana de reconhecer, demarcar e proteger a propriedade comunal indígena. Comentaristas regionais observaram que o arcabouço de restituição da Corte Interamericana nessas decisões é algo mais restritivo do que em sentenças anteriores, particularmente no tratamento dos “terceiros inocentes” (*terceros inocentes*) que adquiriram direitos sobrepostos sem ciência da reivindicação indígena.

Implicação para os avaliadores. As decisões reforçam que a propriedade comunal indígena e afrodescendente é constitucionalmente protegida em todo o sistema interamericano e que a avaliação de tal propriedade em contextos de restituição ou indenização não pode apoiar-se em vendas comparáveis convencionais em condições de mercado (que tipicamente não existem) sem complementação metodológica. Os avaliadores contratados em contextos de restituição indígena devem antecipar que a avaliação deve integrar três componentes distintos: o valor de mercado subjacente de terras de características físicas comparáveis na região; os custos de oportunidade e as benfeitorias atribuíveis ao uso indígena; e o significado cultural e espiritual não econômico do território, que a jurisprudência da Corte Interamericana trata como dimensão indenizável do direito de propriedade. A metodologia para o terceiro componente permanece pouco desenvolvida e é objeto contínuo do raciocínio da Corte Interamericana.

Mapuche / Consejo de Todas las Tierras vs. Chile. Corte Interamericana de Derechos Humanos, sentença de 2024 (comentada na jurisprudência de 2025).

A Corte Interamericana condenou o Estado chileno por violações de direitos humanos contra 135 membros do povo Mapuche em conexão com as ações de reivindicação territorial do Consejo de Todas las Tierras (1989–1992) e com a aplicação da legislação antiterrorista a reivindicações territoriais indígenas. A sentença estende a doutrina de *Norín Catrimán e outros vs. Chile* (2014) e reconhece explicitamente o direito dos povos indígenas à autodeterminação em conexão com reivindicações territoriais.

Implicação para os avaliadores. Embora os entendimentos principais do caso digam respeito a proteções processuais penais e não diretamente à metodologia de avaliação, o reconhecimento da autodeterminação indígena em conexão com o território cria uma camada processual que afeta o trabalho de avaliação em contextos chilenos de restituição territorial indígena e correlatos. Os avaliadores contratados nesses contextos devem assegurar que a metodologia incorpore processos adequados de consulta e reconheça a autodeterminação da comunidade quanto ao modo como seus valores territoriais são caracterizados.

Argentina

Doutrina argentina sobre o Tribunal de Tasaciones de la Nación. Consolidada na jurisprudência até 2025.

A Corte Suprema de Justicia de la Nación (CSJN) e os tribunais federais inferiores continuaram a desenvolver a doutrina relativa às avaliações realizadas pelo Tribunal de Tasaciones de la Nación (Tribunal de Avaliações da Nação, TTN) em matérias de desapropriação e de *expropiación inversa* (desapropriação indireta). A doutrina firmada da CSJN (Fallos 313:82, aplicada reiteradamente ao longo de 2024–2025) sustenta que as avaliações do TTN têm peso presuntivo em razão da capacidade técnica dos membros do tribunal, da base probatória em que se fundam e da uniformidade da prática; o afastamento de uma avaliação do TTN exige a demonstração de erro ou omissão manifestos. Os tribunais federais que aplicaram essa doutrina em 2024–2025 — inclusive o Juzgado Federal de

Oberá em *Sosa v. Entidad Binacional Yacyretá* (junho de 2024, *expropiación inversa* decorrente do reservatório hidrelétrico de Yacyretá) — continuam a apoiar-se nas avaliações do TTN como âncora operativa da indenização.

Implicação para os avaliadores. Para os avaliadores argentinos, a consequência prática é que os laudos privados apresentados em oposição a avaliações do TTN enfrentam ônus probatório substancial. Para afastar ou modificar uma avaliação do TTN, o laudo privado deve demonstrar erro ou omissão metodológica no trabalho do TTN, e não apenas uma conclusão de valor alternativa. Os avaliadores que preparam tais laudos devem concentrar a crítica metodológica na seleção de vendas comparáveis do TTN, na derivação de ajustes e no tratamento das condições de mercado locais, e não apenas na magnitude da diferença de valor.

Tasaciones em moeda estrangeira e paridade da indenização. A doutrina da CSJN segundo a qual a indenização por desapropriação originalmente expressa em moeda estrangeira não pode ser convertida em pesos argentinos a uma paridade artificial de um para um permanece operativa ao longo de 2025. A doutrina protege os desapropriados contra indenizações que, em razão da inflação ou da política cambial, deixem de refletir o valor econômico real do imóvel no momento da indenização. O princípio foi periodicamente reafirmado em 2024–2025 em conexão com processos de desapropriação de longa duração.

Implicação para os avaliadores. Os avaliadores argentinos que preparam avaliações para matérias de desapropriação que atravessem períodos de inflação significativa ou de mudança de política cambial devem documentar explicitamente a base monetária da avaliação e devem considerar a metodologia de indexação adequada entre a data-base da avaliação e a data prospectiva de pagamento. Laudos que não abordem os riscos de inflação e de moeda dificilmente serão persuasivos perante tribunais que apliquem a doutrina indenizatória da CSJN.

Brasil

Supremo Tribunal Federal, RE 922.144 (Tema 865 da repercussão geral). Supremo Tribunal Federal (STF), julgamento de 2024 com aplicação continuada em 2025.

O STF firmou o entendimento de que a indenização em processos de desapropriação é constitucionalmente compatível com o regime federal de precatório estabelecido pelo artigo 100 da Constituição Federal. Especificamente, a diferença entre o valor inicialmente depositado pelo poder público (para fins de *imissão provisória na posse*) e o valor ao final apurado em perícia judicial deve ser paga pelo sistema de precatório, e não como depósito complementar imediato. A decisão pondera o comando constitucional da indenização *justa, prévia e em dinheiro* (artigo 5.^o, XXIV) em face das exigências de pagamento ordenado do sistema de precatório.

Implicação para os avaliadores. A decisão tem consequências práticas diretas para a avaliação do depósito inicial preparada por avaliadores que atuam para as entidades expropriantes. Como qualquer subestimação no depósito inicial será paga posteriormente pelo sistema de precatório (que pode envolver atraso substancial), há um incentivo estrutural para que os autores impugnem as avaliações iniciais como subestimadas. Os avaliadores que preparam avaliações de depósito inicial devem documentar a metodologia com cuidado particular para defender a avaliação contra a perícia judicial posterior, antecipando que a diferença entre o depósito inicial e a determinação judicial final ficará sujeita a juros compensatórios nos termos da jurisprudência do STF e do arcabouço do Decreto-Lei 3.365/1941.

Superior Tribunal de Justiça, REsp 2.129.162 (Tema repetitivo). Superior Tribunal de Justiça (STJ), julgamento de maio de 2025.

O STJ resolveu controvérsia repetitiva a respeito dos honorários advocatícios nos casos em que a entidade expropriante desiste da ação de desapropriação. A Corte estabeleceu que os honorários nos casos de desistência devem seguir o regime do artigo 27, § 1.º, do Decreto-Lei 3.365/1941, e não as regras processuais gerais do *Código de Processo Civil*. A decisão originou-se de ação da Cemig (Companhia Energética de Minas Gerais) que buscava servidão administrativa para linha de distribuição de energia elétrica.

Implicação para os avaliadores. A decisão é processual e não metodológica, mas tem implicações para a estrutura da contratação de avaliadores em matérias de servidão administrativa e de desapropriação. Os avaliadores contratados como peritos judiciais em tais processos devem assegurar que os ajustes de honorários estejam documentados por referência ao arcabouço do decreto-lei aplicável, e não às regras processuais gerais.

Superior Tribunal de Justiça, doutrina sobre *juros compensatórios*. Aplicação continuada em 2025.

O STJ manteve sua doutrina de que os juros compensatórios em matérias de desapropriação somente *incidem* após a decisão judicial sobre o domínio do imóvel expropriado. Os juros aplicam-se à diferença entre o depósito inicial e o valor final da indenização, à taxa historicamente estabelecida de doze por cento ao ano (*Súmula* 618 do STF), e destinam-se a compensar o desapropriado pela perda antecipada da posse antes do pagamento da indenização final.

Implicação para os avaliadores. Os avaliadores brasileiros devem antecipar que a diferença de valor por eles identificada na perícia judicial — entre o depósito inicial do poder público e o valor de mercado apurado — torna-se a base do cálculo dos juros compensatórios. A significância econômica dessa diferença é, portanto, maior do que a diferença nominal sugere, particularmente em processos longos.

Chile

Corte Suprema, Rol Nº 31.423-2025. Corte Suprema do Chile, acórdão de setembro de 2025.

A Corte Suprema chilena rejeitou recurso de cassação interposto por uma incorporadora imobiliária (*inmobiliaria*) contra o Tesouro do Estado (*Fisco de Chile*) em processo especial de desapropriação fundado no artigo 12 do Decreto-Lei Nº 2.186. A autora havia sustentado que a avaliação fixada pela Comissão de Peritos (*Comisión de Peritos*) não refletia o maior e melhor uso (*mayor y mejor uso*) do imóvel nem o valor das construções existentes e das obras complementares, e pleiteara adicionalmente indenização por lucros cessantes (*lucro cesante*) sob o fundamento de contrato de locação existente que geraria receitas futuras de aluguel. A Corte firmou o entendimento de que os argumentos da autora equivaliam a discordância quanto à valoração da prova feita pelo juízo de origem, e não a verdadeira violação das normas legais que regem a avaliação, e que a fundamentação do juízo de origem — que se apoiou na prova da Comissão de Peritos em conjunto com prova documental e testemunhal — era adequadamente motivada.

Implicação para os avaliadores. A decisão ilustra a deferência substancial que a doutrina da Corte Suprema chilena confere à avaliação da Comissão de Peritos nos processos de desapropriação, bem como o alcance limitado da revisão em cassação das conclusões de avaliação com fundamento em mera discordância metodológica. Os avaliadores que preparam avaliações contrapostas em matérias

chilenas de desapropriação devem concentrar-se em erro metodológico demonstrável no trabalho da Comissão, e não em raciocínio de avaliação alternativa, e devem estar cientes de que as pretensões de *lucro cesante* baseadas em contratos de locação existentes enfrentam exigências probatórias substanciais.

Corte Suprema (desapropriação em Coquimbo). Corte Suprema do Chile, acórdão de 17 de julho de 2025.

A Corte Suprema chilena ratificou decisão de segundo grau que elevou a indenização em matéria de desapropriação em Coquimbo para UF 2.969,75 (com base em laudo de avaliador privado), contra o montante sustentado pelo Estado de UF 2.349,07 (baseado em vendas registradas junto ao *Conservador de Bienes Raíces*, registro de imóveis). A Corte enfatizou que a utilização, pelo tribunal de segundo grau, do laudo do avaliador privado, considerado em conjunto com prova testemunhal e com laudo adicional de avaliação de imóvel vizinho, foi exercício admissível da *sana crítica* (valoração racional da prova) e que a revisão em cassação não alcança o reexame de tal valoração probatória.

Implicação para os avaliadores. A decisão de Coquimbo é a contrapartida do Rol N° 31.423-2025: assim como a avaliação da Comissão prevalece quando impugnada por uma incorporadora, a avaliação mais alta de um avaliador privado também pode prevalecer quando sustentada por fundamentação metodológica adequada e prova corroborante. A lição prática para os avaliadores chilenos é que a metodologia e o suporte documental de uma avaliação privada são decisivos, e que a prova de vendas registradas, isoladamente, é insuficiente para afastar uma avaliação privada bem fundamentada que se apoie em comparáveis adicionais e em prova contextual.

Tribunal Constitucional, outubro de 2025. *Tasación de bienes raíces* (imóveis) nos termos do Código Tributário.

O Tribunal Constitucional chileno rejeitou impugnação constitucional às disposições do *Código Tributario* (Código Tributário) que regem a *tasación* (avaliação) de imóveis (*bienes raíces*) para fins do *impuesto territorial* (imposto territorial). Ministros vencidos sustentaram que as normas impugnadas ofereciam mecanismos inadequados para que a parte executada se opusesse à *tasación*, comprometendo a garantia constitucional de procedimento racional e justo.

Implicação para os avaliadores. A decisão preserva o arcabouço administrativo de *tasación* vigente para o imposto imobiliário chileno, mas a fundamentação vencida identifica uma vulnerabilidade do arcabouço atual que poderá ser retomada em futuras impugnações constitucionais ou administrativas. Os avaliadores chilenos que preparam avaliações relacionadas a disputas de imposto imobiliário devem acompanhar o ambiente normativo posterior à decisão, pois reformas administrativas para aprimorar os mecanismos de objeção do contribuinte parecem plausíveis.

Colômbia

Corte Constitucional, aprovação da Jurisdicción Agraria y Rural. Corte Constitucional colombiana, decisão de 2025.

A Corte Constitucional colombiana aprovou, em controle de constitucionalidade, a legislação que cria a nova Jurisdicción Agraria y Rural (Jurisdição Agrária e Rural), uma jurisdição especializada em terras rurais dentro do Poder Judiciário colombiano. A Corte aprovou artigos centrais que definem o objeto da reforma, sua integração ao Poder Judiciário, a criação de tribunais e juízos especializados e os aspectos orçamentários. A nova jurisdição está prevista para iniciar suas operações em 2027 e é

concebida para resolver disputas de restituição de terras, disputas de propriedade agrária e matérias decorrentes da implementação do acordo de paz de 2016. Na data da aprovação, disputas agrárias envolvendo mais de 61.000 hectares encontravam-se pendentes no sistema administrativo e judicial existente.

Implicação para os avaliadores. A nova jurisdição tornar-se-á o foro principal das disputas de avaliação decorrentes de processos de restituição de terras e de matérias de propriedade rural na Colômbia. Os avaliadores colombianos que atuam em contextos de restituição (inclusive os que trabalham para a Unidad de Restitución de Tierras e para o Instituto Geográfico Agustín Codazzi) devem antecipar que os novos tribunais desenvolverão sua própria jurisprudência de avaliação e que os padrões metodológicos poderão evoluir a partir do atual arcabouço conduzido administrativamente. Os profissionais devem acompanhar o início operacional em 2027 e a formação do acervo inicial de casos dos tribunais.

Arcabouço de avaliação baseado no IGAC para a restituição de terras. Aplicação continuada em 2025.

A prática colombiana de restituição de terras continua a apoiar-se no *avalúo comercial* (avaliação de mercado) elaborado pelo Instituto Geográfico Agustín Codazzi (IGAC) como base principal da indenização em contextos de restituição e de substituição, complementado por *avalúos* elaborados pelas Lonjas de Propiedad Raíz (entidades de profissionais de imóveis) quando exigido. O arcabouço está articulado na decisão SU-034 de 2018 da Corte Constitucional (sobre os requisitos processuais de cumprimento) e em decisões posteriores, inclusive a T-120 de 2024 (sobre a relação entre a avaliação do IGAC e as preferências individuais dos beneficiários em situações de indenização em espécie).

Implicação para os avaliadores. Para os avaliadores que atuam na restituição de terras na Colômbia, o arcabouço institucional coloca as avaliações do IGAC no centro operativo do cálculo da indenização, com os *avalúos* das Lonjas de Propiedad Raíz servindo como mecanismo suplementar ou de resolução de controvérsias. A futura Jurisdicción Agraria y Rural pode alterar essa arquitetura institucional, e o peso relativo das avaliações do IGAC frente às das Lonjas é uma das questões que provavelmente serão desenvolvidas pelos novos tribunais.

México

Suprema Corte de Justicia de la Nación, incidência de ISR sobre a diferença entre *avalúo* e preço. Suprema Corte do México, decisão anunciada no início de 2026.

O Pleno da Suprema Corte de Justicia de la Nación (Suprema Corte de Justiça da Nação do México, SCJN) declarou constitucionais os artigos 125, 130(IV) e 132 da Ley del Impuesto sobre la Renta (Lei do Imposto sobre a Renda, ISR), que preveem que, quando o *avalúo* de uma aquisição imobiliária excede o preço contratual de compra em mais de dez por cento, a diferença é tratada como receita acumulável do adquirente e fica, portanto, sujeita ao ISR. A Corte firmou o entendimento de que a regra não viola os princípios constitucionais da proporcionalidade e da equidade tributárias, ao raciocínio de que a aquisição de imóvel por preço materialmente inferior ao valor de mercado representa aumento objetivo do patrimônio do adquirente e, por isso, constitui adequadamente benefício econômico tributável.

Implicação para os avaliadores. A decisão tem consequências práticas substanciais para as transações imobiliárias mexicanas e para a prática avaliatória que as apoia. Os *avalúos* realizados por

corredores públicos (corretores públicos habilitados) ou por outros profissionais autorizados passam a ter consequências fiscais diretas para os adquirentes quando o *avalúo* exceder o preço contratual em mais de dez por cento. Os avaliadores devem antecipar maior escrutínio da metodologia em *avalúos* transacionais e demanda crescente por *avalúos* que documentem com rigor particular a base de vendas comparáveis. A decisão também eleva implicitamente a consequência da escolha metodológica na seleção entre métodos alternativos de avaliação, uma vez que a escolha pode agora afetar a carga tributária do adquirente.

Suprema Corte de Justicia de la Nación, Amparo en revisión 391/2024. Segunda Sala da Suprema Corte do México, acórdão de 25 de junho de 2025.

A Segunda Sala da Suprema Corte mexicana julgou impugnação constitucional apresentada por uma mineradora contra Decreto que afeta o regime jurídico da desapropriação de lotes de mineração e dos direitos acessórios correlatos (ocupação temporária, servidões, preferências em concessões de água e disposições de renovação de arrendamento). A Corte firmou o entendimento de que o Decreto não violou o princípio da irretroatividade, porque as disposições afetadas diziam respeito a meras expectativas de direitos futuros, e não a direitos adquiridos.

Implicação para os avaliadores. A decisão é significativa para a avaliação de direitos minerários e de propriedades de mineração no México. A caracterização, pela Corte, dos direitos de prorrogação de lotes de mineração e das disposições correlatas como “meras expectativas”, e não como direitos adquiridos, afeta a avaliação de concessões de mineração e a aferição do risco regulatório em sua avaliação. Os avaliadores que preparam avaliações de propriedades de mineração ou de direitos correlatos no México devem abordar explicitamente a dimensão de risco regulatório da avaliação e devem evitar tratar valores contingentes ou dependentes de prorrogação como se estivessem plenamente consolidados.