

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 apóstrofes (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 lambda selecionado, lambda.min/lambda.1se, gamma (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 λ_{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).