

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).

```