



Promoção:



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia



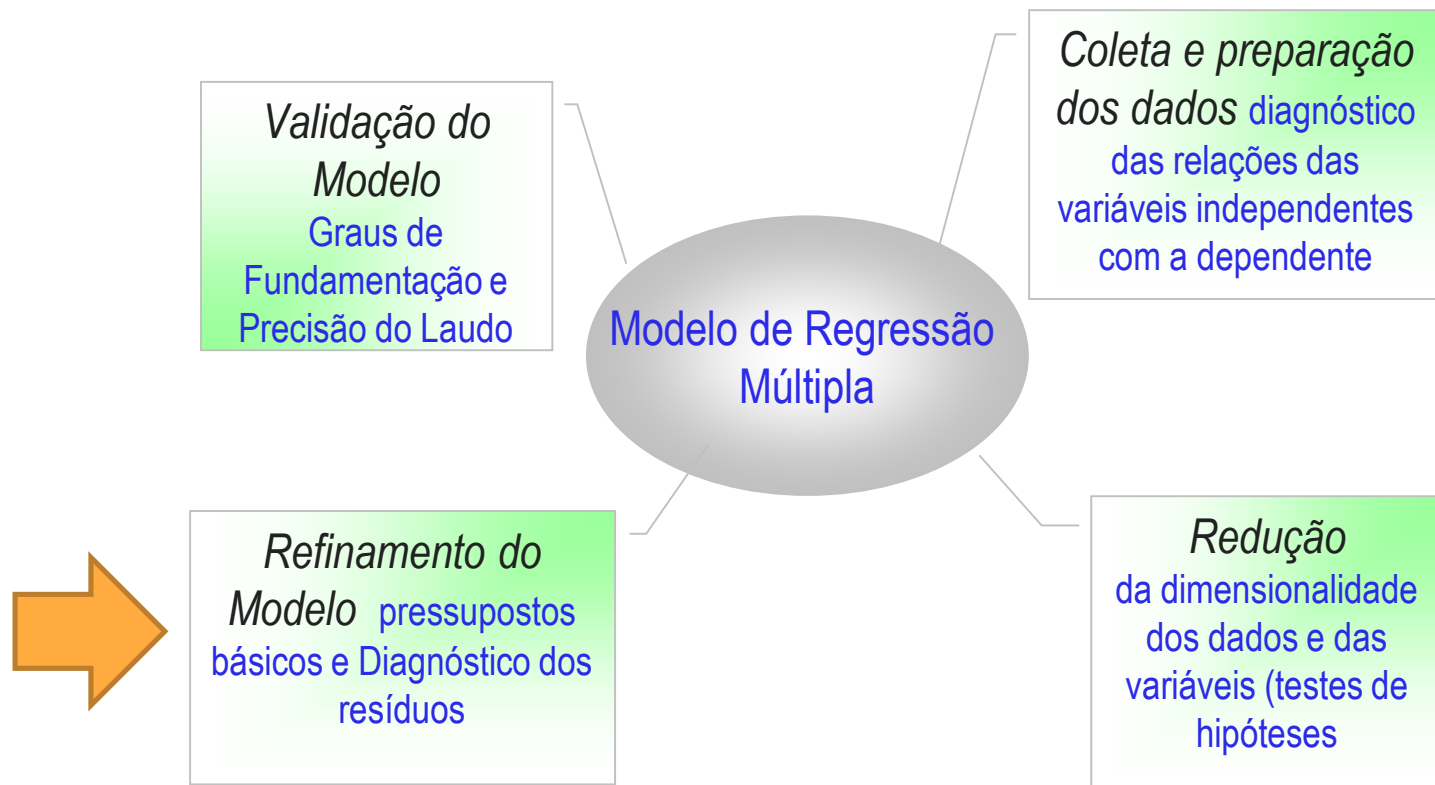
IBAPE BAHIA
Instituto Brasileiro de
Avaliações e Perícias de
Engenharia da Bahia

Painel: “**PARTICULARIDADES DA REGRESSÃO LINEAR**”

Arq. Urb. Ana Maria de Biazzi Dias de Oliveira
anabiazzi@uol.com.br

Arquiteta e Urbanista, Mestre em Engenharia Civil e Urbana pela Poli-USP. Certificada em Engenharia de Avaliações pelo IBAPE. Ex-Presidente do IBAPE/SP (biênio 2009-2010). Professora da disciplina “Inferência Estatística” de cursos de pós-graduação e especialização em Engenharia de Avaliações e Perícias. Coautora do capítulo “Métodos Científicos e a Engenharia de Avaliações” do Livro - Engenharia de Avaliações do IBAPE/SP.

Principais Etapas





A questão do respeito aos pressupostos da Regressão Linear é vastamente discutida na literatura.

Fundamentais nas derivações das fórmulas e procedimentos de Regressão Linear Múltipla, dependendo da situação, pequenas violações **não causam distorções relevantes nos valores das estimativas pontuais** e na significância associada a cada variável independente.

O que são pequenas ou grandes violações ?



Motivação

Estamos condicionados à construir modelos visando apenas interpretar os resultados apenas perante o atendimento das Normas da ABNT?

Será que



As restrições/imposições contidas na **ABNT NBR 1463-2** permitem extrair conclusões mais favoráveis do que aquelas que a Teoria Estatística de fato permitiria? Ou vice versa: Premissas impositivas demais perante a Estatística?

Usamos corretamente as **Recomendações do Anexo A** quanto aos pressupostos básicos de normalidade, homocedasticidade, não-multicolinearidade, inexistência de pontos atípicos....



Refinamento do Modelo **Pressupostos Básicos**

Há vários testes estatísticos sugeridos para a determinação de violações importantes dos pressupostos, que podem levar a conclusões diferentes conforme o teste utilizado,

Os testes estatísticos nunca geram evidências de normalidade dos resíduos, homocedasticidade, etc. Podem gerar alertas de não-normalidade, heterocedasticidade, etc, conforme o caso. Tais alertas podem não ser suficientes para saber se houve real comprometimento das inferências e estimativas obtidas pelo modelo de Regressão Linear Múltipla. Se os testes não gerarem tais alertas, ainda assim não se pode garantir a normalidade, a homocedasticidade, etc., principalmente no caso de amostras pequenas.

a norma precisa disciplinar melhor o uso desses teste., principalmente quando o tamanho da amostra é pequeno.



Subjetividade na decisão

- baseada em **exame dos gráficos dos resíduos**, especialmente amostras pequenas.
- Na escolha de uma entre as múltiplas alternativas para a tomada de decisão, por exemplo, aferição da normalidade dos resíduos, em A.2.1.2, **há várias formas sugeridas**, bastando que utilize uma delas (corre-se o risco de escolher a mais conveniente). **Diferentes qualidades entre as múltiplas alternativas para a tomada de decisão** - por exemplo, a forma “b” recomendada para aferição da normalidade dos resíduos pode ser insuficiente, diferentemente da forma “c”.



- **Ausência de Colinearidade** (ou multicolienaridade) **entre as variáveis independentes.**

*Para verificação da multicolinearidade deve-se, em primeiro lugar, analisar a **Matriz das Correlações**, que espelha as dependências lineares de primeira ordem entre as **variáveis independentes**, com atenção especial para resultados superiores a 0,80*

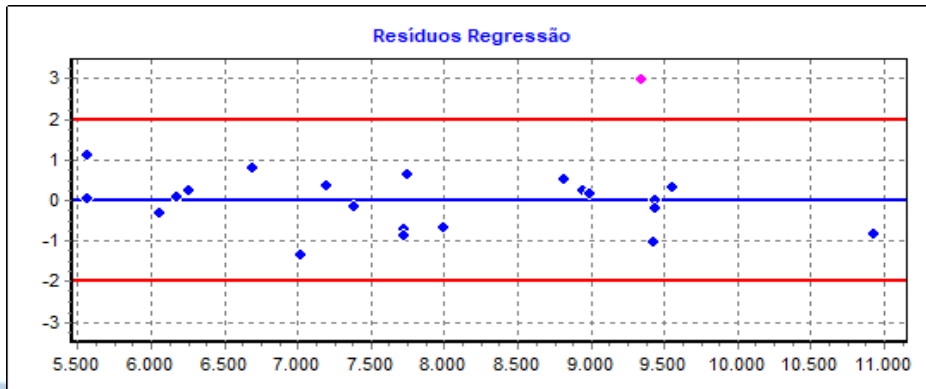
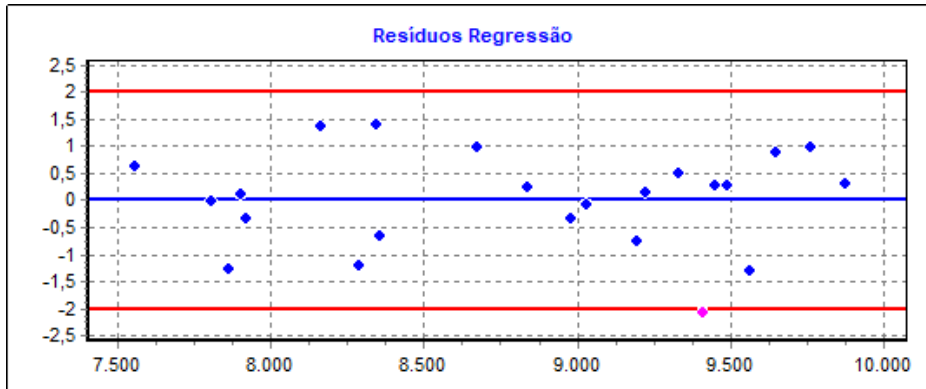
- imprecisão no cálculo dos parâmetros da regressão
- variâncias das estimativas dos parâmetros podem ser muito grandes e acarretar a aceitação da hipótese nula e a eliminação de variáveis fundamentais
- inversão de sinais em relação ao esperado



- Os erros são variáveis aleatórias com distribuição normal

É viável exigir normalidade para pequenas amostras ?

O gráfico mostra uma curva normal padronizada com o eixo horizontal rotulado 'Z - Normal Padronizada'. O eixo tem marcas em -1,96, -1,64, -1,00, 0, 1,00, 1,64 e 1,96. Duas linhas verticais vermelhas estão posicionadas exatamente nos pontos -1,96 e 1,96. À esquerda do ponto -1,96, há um texto: 'Outliers Elementos Discrepantes acima de -2 Desvios Padrão'. À direita do ponto 1,96, há um texto: 'Outliers Elementos Discrepantes acima de 2 Desvios Padrão'. Abaixo do eixo, há três setas horizontais azuis que indicam a porcentagem de dados dentro de certos intervalos: a primeira seta, entre -1,00 e 1,00, está rotulada '68%'; a segunda seta, entre -1,64 e 1,64, está rotulada '89,90%'; a terceira seta, entre -1,96 e 1,96, está rotulada '95%'. Na parte inferior direita, há um texto: 'Por este critério - 5% de Outliers (- ou +) e mantêm as características de Distribuição Normal'.



Nos dois gráficos, existe um *outlier*, representando menos que 5% da amostra (mais de 20 elementos). Entretanto.... No graffito 1 a distribuição é aproximadamente normal (68%, 90% e 95% (se retirado não haverá grandes alterações). Já no Graffito 2, a distribuição é 78, 95 e 95%. (não é normal) se retirado haverá melhoria no modelo)



Pontos influenciantes ou outliers

A Norma cita sem distinguir

Pontos influenciantes ou “outliers”

A existência desses pontos atípicos pode ser verificada pelo gráfico dos resíduos versus cada variável independente, como também em relação aos valores ajustados, ou usando técnicas estatísticas mais avançadas, como a estatística de Cook ou a distância de Mahalanobis para detectar pontos influenciantes.

É importante distinguir os dois conceitos: outliers e pontos influentes. As formas de detecção de um e de outro são distintas, bem como os efeitos sobre o modelo.

No uso das medidas para avaliação dos pontos influentes (exemplo: Cook)



Observações Atípicas: Outlier x Pontos Influenciante

Elementos discrepantes (*Outliers*) e pontos influentes, embora podendo estar relacionados, não são o mesmo conceito: As formas de detecção são distintas, bem como os efeitos sobre o modelo.



Outlier:

Conceito sem definição rigorosa, procura designar observações que se destacam da relação linear, identificados pelo número de desvios padrão de seus afastamentos em relação a 0 (zero) Detectado pela Diferença entre valores previstos pelo modelo e os valores observados na amostra.



Ponto influenciante

Um valor atípico é um caso que difere substancialmente da maioria dos dados. observações distantes da tendência geral podem afetar o próprio ajustamento do modelo, e não serem facilmente identificáveis a partir dos seus resíduos. Pode introduzir tendenciosidade no modelo, pois afeta os valores dos coeficientes de regressão estimados.

