



XIX COBREAP | Foz do Iguaçu

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

AGOSTO 2017

Arq. Urb. Ana Maria de Biazzzi Dias de Oliveira
anabiazzi@uol.com.br (55 11-981097733)

Aspectos **polêmicos** nas Avaliações
utilizando **Inferência Estatística**

Motivação

Inferência estatística (modelos de regressão múltipla) - tem uso na avaliação de bens desde em **1989** (incorporada na ABNT: NBR- 5676/89 e distinguida com “**nível rigoroso**”)

A partir de 2.001 (ABNT- **NBR 14653**) - reformulações com **níveis de fundamentação e níveis de precisão**.

Softwares aplicativos e direcionados, sem os quais a Engenharia de Avaliações não estaria no estágio atual - interface amigável e aplicação da inferência estatística, sem se “entediar” as deduções matemáticas.

Motivação

Estamos condicionados à construir modelos construídos visando interpretar os resultados 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?

Estamos usando corretamente as **Recomendações do Anexo A** quanto aos pressupostos básicos quanto à sua especificação, normalidade, homocedasticidade, não-multicolinearidade, inexistência de pontos atípicos....

Principais pontos polêmicos

- Montagem do modelo e definição das variáveis
- O que é um modelo significativo?
- Até que ponto os testes estatísticos são suficientes para saber o real comprometimento do modelo?
- Colinearidade multicolinearidade
- Por quê é importante analisar os resíduos no método dos mínimos quadrados?
- Distinção entre *outliers* e pontos influentes
- Os problemas de estimativas com amostras pequenas – a questão da precisão dos resultados associados à Amplitude do Intervalo de Confiança- bootstrap

Como quantificar a escala para variável qualitativa para categorias múltiplas ?

.... A Norma indica a seguinte **ordem de prioridade**:

Dicotômica?

Estabelecida por um conjunto de variáveis ("*dummies*") -
Dificuldade **quantidade de dados** para preservar os graus de liberdade necessários à modelagem estatística.

Proxy?

Substitui outra de difícil mensuração e que se presume guardar com ela relação de pertinência. Tem que ser obtida por meio de indicadores publicados ou inferidos em outros estudos.

Código Ajustado?

Escala extraída dos elementos amostrais originais por meio de modelo de regressão, com a utilização de variáveis dicotômicas..... Dificuldade – analisar **só as dicotômicas ou com outras variáveis?**

Código Alocado?

“escala lógica ordenada” ... Dificuldade : **a amplitude da escala pode interferir no resultado**

Variáveis dicotômicas

A categoria base é representada pelo intercepto, se há n categorias, devem haver $n-1$ dummies

Exemplo: três padrão construtivo: a base será o medio

$D_1 = 1$ se padrão Superior; 0 em caso contrário;

$D_2 = 1$ se padrão Alto; 0 em caso contrário

Padrão	X_1	X_2	Y
Médio	0	0	b_0
Superior	1	0	b_0+b_1
Alto	0	1	b_0+b_2

As medidas estabelecidas nos parâmetros medem as diferenças proporcional ao padrão superior

Permite gerar os Códigos Ajustados

Códigos Ajustados (x_1)

b_0/b_0

$(b_0+b_1)/b_0$

$(b_0+b_2)/b_0$

$$Y = b_0 + b_1 * x_1 + b_2 * x_2$$

$$Y = b_0 + b_1 * x_1$$

Variáveis Proxy Padrão Construtivo

Custo unitário básico no Estado de São Paulo*, abril de 2015 em R\$/m²

Padrão Baixo			Padrão Normal			Padrão Alto		
	Custo m²	% mês		Custo m²	%mês		Custo m²	% mês
R-1	1.166,54	0,33	R-1	1.434,26	0,33	R-1	1.720,88	0,41
PP-4	1.074,25	0,37	PP-4	1.350,94	0,31	R-8	1.384,73	0,36
R-8	1.022,91	0,35	R-8	1.179,19	0,34	R-16	1.488,17	0,29
PIS	796,03	0,30	R-16	1.143,85	0,34			

(*) Conforme Lei 4.591 de 16 de dezembro de 1964 e disposto na NBR 12.721 da ABNT. Na formação do Custo Unitário Básico não foram incluídos os itens descritos na seção 8.3.5 da NBR 12.721/06

Custo da construção comercial, industrial e popular no Estado de São Paulo, abril de 2015 em R\$/m²

CAL (comercial andares livres) e CSL(comercial - salas e lojas), GI (galpão industrial) e RP1Q (residência popular)

Padrão Normal			Padrão Alto		
	Custo m²	% mês		Custo m²	% mês
CAL-8	1.358,24	0,32	CAL-8	1.442,17	0,30
CSL-8	1.175,63	0,31	CSL-8	1.271,48	0,30
CSL-16	1.565,79	0,31	CSL-16	1.691,24	0,30
	Custo m²	% mês			
RP1Q	1.277,45	0,38			
GI	664,52	0,35			

Proxy - CUB

1.384,73

1.488,17

1.720,88

(*) Conforme Lei 4.591 de 16 de dezembro de 1964 e disposto na NBR 12.721 da ABNT. Na formação do Custo Unitário Básico não foram incluídos os itens descritos na seção 8.3.5 da NBR 12.721/06

Variáveis qualitativas expressas por Códigos

Alocados

Associar valores numéricos às diferentes categorias.

Os códigos numéricos são apenas uma maneira abreviada de se referir às categorias **e não têm qualquer sentido considerá-los como valores em um estudo de correlação.**

Fundamental **verificar se os códigos correspondem às amplitudes das relações.**

Código Alocado
1
2
3

$$Y = b_0 + b_1 * x_1$$



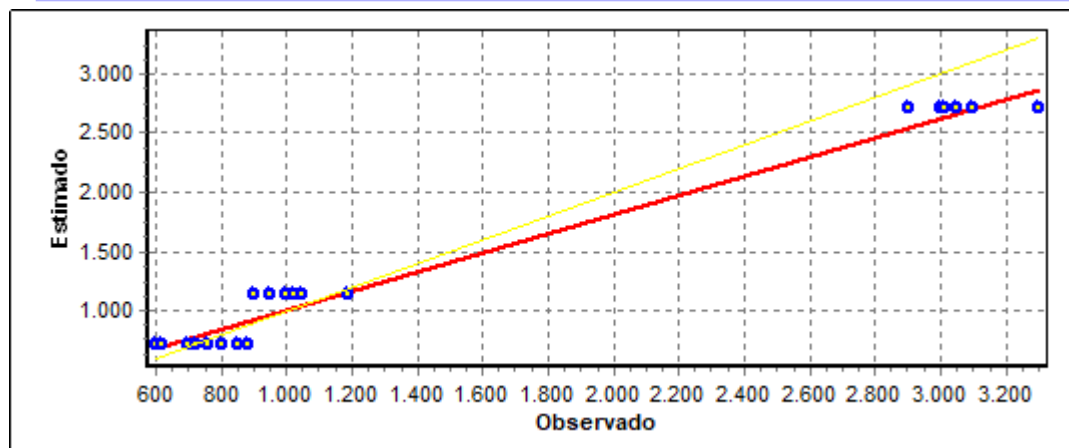
Importante: estes **valores numéricos não tem significado** como tal. Ou seja, ao assumir os valores 1, 2 e 3 para as categorias médio, superior, alto, os números 1, 2 e 3 são simbólicos para representar as categorias, o que não significa padrão superior tenha o dobro do valor do médio, ou o alto, o triplo, do médio.

		x1	x2	Y
	Numero de vagas	2 Vagas	3 vgs	Aluguel
1	2	1	0	1.020,00
2	2	1	0	1.190,00
3	2	1	0	950,00
4	2	1	0	1.000,00
5	2	1	0	1.050,00
6	2	1	0	900,00
7	1	0	0	800,00
8	1	0	0	850,00
9	1	0	0	600,00
10	1	0	0	620,00
11	1	0	0	700,00
12	1	0	0	800,00
13	1	0	0	880,00
14	1	0	0	720,00
15	1	0	0	760,00
16	3	0	1	2.900,00
17	3	0	1	3.300,00
18	3	0	1	3.100,00
19	3	0	1	3.050,00
20	3	0	1	3.300,00
21	3	0	1	3.010,00
22	3	0	1	3.000,00

R\$ 1.018

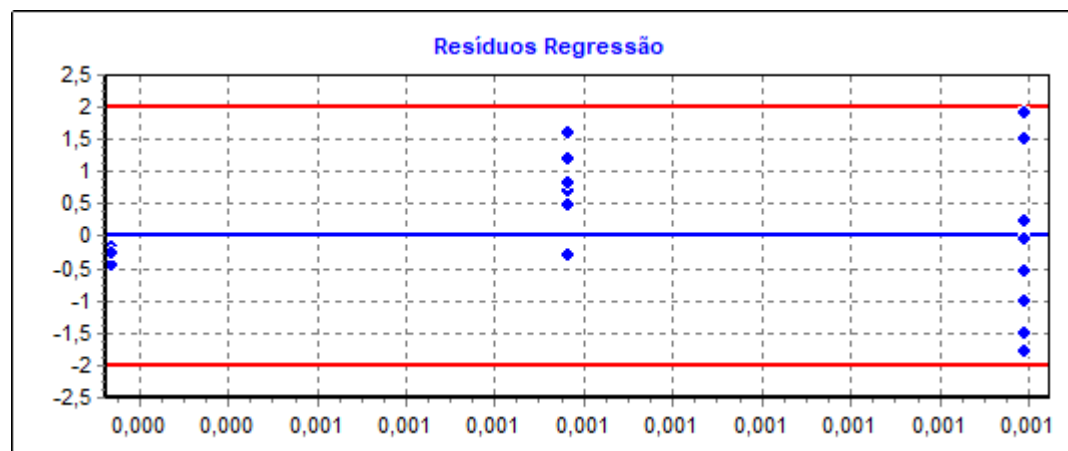
R\$ 747

R\$ 3.094



$$AI = b_0 + b_1 * N. \text{ de Vagas}$$

1	717
2	1.133
3	2.703



R\$	747,00		1,00
R\$	1.018,00	36%	1,34
R\$	3.094,00	313%	4,13

	Dicotomicas	
	x1	x2
	2 Vagas	3 Vagas
1 vaga	0	0
2 Vagas	1	0
3 Vagas	0	1

$$AI = b_0 + b_1 * 2 V + b_2 * 3V$$

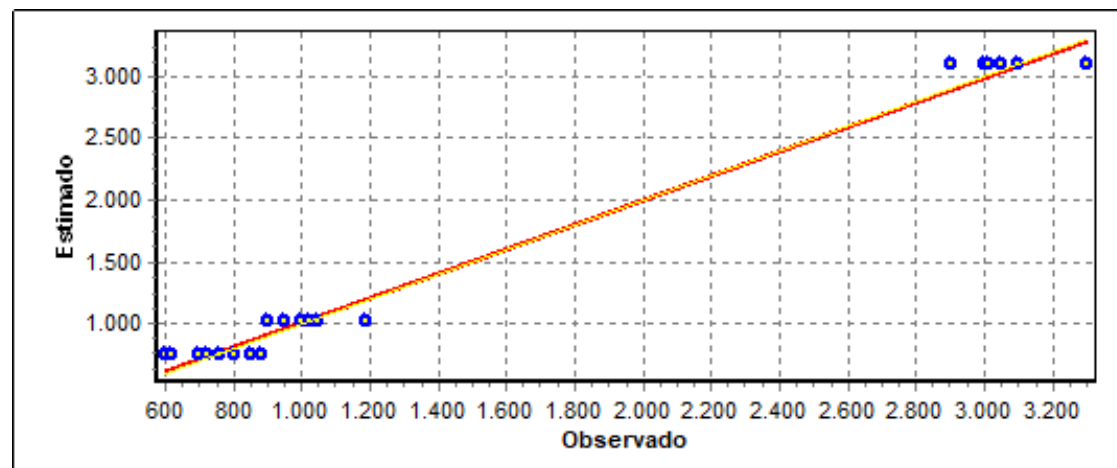
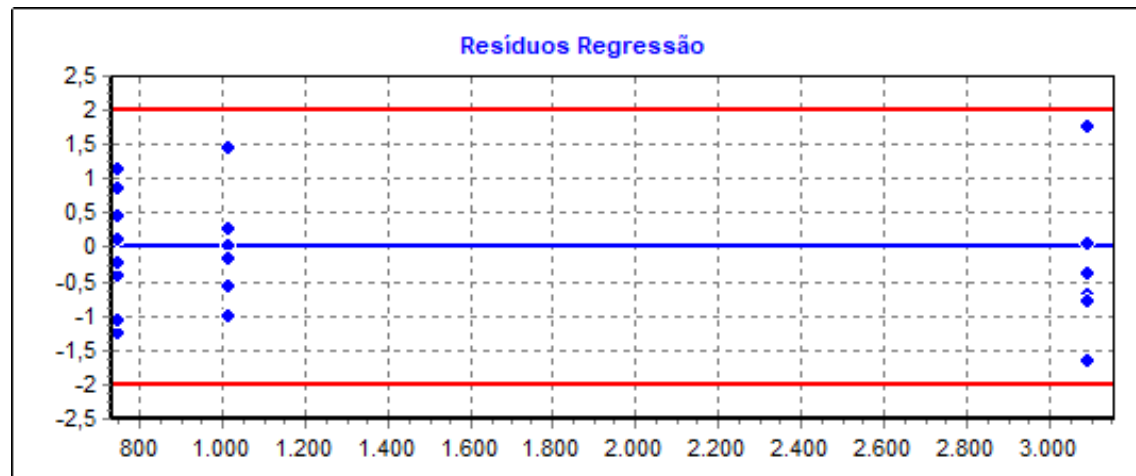
Aluguel =

+747.77

+270.55 * 2 Vagas

+2346.508 * 3 Vagas

		C Ajust
R\$ 747		1,00
R\$ 1.018	36%	1,36
R\$ 3.094	313%	4,13



O que é um modelo de regressão significativo?

Tabela 1- Graus de de Fundamentação NBR 14.653-2

	Descrição	Grau		
		III	II	I
5	Nível de significância (somatório do valor das duas caudas) máximo para a rejeição da hipótese nula de cada regressor (teste bicaudal)	10 %	20 %	30 %
6	Nível de significância máximo admitido para a rejeição da hipótese nula do modelo através do teste F de Snedecor	1 %	2 %	5 %

É suficiente apenas “enquadrar” nos níveis de significância para obter bons resultados?

Regressão linear inferencial - Teste F

Testa a significância do modelo de regressão, permite avaliar se as previsões feitas a partir do modelo de regressão são significativamente mais acertadas o **uso da média da variável dependente**.

Este teste de significância é uma **Análise de Variância - ANOVA** e equivale a um teste simultâneo da significância de todos os parâmetros de regressão.

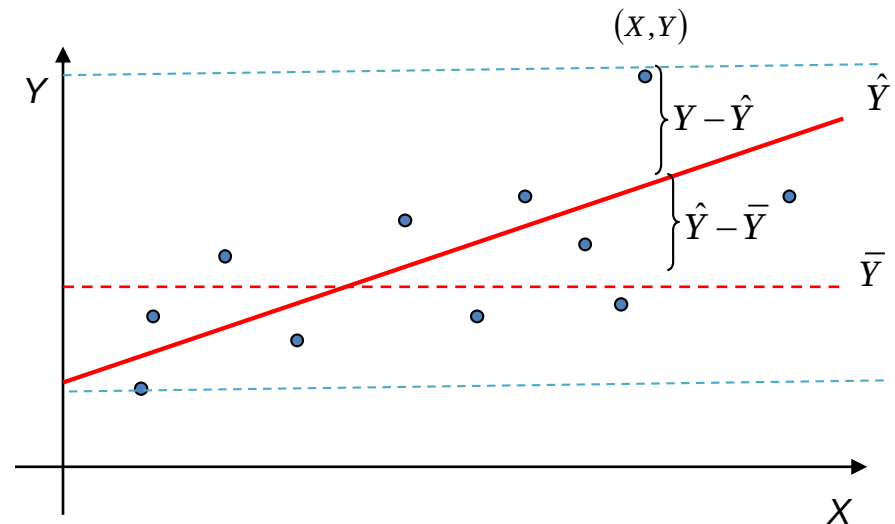
O teste global do modelo de regressão não nos informa se todos ou se apenas alguns dos preditores contribuem significativamente. Para isso, é necessário testar a significância individual de cada preditor/regressor (teste t).

O teste de significância dos coeficientes da regressão permite identificar aqueles cujo contributo para a explicação da variável dependente **não é diferente de zero**. Com base nesta informação, exclui-se do modelo de regressão preditores que não tenham contributo significativo. Garante-se assim que o modelo seja mais eficaz.

Decomposição da soma dos quadrados

Objetivo: estudar a variação da variável dependente Y .

- Que parcela da variação é causada pela variação de X ?
- Que parcela da variação não é “explicada” pela variação de X ?



$$Y - \bar{Y} = (\hat{Y} - \bar{Y}) + (Y - \hat{Y}) \Rightarrow y = \hat{y} + \varepsilon$$

Análise de variância (ANOVA)

A análise de variância (ANOVA) compara **se o modelo estimado é melhor** do que o **modelo nulo** (sem nenhuma variável independente) – **media da variável dependente**.

A **estatística F** é calculada a partir da divisão da média dos quadrados atribuída à regressão pela média dos quadrados dos resíduos.

Fonte de variação de Y	Soma de Quadrados	Graus de Liberdade	Quadrado médio das variações	Razão F	Valor -P
E = Explicada pela regressão $X_1...X_n$	SQE	K	$QME = SQE/k$	$F = QME/QMR$	Significância obtida da Tabela F
R= Residual não explicada pela regressão	$SQR = SQT - SQE$	n-k-1	$QMR = SQR/(n-k-1)$		
T=Total	$SQT = (SQE + SQR)$	TOTAL	TOTAL		

Análise de variância (ANOVA)

Fonte de variação de Y	Soma de Quadrados	Graus de Liberdade	Quadrado médio das variações	Razão F	Valor -P
E = Explicada pela regressão $X_1 \dots X_n$	SQE	K	$QME = SQE/k$	$F = QME/QMR$	Significância obtida da Tabela F
R= Residual não explicada pela regressão	$SQR = SQT - SQE$	n-k-1	$QMR = SQR/(n-k-1)$		
T=Total	$SQT = (SQE + SQR)$	TOTAL	TOTAL		

R^2

F

Teste de significância do Modelo

$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$

$H_1 : \text{pelo menos um } \beta_j \neq 0,$

verificar se existem variáveis colineares
Importante para selecionar variáveis importantes

Amostras Homogêneas

As Normas admitem uso do modelo de regressão nulo

9.2.1.6.1 *No caso de amostras homogêneas⁷, será adotada a Tabela 1, com as seguintes particularidades:*

- a) serão admitidos os itens 3 e 4 apenas no Grau III, de forma a ficar caracterizada a homogeneidade;*
- b) será atribuído o Grau III para os itens 5 e 6, por ser nulo o modelo de regressão.*

Teste de significância dos coeficientes de regressão

... além da **nível de significância** e dos **Graus de Liberdade**



A.1.4 Na modelagem devem ser expostas as hipóteses relativas aos comportamentos das variáveis dependentes e independentes, com base no conhecimento a respeito do mercado, quando serão formuladas as hipóteses nula e alternativa para cada parâmetro.

Interpretar resultados de regressão e observar:

- a correspondência entre o **sinal dos coeficientes** e a relação teoricamente esperada: se os resultados dão evidências em favor das hipóteses.
- se resultados sugerem a rejeição da hipótese quando o resultado apontam na direção contrária. Nesse caso, revisar o que poderia estar produzindo tal correlação inesperada.
- o peso explicativo dos coeficientes e interpretar se as estatísticas estimadas se relacionam com a sua hipótese de pesquisa.

Por exemplo

Valor unitário = 7429.22

- 10.67119 * **Área** (82 a 131) (3,67%)
- + 2.3113 * **Padrão Cub** (três níveis) (1,65%)
- + 959.57 * **Vagas** (dicotômica 0/1 -2 ou 3 vagas) (0,02%)
- 53.836 * **Idade Estimativa** (1 a 15 anos) (10,75%)
- 207.318 * **Varanda Gourmet** (sim 1, não 0) (51,37%)

Peso dos coeficientes: por exemplo vagas (no valor médio projetado para o apartamento de \$8.315 \$959 representa 11,51%

Hipótese incoerente (ou a hipótese esta errada, neste caso modelo certo, ou **Matriz de correlações isoladas** das variáveis independentes identifica **correlação de -0,87**, entre **idade e A Varanda ser ou não Gourmet**

Suspeita de Colinearidade?

Multicolinearidade

O modelo de regressão de mínimos quadrados ordinários utiliza apenas a variação única de cada variável para estimar os coeficientes, ignorando a variância compartilhada. Eis a essência do problema: quanto maior a correlação entre as variáveis independentes, menos informação estará disponível para estimar os coeficientes associados às variáveis explicativas.

Colinearidade ou multicolinearidade (A.2.1.2 ABNT 14.653-2)

A.2.1.5.1 Uma forte dependência linear entre duas ou mais variáveis independentes provoca degenerações no modelo e **limita** a sua utilização. **As variâncias das estimativas dos parâmetros podem ser muito grandes e acarretar a aceitação da hipótese nula e a eliminação de variáveis fundamentais.**

A.2.1.5.2 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.** Como também é possível ocorrer multicolinearidade, mesmo quando a matriz de correlação apresenta coeficientes de valor baixo, recomenda-se, também, verificar o correlacionamento de cada variável com subconjuntos de outras variáveis independentes, por meio de regressões auxiliares, como pela **análise de variância por partes.**

A.2.1.5.3 Para tratar dados na presença de multicolinearidade, é recomendável que sejam tomadas medidas corretivas, como a ampliação da amostra ou adoção de técnicas estatísticas mais avançadas, a exemplo do uso de **regressão de componentes principais.**

A.2.1.5.4 Nos casos em que o imóvel avaliando segue os padrões estruturais do modelo, a existência de multicolinearidade pode ser negligenciada.

O coeficiente de correlação simples entre duas variáveis aleatórias independentes acima de 0,8 – é uma referência de indício de problema de colineariedade

A retirada de variáveis explicativas do modelo, **não deve ser feita por causa de seu baixo valor de t, sob risco de retirar-se importante variável**

No modelo do apartamento, se retirar a variável idade

.... A **variável Varanda** apresenta relação teoricamente esperada de aumentar o valor do apartamento, dando evidência à favor da hipótese.

Valor unitário =

+6835,668718

-11,95707915 * Área privativa

+2,368808338 * Padrão construtivo - CUB

+1002,512992 * Vagas garagem

+201,8970559 * Varanda gourmet

Portanto qualquer uma das duas variáveis poderá permanecer ou ser retirada no modelo

Passos para avaliar a multicolineariedade:

1) Testar a amostra

Há indicação de multicolineariedade, por exemplo, quando o **teste t indica insignificância estatística** dos estimadores e **R² ou estatística F são altos**.

2) Nessa situação, a matriz de correlação deve ser investigada. Retirar as variáveis independentes altamente correlacionadas exceto uma.

É importante ressaltar que:

1. É possível haver variáveis independentes altamente correlacionadas (altos coeficientes de correlação) e a regressão não ter problemas de multicolinearidade.
2. Se o **teste t indicar significância do estimador**, é sinal que a multicolinearidade não é séria para fins de previsão.

No exemplo do modelo do apartamento, descartando variável varanda

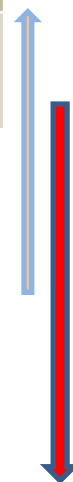
Verificar ocorrência de multicolinearidade, pela *análise de variância por partes*

Análise da Variância				
Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F calculado
Explicada	11.400.541	5	2.280.108	19,43
Não explicada	1.994.581	17	117.328	
Total	13.395.123	22		

R^2	85,11%
R^2 Ajustado	80,73%

Análise da Variância				
Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F calculado
Explicada	11.348.347	4	2.837.086	24,95
Não explicada	2.046.776	18	113.709	
Total	13.395.123	22		

R^2	84,72%
R^2 Ajustado	81,32%



Coeficiente de determinação $0 \leq R^2 \leq 1$

Nada nas hipóteses do modelo linear clássico exige que o R^2 esteja acima de qualquer valor em particular.

O R^2 é simplesmente uma estimativa do quanto da variação em y é explicado pela variação de $x_1, x_2, \dots, x_k$.

- 1) R^2 sempre aumentará se adicionar mais variáveis ao modelo, porém não significa necessariamente que o modelo ficará melhor.
- 2) Analisar R^2 pode identificar o ponto em que a adição de um novo parâmetro não aumente sensivelmente o valor de R^2 .

O R^2 ajustado auxilia na escolha de modelo sem variáveis independentes redundantes r leva em conta a chance de contribuição de cada variável incluída, subtraindo-se o valor que seria esperado se nenhuma variável independente fosse associada à variável dependente.

Detectando a Colinearidade

FIV - Fator Inflacionário da Variância

FIV_j é usado para medir a colinearidade:

$$FIV_j = \frac{1}{1 - R_j^2}$$

onde R_j^2 é o coeficiente de determinação de um modelo de regressão que usa X_j como variável dependente e todas as demais variáveis X como variáveis independentes

Regra para o VIF

até 1 - sem multicolinearidade

de 1 até 10 - multicolinearidade aceitável

acima de 10 - multicolinearidade problemática

Independente dos testes - Necessário analisar os problemas no modelo

Para que servem os testes estatísticos?

Os testes estatísticos nunca geram evidências de normalidade dos resíduos, homocedasticidade, etc.

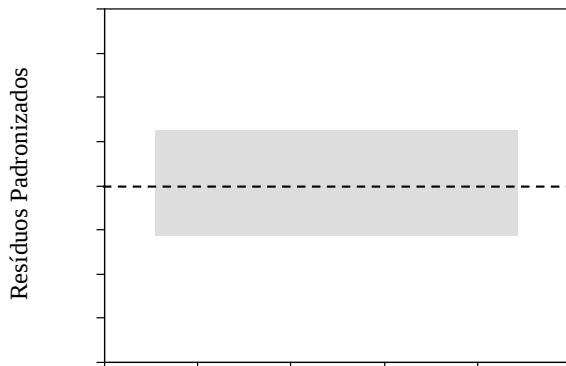
Podem gerar alertas de não-normalidade, heterocedasticidade, etc, conforme o caso.

Tais alertas **podem não ser suficientes para o avaliador saber se houve real comprometimento das inferências e estimativas obtidas pelo modelo de Regressão Linear Múltipla.**

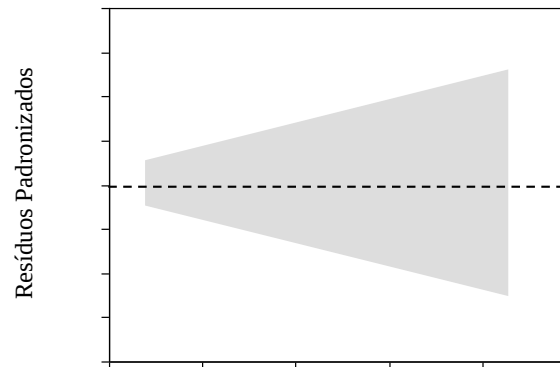
Se os testes não gerarem tais alertas, ainda assim não se pode garantir a normalidade, a homocedasticidade, etc., principalmente no caso de amostras pequenas.

Fundamental - Análise de Resíduos

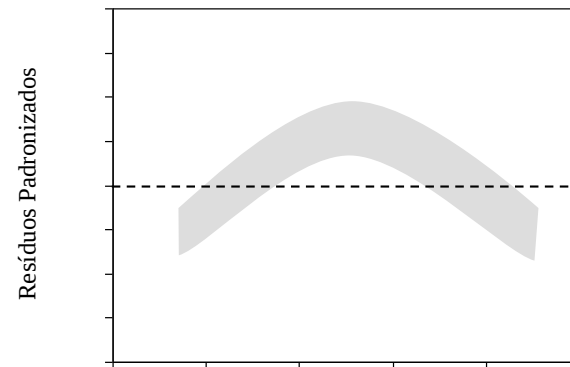
"ideal"



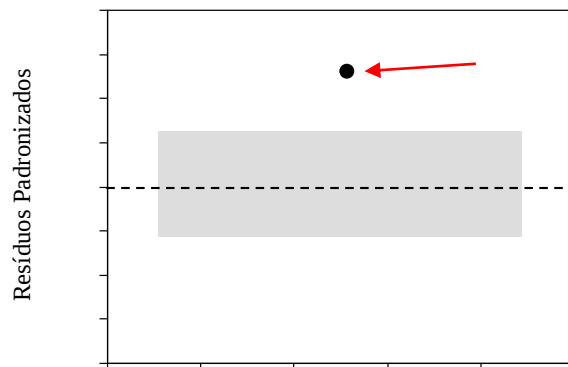
σ^2 não constante



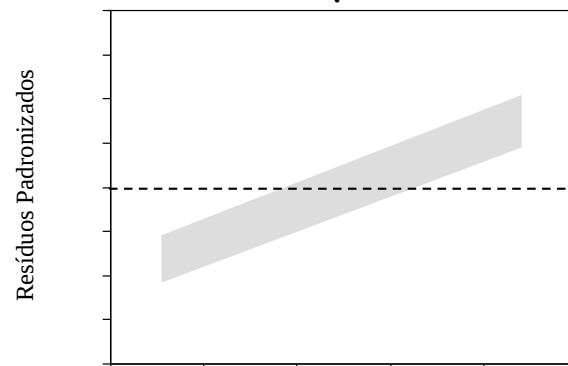
não linearidade



X
"outlier"



não independência



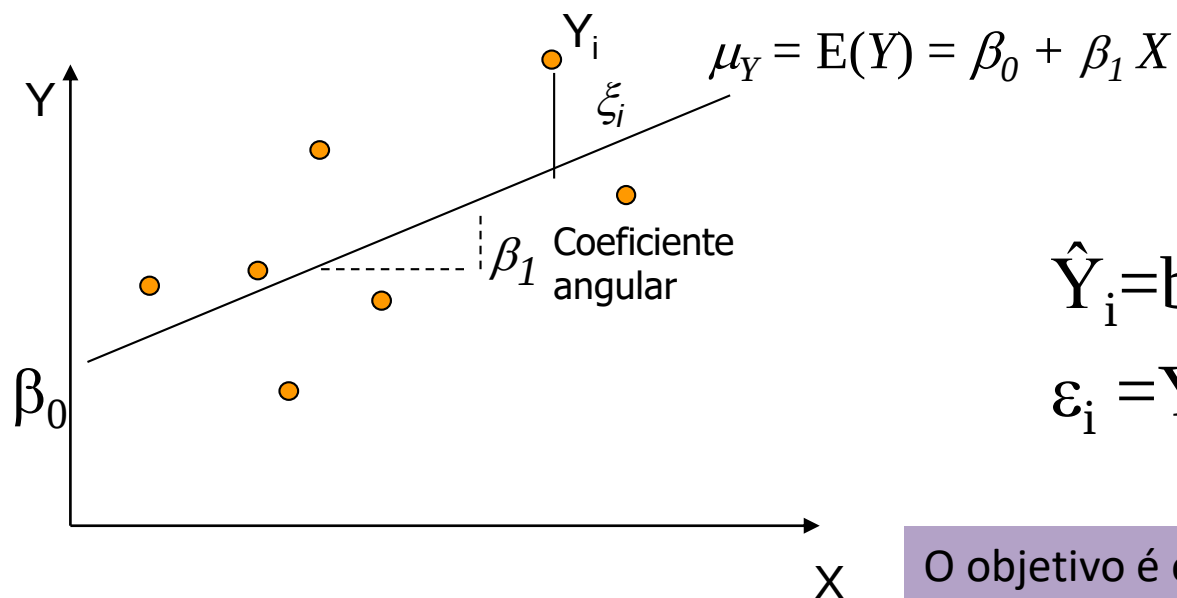
■ A **linearidade** é facilmente examinada utilizando:

- Gráficos de Dispersão
- Gráfico de resíduos

OBS: O que justifica a transformação da variável não é a maior aderência e sim a análise dos resíduos.

Porque é tão importante as premissas básicas modelo e o diagnóstico dos **resíduos**?

O modelo ajustado utilizando a forma funcional de mínimos quadrados ordinários é assim denominada porque **minimiza os erros de estimação** entre os valores observados e os valores preditos, ou seja, minimiza os resíduos.



$$\hat{Y}_i = b_0 + b_1 X_i \quad \text{Modelo estimado}$$

$$\varepsilon_i = Y_i - \hat{Y}_i \quad \text{Resíduo}$$

O objetivo é encontrar a equação tal que a **soma dos quadrados dos desvios** seja mínima

Normalidade

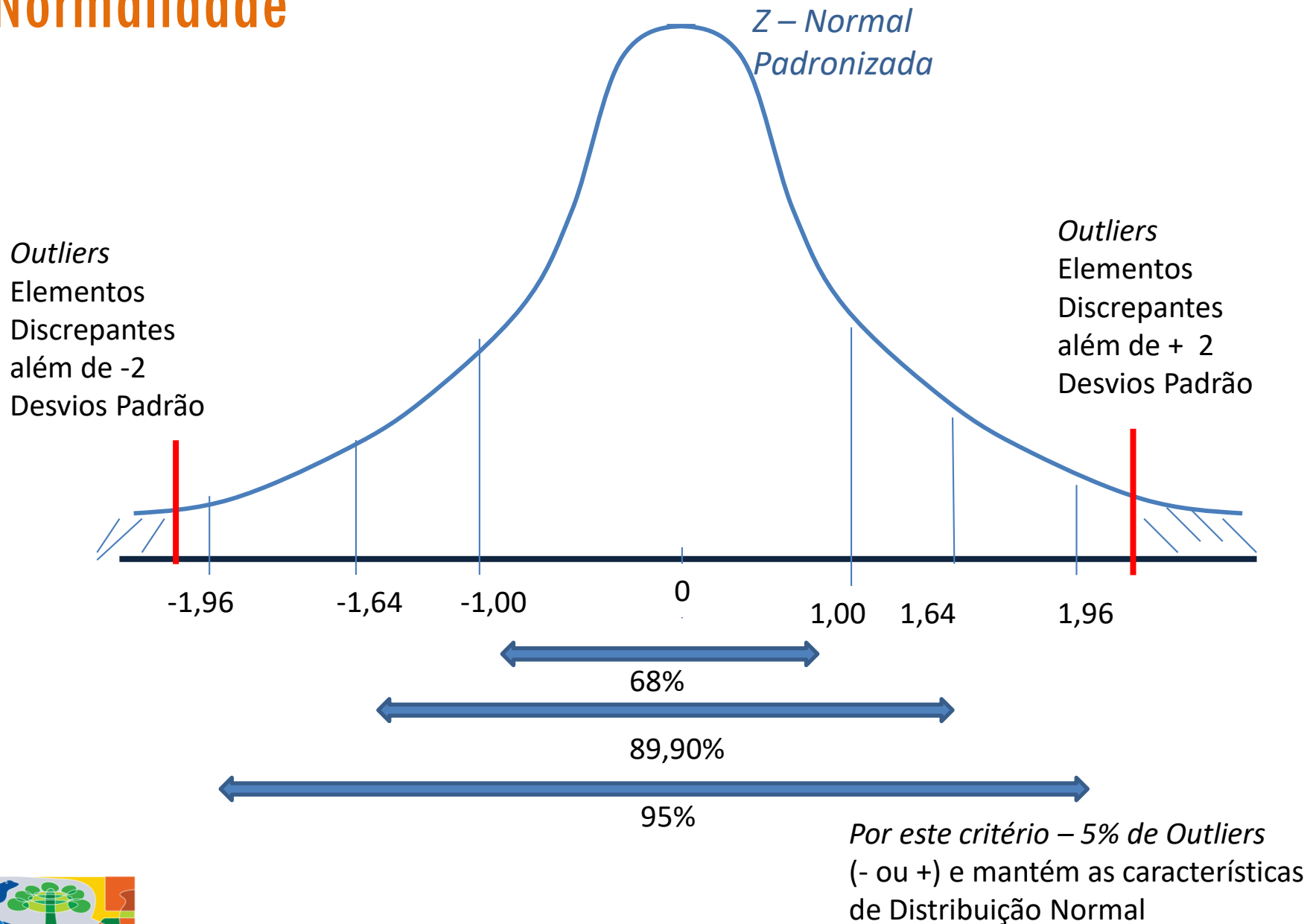
Por que a premissa de normalidade?

Propriedade da distribuição Normal: qualquer função linear de variáveis com distribuição normal também é normalmente distribuída.

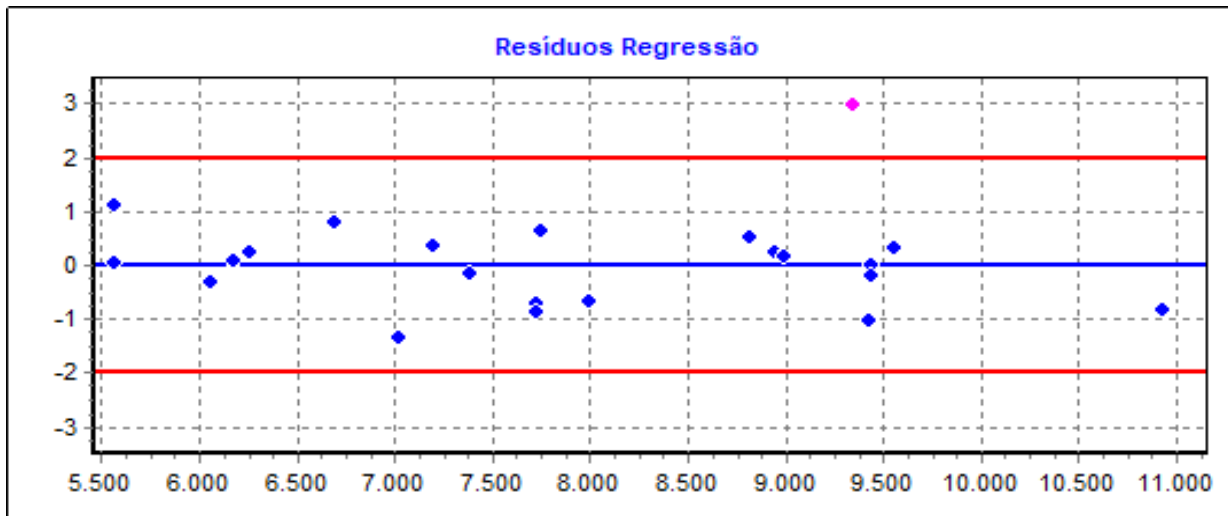
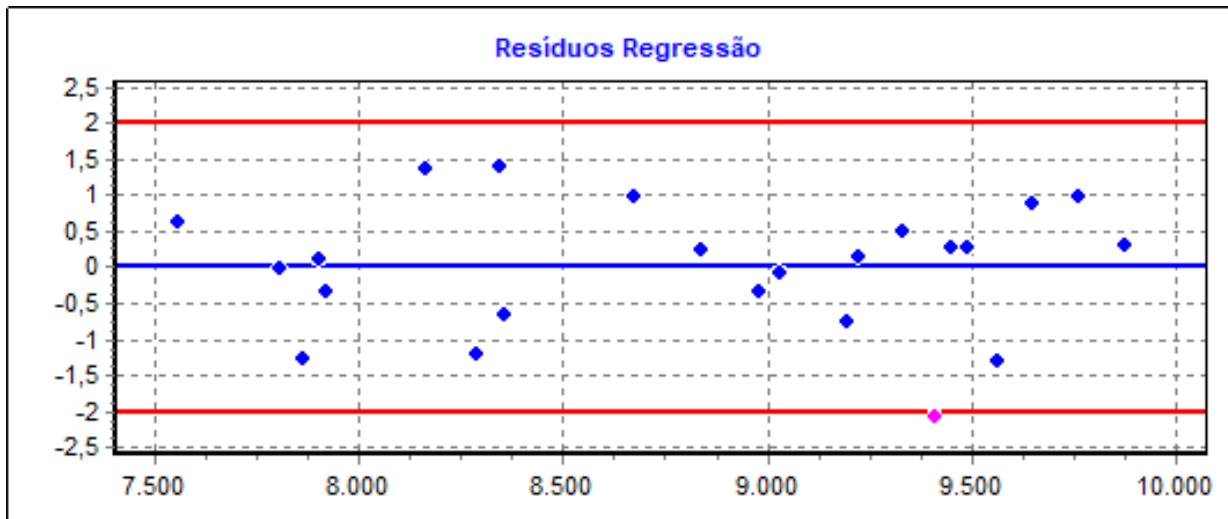
Espera-se que u_i seja resultado da influência de variáveis omitidas ou negligenciadas e que esse efeito seja pequeno e aleatório.

Pelo Teorema do Limite Central a soma de um grande número de variáveis aleatórias, independentes e com distribuição idêntica, tende à distribuição normal.

Normalidade



Será que basta aceitar 5% de outliers para concluir pela Normalidade?



Nos dois gráficos, existe um outlier, representando menos que 5% da amostra (mais de 20 elementos).

Entretanto.... No gráfico 1 a distribuição é aproximadamente normal (68%, 90% e 95% (se retirado não haverá grandes alterações).

Já no Gráfico 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.

Pontos inflenciantes ou ouliers

Existem varias razões para o aparecimento de pontos discrepantes e geralmente podem ser consequência de:

erros de mensurarão decorrentes de registro incorreto

inadequação do elemento amostral;

a observação é legítima e está dentro da especificação do modelo e apesar de nada de improvável estar ocorrendo, forma um ponto discrepante em comparação com os demais.

A identificação de “outliers” e análise das causas que levaram ao seu surgimento podem resultar em aperfeiçoamento do modelo com incremento de novos conceitos

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:

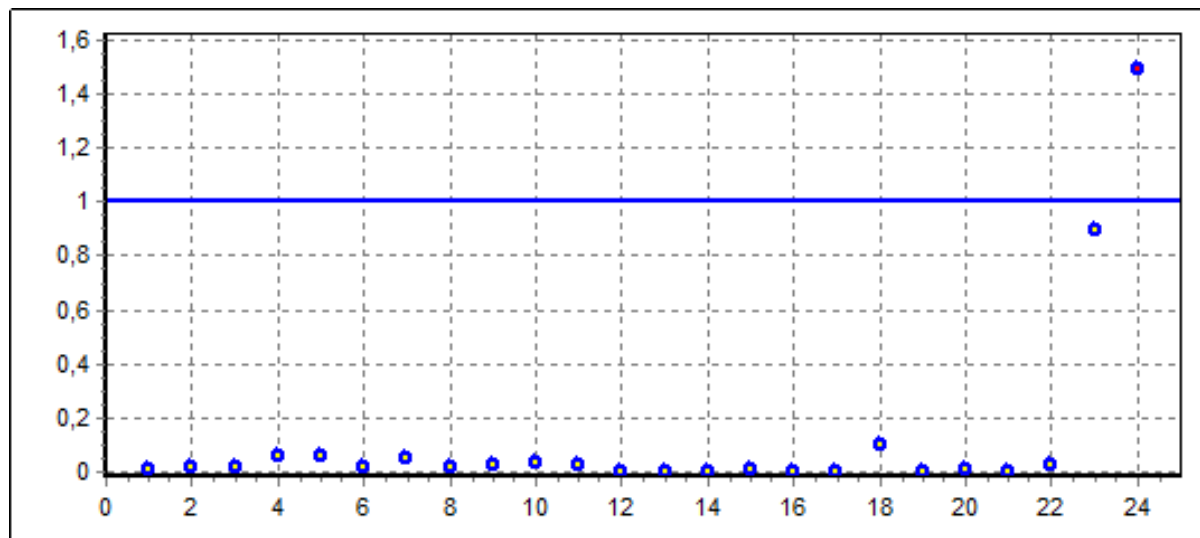
Conceito sem definição rigorosa, procura designar observações que se destacam da relação linear, identificados pelo número de desvios padrão de seus afastamentos em relação a 0 (zero). Detectado pela Diferença entre valores previstos pelo modelo e os valores observados na amostra.



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.

Detecção de pontos de influência



Existem variações nas formulas
Alguns programas a identificação pode ser 0,5 ou 2

Distância de Cook (se $D > 1$, o ponto é influente na regressão – poderá também ser *outlier*)

O elemento 24, que é um valor atípico muito influente (distancia 1,48). Já o elemento 23

$D = 0,8916$, tem um valor de influência na regressão, mas está dentro da gama de valores usuais

Transformações (automáticas) das variáveis

Nos casos em que os resíduos não tem distribuição aproximadamente normal e/ou a variância não é constante (é usual que esses problemas ocorram juntos), recomenda-se a transformação da variável dependente. Em algumas situações a transformação simultânea da variável independente, X , e da variável dependente, Y , é necessária.

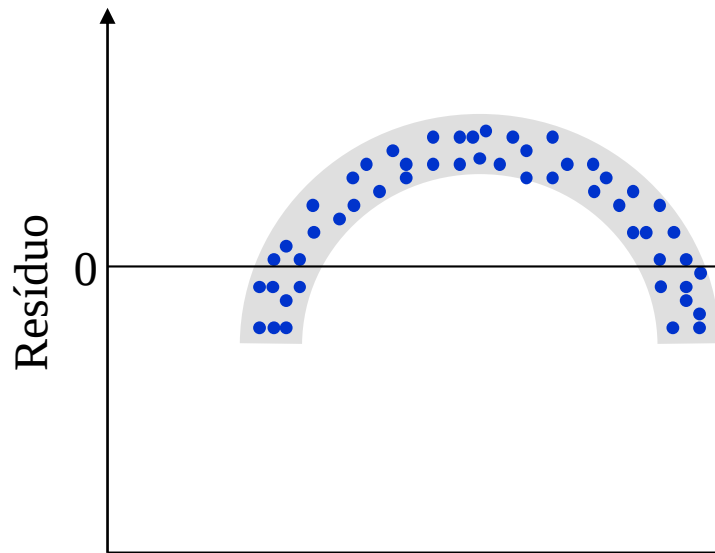
As transformações de dados fornecem um meio para modificar as variáveis devido:

- Para corrigir violações das suposições estatísticas inerentes ao método do mínimos quadrados
- Para melhorar a relação correlação entre as variáveis

Se num gráfico de pontos dos resíduos (erros) versus a resposta prevista, o espalhamento não é homogêneo.

- Então os resíduos são ainda uma função das variáveis previsoras.
- A transformação da resposta pode resolver o problema

Transformações automáticas das variáveis



Não Linearidade

OBS: O que justifica a transformação da variável não é a maior aderência e sim a análise dos resíduos.

- Se uma relação não-linear é encontrada, a abordagem mais direta é transformar uma ou as duas variáveis de modo a ter linearidade.
- Várias transformações podem ser usadas:

TIPO	EQUAÇÃO	TRANSFORMAÇÃO	VARIÁVEL X	VARIÁVEL Y
Linear	$Y = a + bx$	$Y = a + bx$	x	y
Exponencial	$Y = a \cdot e^{bx}$	$\ln y = \ln a + bx$	x	$\ln Y$
Logarítmica	$Y = a + b \cdot \ln x$	$Y = a + b \cdot \ln x$	$\ln x$	y
Potência	$Y = a \cdot x^b$	$\ln y = \ln a + b \cdot \ln x$	$\ln x$	$\ln y$

A transformação logarítmica.

O uso de logaritmos das variáveis dependentes ou independentes é o artifício mais comum para permitir relações não lineares entre a variável explicada e as variáveis explicativas.

Log estreita amplitude dos valores, tornando estimativas menos sensíveis a observações extremas (*outliers*).

Interpretando modelos com log

- Se o modelo é $\ln(y) = b_0 + b_1 \ln(x) + u$:

b_1 é a elasticidade de y em relação a x .

Ou seja: escala das variáveis não importa, pois o que se mede são variações percentuais. Ele dá uma estimativa direta da elasticidade.

- Se o modelo é $y = b_0 + b_1 \ln(x) + u$:

b_1 é aproximadamente a variação em y de uma variação de 100% em x .

- Se o modelo é $\ln(y) = b_0 + b_1 x + u$:

b_1 é aproximadamente a variação percentual em y dada uma variação de 1 unidade em x .

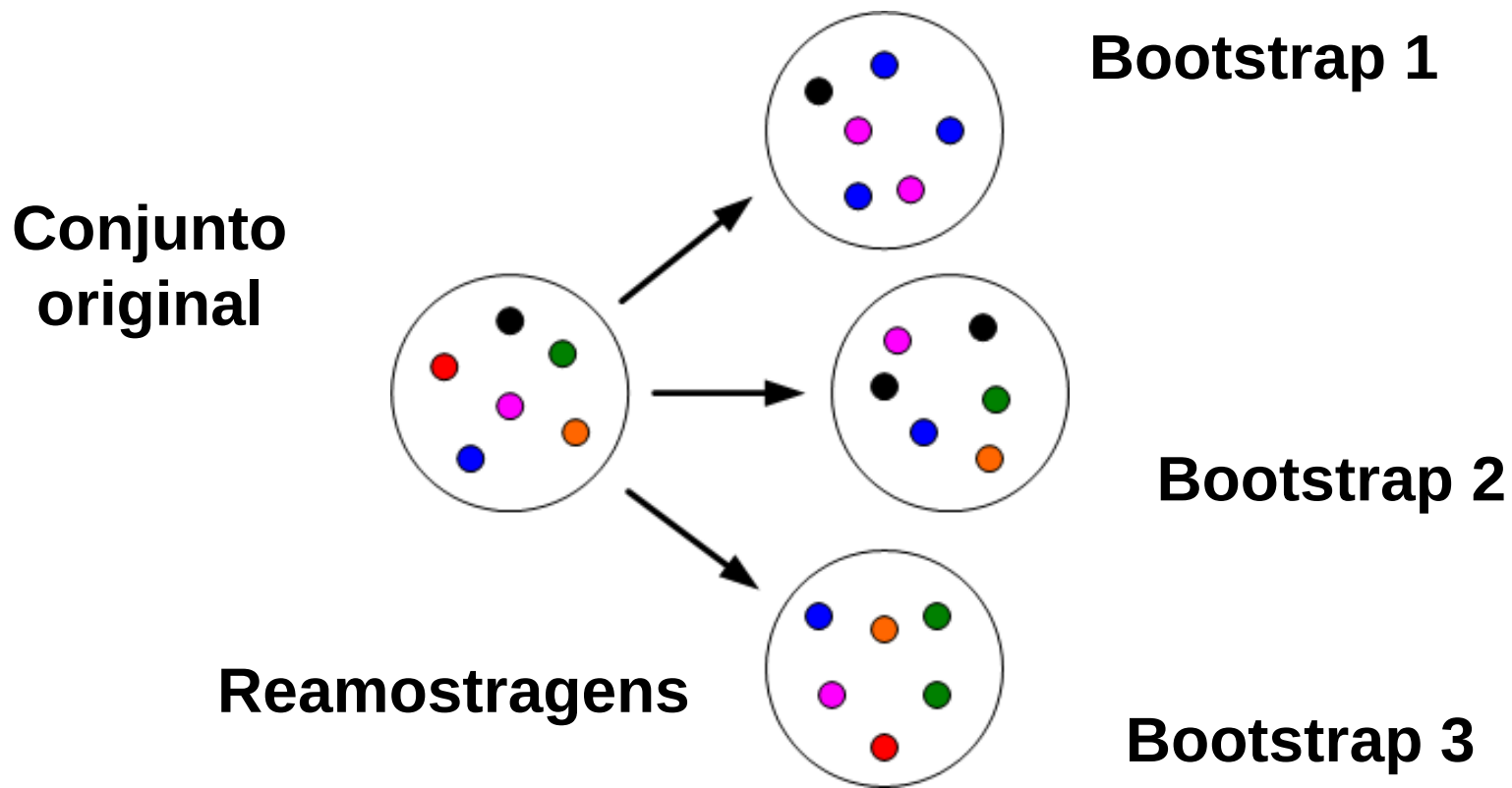
Estimação de Intervalos de Confiança via Reamostragem *Bootstrap*

A ideia básica da técnica *bootstrap* é tratar a amostra original como se fosse a população de interesse e, retirar várias amostras com reposição dela, ou seja, *reamostrar* a amostra original com reposição e, para cada *reamostra*, calcular a estimativa de interesse.

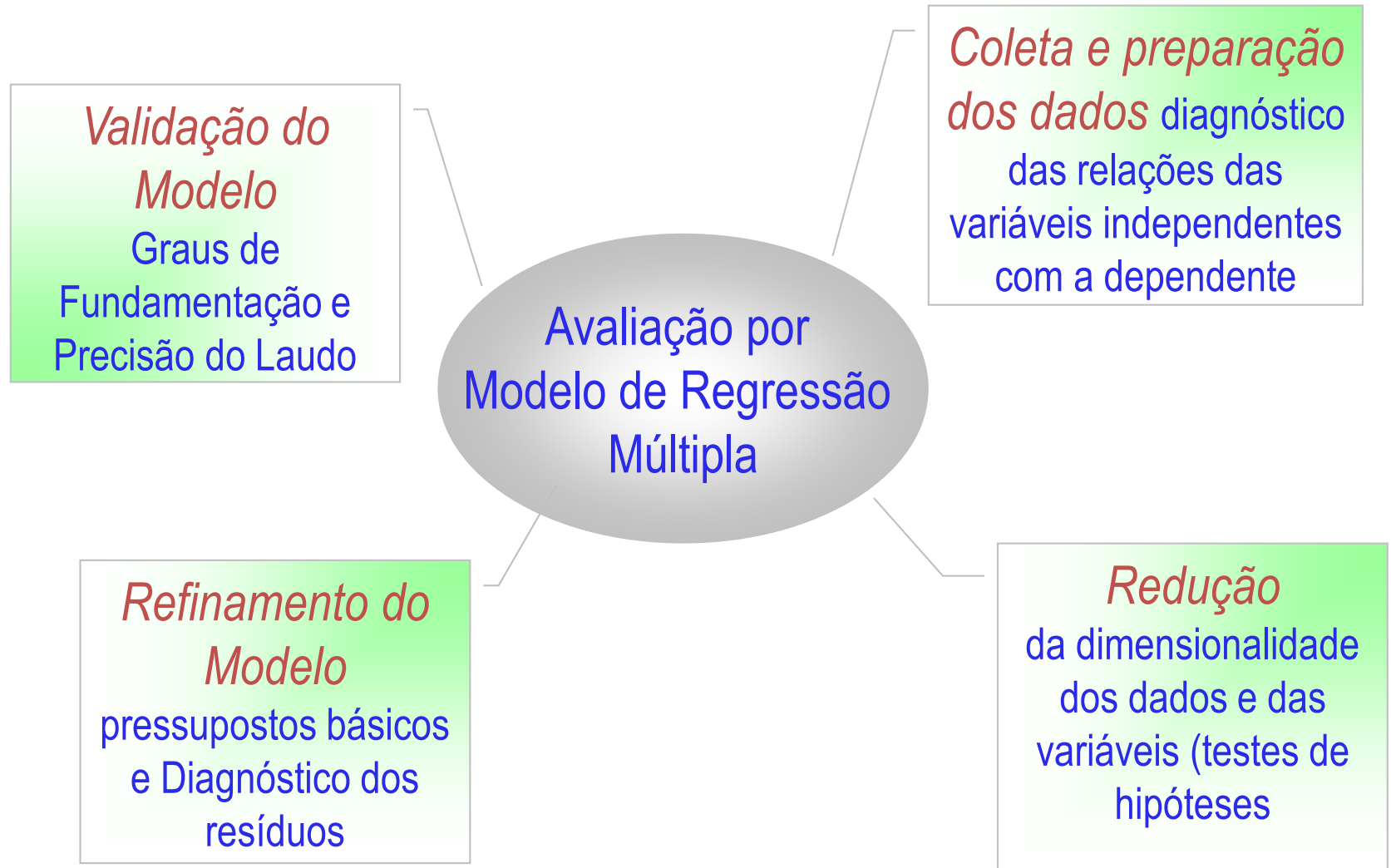
Ao final da reamostragem tem-se várias estimativas distintas para cada reamostra, porque se faz amostragem com reposição em cada caso, criando assim, uma aproximação da população original.

Estimação de Intervalos de Confiança via *Reamostragem Bootstrap*

Logo que eles estejam todos calculados, pode-se realizar o teste de normalidade dos valores e até mesmo construir intervalos de confiança e realizar testes de hipóteses.



Principais Etapas



Obrigada

“ Uma resposta aproximada para o problema certo vale muito mais do que uma resposta exata para um problema aproximado. ”

(John Wilder Tukey)

Bibliografia recomendável:

Estatística:

Estatística Aplicada à Administração – William J. Stevenson – Editora Harbra, Ed. 2001

Estatística Básica – Wilton O. Bussab – Editora Saraiva – Ed. 2002

Estatística Aplicada Serie Essencial – Douglas Downing & Jeffrey Clark – Editora Saraiva

Estatística Aplicada à Administração e Economia. ANDERSON et al. (2007) 2a ed. São Paulo: Cengage Learning.

Estatística Básica. MORETTIN, P. A. & BUSSAB, W. O. (2010) 6a ed. São Paulo: Saraiva.

Estatística Aplicada e Probabilidade para Engenheiros, Douglas C. Montgomery - Editora LTC

Análise Multivariada de Dados, vários autores (Hair, Black, Babin, Anderson, Tathan)

Econometria

Econometria Básica – Damodar N. Gujarati – Makron Books – 5ª. Edição

Introdução á Econometria - Uma Abordagem Moderna; Autor: Jeffrey M. Wooldridge

Editora: THOMSON

Econometria – Carter Hiil, William Griffiths, George Judge – Editora Saraiva – 2ª. Ed. 2.002