



XIX COBREAP | Foz do Iguaçu

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

AGOSTO 2017

Sérgio Antão Paiva

INTERPRETAÇÃO DE RESULTADOS ESTATÍSTICOS EM MODELOS DE REGRESSÃO MÚLTIPLA

ENFOQUE DA COMPARAÇÃO

- **Princípio da semelhança:** numa mesma data, dois bens semelhantes, em mercados semelhantes, têm valores semelhantes
- **Método comparativo direto de dados de mercado:** o valor do imóvel corresponde ao valor de imóveis semelhantes

AMOSTRA E POPULAÇÃO

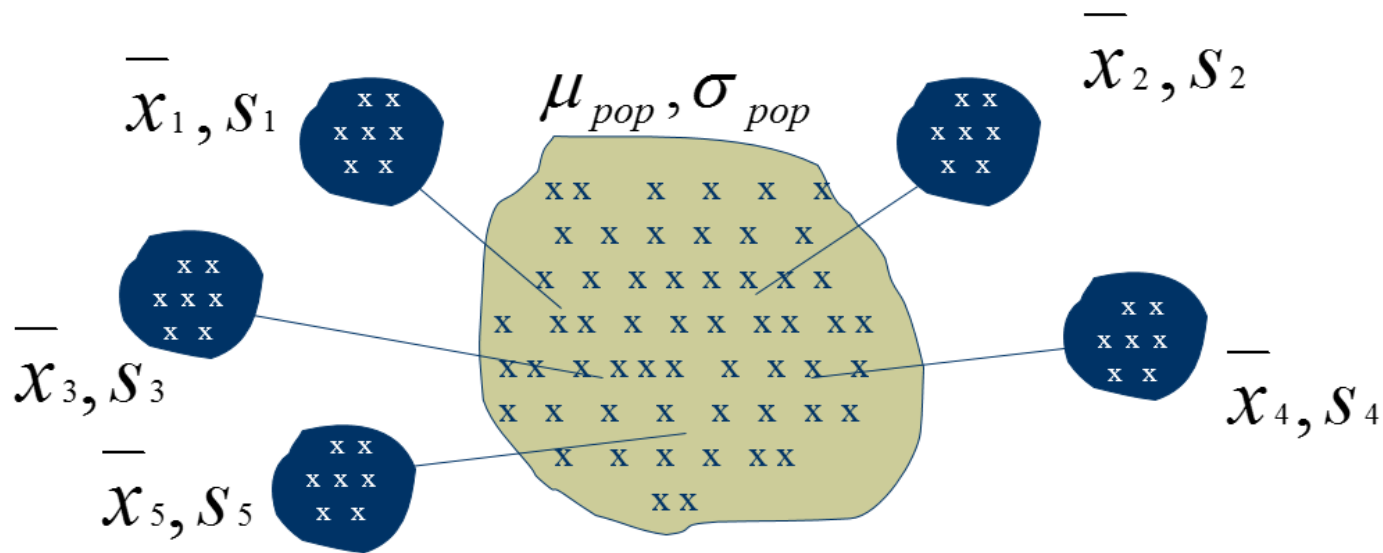
- Em geral, a quantidade de dados do mercado imobiliário é pequena
- Se os dados correspondem à **população** e os imóveis forem **homogêneos**:
 - Valor corresponde à **média** aritmética dos preços observados na população

média:
$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

AMOSTRA E POPULAÇÃO

Problema: inferir o valor dos **parâmetros** da população, a partir de **estatísticas** da amostra

Há possibilidade de cometer **erro**



AMOSTRA E POPULAÇÃO

Se dados correspondem a uma **amostra** (e não a uma **população**)

- Modelo estatístico inferencial
- Estimativa pontual
- Estimativa intervalar
 - Intervalo de confiança para a média da população

AMOSTRA E POPULAÇÃO

- **Teorema central do limite**
 - **População** qualquer com **média** μ e **desvio padrão** σ
 - **Amostras** de **n** elementos
 - À medida que **n** cresce:
 - A distribuição das médias amostrais $\bar{X}$ aproxima-se de uma distribuição normal com:

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \sigma/\sqrt{n}$$

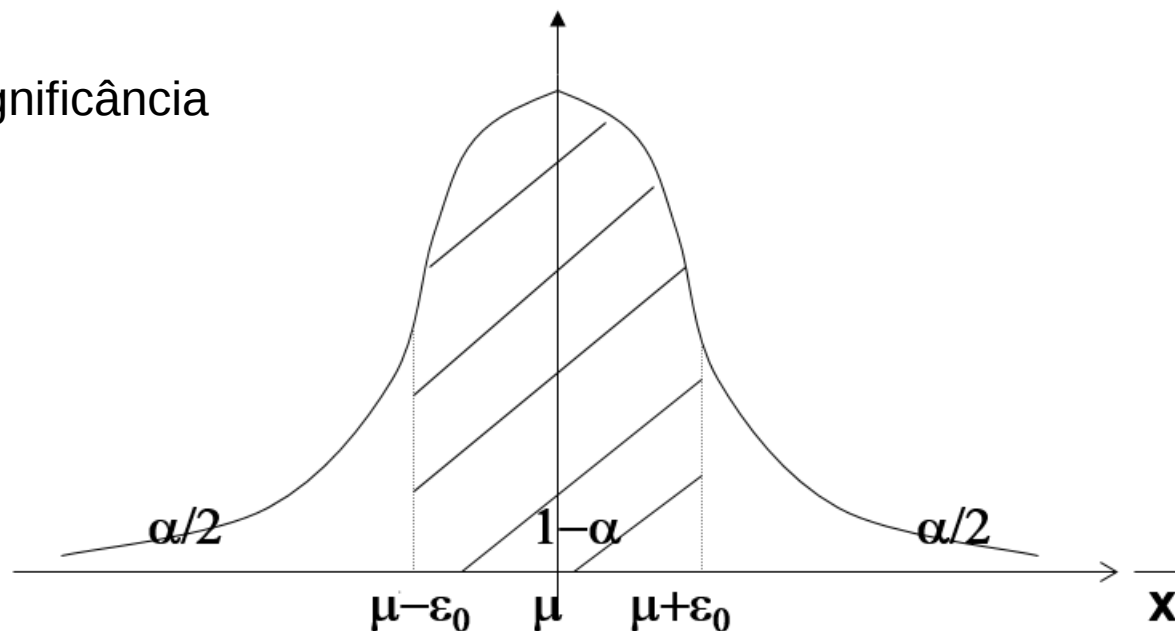
AMOSTRA E POPULAÇÃO

- **Intervalo de confiança** para a média amostral com probabilidade de $1-\alpha$

$1 - \alpha$ = confiança

α = significância

$\alpha / 2$ = metade da significância



AMOSTRA E POPULAÇÃO

Intervalo de confiança para a média amostral com probabilidade de $1-\alpha$

$$\Pr[(\mu - \varepsilon_0) \leq \bar{X} \leq (\mu + \varepsilon_0)] = (1 - \alpha)$$

$$(\mu - \varepsilon_0) \leq \bar{X} \rightarrow \mu \leq (\bar{X} + \varepsilon_0)$$

$$\bar{X} \leq (\mu + \varepsilon_0) \rightarrow \mu \geq (\bar{X} - \varepsilon_0)$$

$$\Pr[(\bar{X} - \varepsilon_0) \leq \mu \leq (\bar{X} + \varepsilon_0)] = (1 - \alpha)$$

AMOSTRA E POPULAÇÃO

- **Intervalo de confiança** para a média amostral com probabilidade de $1-\alpha$
- Transformação dos limites do intervalo para a utilização da curva normal padronizada

$$Z_{\text{sup}} = \frac{(\bar{X} + e_0) - \bar{X}}{s / \sqrt{n}} \rightarrow e_0 = \frac{Z_{\text{sup}} \cdot s}{\sqrt{n}}$$

$$Z_{\text{inf}} = \frac{(\bar{X} - e_0) - \bar{X}}{s / \sqrt{n}} \rightarrow e_0 = -\frac{Z_{\text{inf}} \cdot s}{\sqrt{n}}$$

$$\lim_{\text{inf,sup}} = \bar{X} \pm e_0 = \bar{X} \pm \frac{Z \cdot s}{\sqrt{n}}$$

AMOSTRA E POPULAÇÃO

Amostra aleatória simples

- As n observações que compõem a **amostra** ($X_1, X_2, \dots, X_n$) são **independentes**
- A distribuição dos X_i corresponde à **distribuição da população**, com média μ e variância σ^2

AMOSTRA E POPULAÇÃO

Particularidades do mercado imobiliário

O mercado imobiliário caracteriza-se como um **mercado imperfeito**, com **bens heterogêneos**, **estoque limitado**, **liquidez diferenciada** e grande influência de **fatores externos**.

Fonte: Projeto de revisão da NBR 14.653-1

AMOSTRA E POPULAÇÃO

Particularidades do mercado imobiliário

Nesse mercado, o avaliador tem acesso, muitas vezes, apenas aos **preços** desejados pelos vendedores (as **ofertas**) ou a **preços de transações sujeitos a deformações** decorrentes dos interesses do informante, influenciando na **amostra** que é coletada pelo avaliador no método comparativo direto de dados de mercado.

Fonte: Projeto de revisão da NBR 14.653-1

AMOSTRA E POPULAÇÃO

Particularidades do mercado imobiliário

Características do mercado e prazos limitados para as pesquisas levam à coleta de amostras que não atendem na íntegra aos pressupostos formais das **amostras aleatórias simples**

Amostras coletadas são mais bem descritas como **amostras acidentais** e o avaliador deve se esforçar para que elas contenham a maior **representatividade** possível em relação à população e **não sejam tendenciosas**

AMOSTRA E POPULAÇÃO

Objetivo da pesquisa é a coleta de uma amostra de dados de mercado **representativos**:

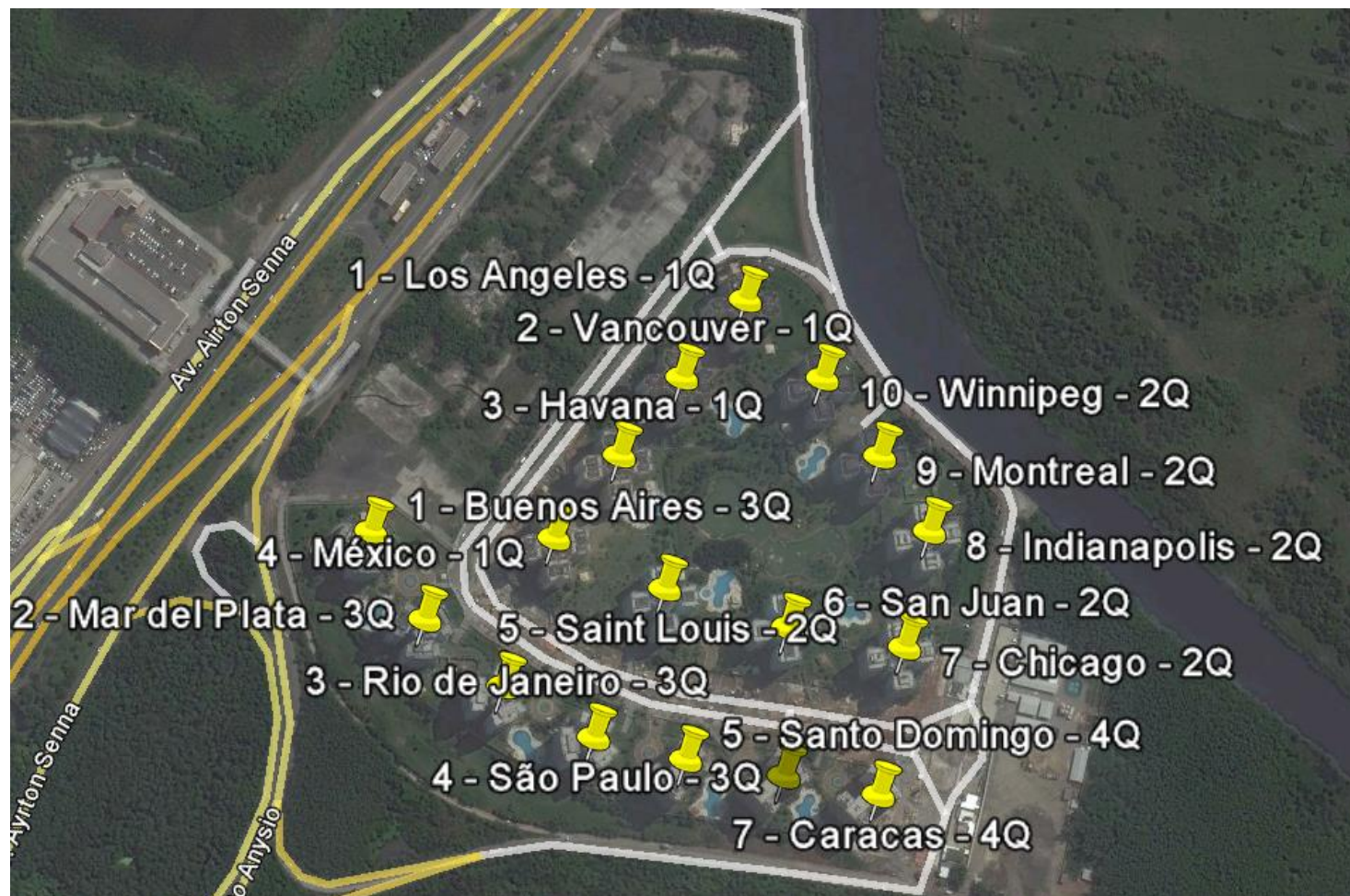
- **Semelhantes** ao avaliando
- **Contemporâneos**
- De **fontes** de informação **diversificadas**
- Em **quantidade suficiente** para o modelo estatístico-inferencial

AMOSTRA E POPULAÇÃO

Exemplo 1

Apartamentos de 1 quarto na
Vila Pan-Americana

VILA PANAMERICANA



Efeito Tschebotarioff

AMOSTRA E POPULAÇÃO

Número insuficiente de dados:

- Afrouxar os **critérios de semelhança**
 - Aumentar o nível de abstração
- **Tratar as diferenças** por meio de **variáveis explicativas (variáveis independentes)** em modelo de regressão linear (ou redes neurais, ou EDO/DEA, etc.)

AMOSTRA E POPULAÇÃO

Coleta de uma amostra **representativa**:

- **Esgotamento** da pesquisa, na medida do possível (**amostra acidental**)
- Segurança quanto às **características** físico-locacionais e negociais dos dados de mercado
- Segurança quanto aos **preços observados** (ofertas e transações) dos dados de mercado

AMOSTRA E POPULAÇÃO

Dados de mercado heterogêneos

- Se corresponderem à **população**
 - Modelo de regressão linear da população
 - Valor ajustado na função de regressão
- Se corresponderem a uma **amostra**
 - Modelo de regressão linear da amostra
 - Modelo estatístico inferencial
 - **Intervalo de confiança** para a média da população, para um determinado conjunto de X_i
 - **Intervalo de previsão** para uma observação isolada, para um determinado conjunto de X_i

MODELO

- Modelo é uma representação **simplificada** da realidade
- O modelo considera o que é **relevante** e abstrai os aspectos **irrelevantes**
- Se o ponto de partida é uma **amostra**, da qual se pretende estimar os **parâmetros da população**, o modelo é estatístico-inferencial (ex.: **modelo de regressão linear**)

MODELO

QUALIDADES DESEJÁVEIS

Plausibilidade

- Sentido e magnitude da variação são coerentes com a teoria ou com o que é observado no mundo real

Poder de **explicação**

- Modelo deve explicar grande parte das variações [na regressão linear: R^2]

MODELO

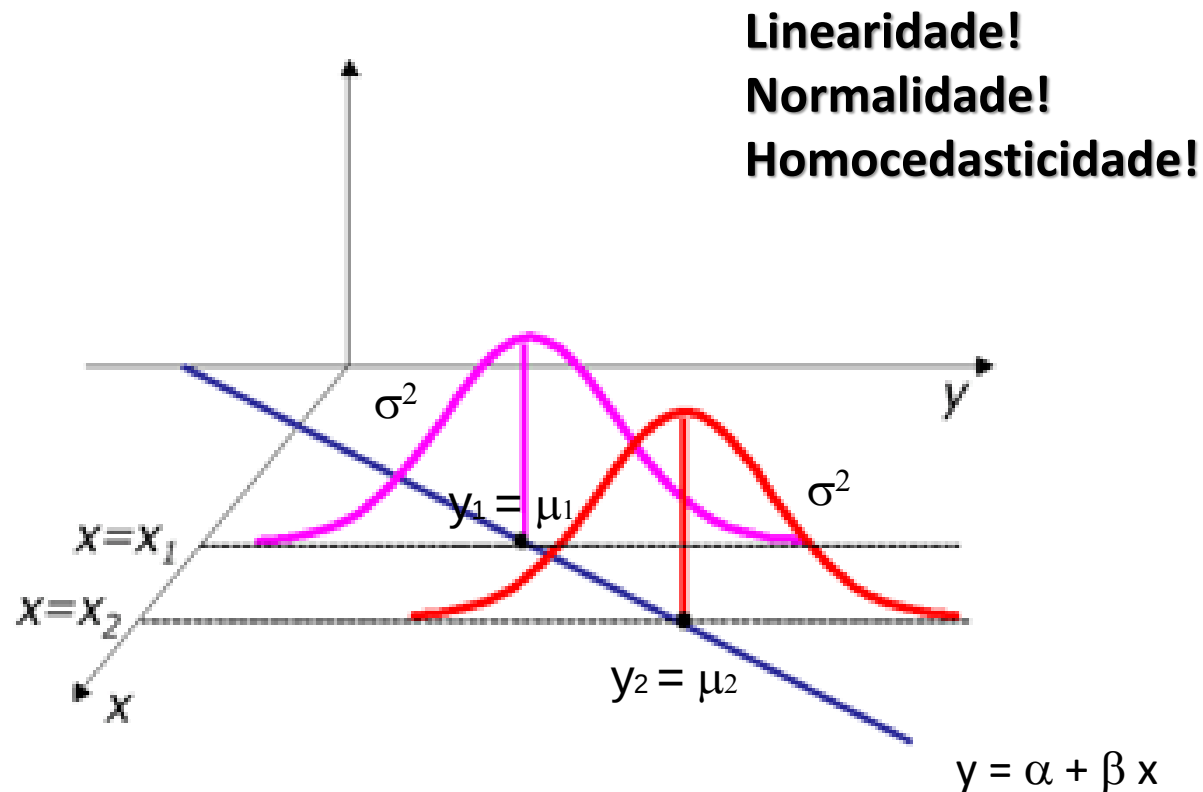
QUALIDADES DESEJÁVEIS

- Precisão na estimativa dos **parâmetros**
[na NBR 14.653-2: grau de precisão da **estimativa**]
- Capacidade de **previsão**
- Simplicidade

MODELO DE REGRESSÃO LINEAR

Regressão simples $Y = \beta_0 + \beta_1 X_1 + \varepsilon$

A regressão da **população**: a verdadeira reta de regressão

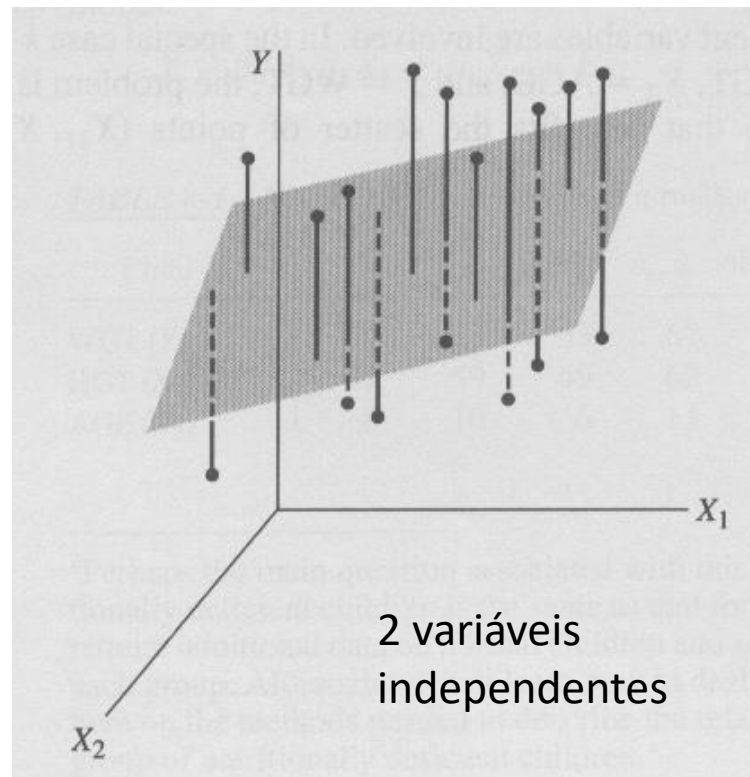


MODELO DE REGRESSÃO LINEAR

População:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

k variáveis
independentes



Amostra:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k + e$$

ESTIMADOR DE MÍNIMOS QUADRADOS ORDINÁRIOS (MQO)

$$\sum_{i=1}^n e_i = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki}))^2$$

- Modelo minimiza a soma dos resíduos quadráticos
- Resíduo: diferença entre o valor **observado** Y_i e o valor **ajustado** no modelo $\hat{Y}_i$

Estimadores de mínimos quadrados

- Solução: valores de b_j para os quais a soma dos resíduos quadráticos é mínima

ESTIMADOR DE MÍNIMOS QUADRADOS ORDINÁRIOS (MQO)

Regressão múltipla (forma matricial)

$$\frac{\partial[(Y - Xb)'(Y - Xb)]}{\partial b} = 0$$

$$\frac{\partial(Y'Y - Y'Xb - b'X'Y + b'X'Xb)}{\partial b} = 0$$

$$-2X'Y + 2X'Xb = 0$$

$$X'Xb = X'Y$$

$$b = (X'X)^{-1} X'Y$$

REGRESSÃO LINEAR SIMPLES

Exemplo 2

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

PRESSUPOSTO DA LINEARIDADE

- A relação entre a variável dependente e as variáveis independentes **deve ser linear** (ou *linearizável*)
- Do contrário: modelo de regressão **não linear**
- **Transgressão:** MQO não é o melhor estimador linear não-viesado

PRESSUPOSTO DA LINEARIDADE

NBR 14.653-2

A.2.1 Verificação dos pressupostos do modelo

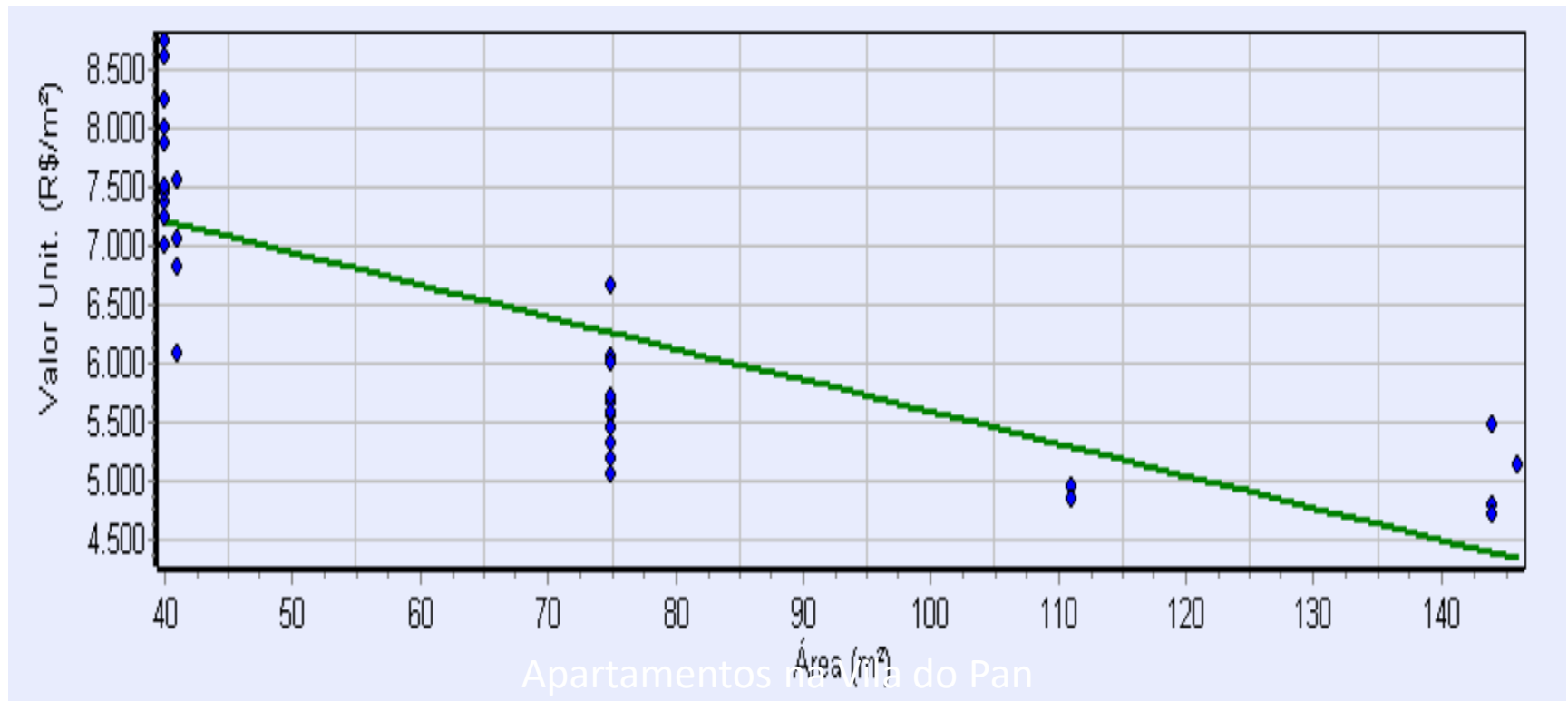
A.2.1.1 Linearidade

Recomenda-se ser analisado, primeiramente, o comportamento gráfico da variável dependente em relação a cada variável independente, em escala original. Isto pode orientar o avaliador na transformação a adotar. Existem formas estatísticas de se buscar a transformação mais adequada, como, por exemplo, os procedimentos de Box e Cox.

As transformações utilizadas para linearizar o modelo devem, tanto quanto possível, refletir o comportamento do mercado, com preferência pelas transformações mais simples de variáveis, que resultem em modelo satisfatório.

Após as transformações realizadas, se houver, examina-se a linearidade do modelo, pela construção de gráficos dos valores observados para a variável dependente *versus* cada variável independente, com as respectivas transformações.

PRESSUPOSTO DA LINEARIDADE



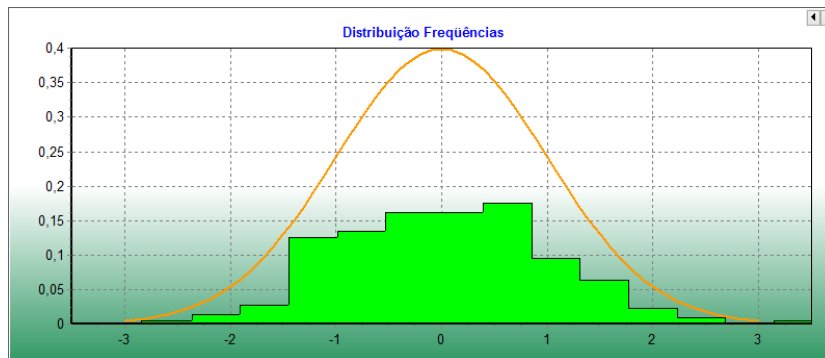
Exemplo 2

PRESSUPOSTO DO NÚMERO DE DADOS EM RELAÇÃO AO NÚMERO DE PARÂMETROS

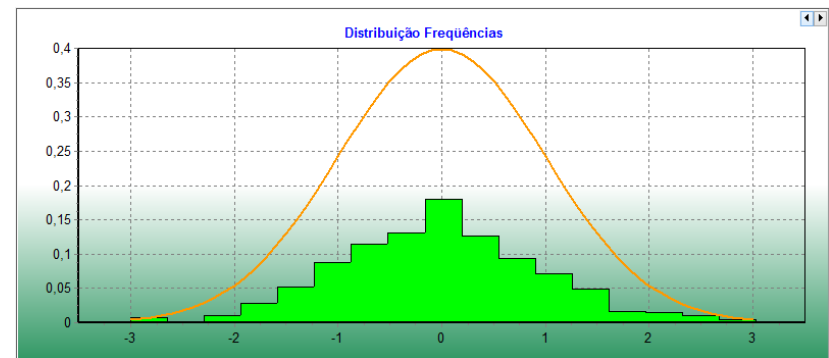
- O número de dados deve ser superior ao número de parâmetros estimados
- **Transgressão:** o cálculo matemático dos parâmetros torna-se impossível

PRESSUPOSTO DO NÚMERO DE DADOS EM RELAÇÃO AO NÚMERO DE PARÂMETROS

É desejável maximizar o tamanho da amostra, em relação ao número de parâmetros, pois a distribuição amostral de uma variável aleatória converge para a distribuição normal com o aumento do tamanho da amostra (TCL)



222 dados



720 dados

PRESSUPOSTO DO NÚMERO DE DADOS EM RELAÇÃO AO NÚMERO DE PARÂMETROS

NBR 14.653-2

- a) para evitar a micronumerosidade, o número mínimo de dados efetivamente utilizados (n) no modelo deve obedecer aos seguintes critérios, com respeito ao número de variáveis independentes (k):

$$\underline{n \geq 3(k + 1)}$$

Tabela 1 — Grau de fundamentação no caso de utilização de modelos de regressão linear

Item	Descrição	Grau		
		III	II	I
1	Caracterização do imóvel avaliando	Completa quanto a todas as variáveis analisadas	Completa quanto às variáveis utilizadas no modelo	Adoção de situação paradigma
2	Quantidade mínima de dados de mercado, efetivamente utilizados	6 ($k + 1$), onde k é o número de <u>variáveis independentes</u>	4 ($k + 1$), onde k é o número de <u>variáveis independentes</u>	3 ($k + 1$), onde k é o número de variáveis independentes



REGRESSÃO LINEAR MÚLTIPLA

Exemplo 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

Novas variáveis: andar, nº de quartos e estado
da infraestrutura

PRESSUPOSTO DO NÚMERO DE DADOS EM RELAÇÃO AO NÚMERO DE PARÂMETROS

Variáveis			Dados	
Total	Consideradas		Total	Considerados
6	4		38	38
Graus de Liberdade			34	
Determinação				Correlação
Linear		Não Linear		Linear
0,861038		0,853178		0,927922
Fisher-Snedecor			Desvio Padrão	
F calculado		Significância		Linear
70,22		0,01		Não Linear
				0,07
				488,47
Normalidade dos Resíduos			D-Watson (Dependente)	
-1 a +1	-1,64+1,64	-1,96+1,96	D Calculado	2,02
73 %	92 %	97 %	Não auto-regressão 90%	
Cálculo			Outliers	
Tipo	Critério	Equação	Linear	Não Linear
Geral	Linear	12	1	3

Exemplo 3: $6(k+1) = 3 \times 4 = 24 < 38$ Grau III

PRESSUPOSTO DA AUSÊNCIA DE ERROS DE MEDIDA

- Os valores das **variáveis independentes** foram medidos **sem erro**, ou seja, não contêm perturbações aleatórias
- **Transgressão**: as estimativas de a e b_i são viesadas, os testes de hipótese e o intervalo de confiança são afetados

* Na prática, essa condição é atendida porque as **variáveis independentes** podem ser medidas de forma **muito mais precisa** do que a dependente

PRESSUPOSTOS PARA O ERRO

O erro aleatório ε pode ser devido a:

- Elementos erráticos ou imprevisíveis do **comportamento humano**
- Variáveis independentes **desconhecidas** ou **não mensuráveis** (porém, individualmente **irrelevantes**)
- **Erros** de medida aleatórios na **variável dependente Y**
- Mas também: **especificação imperfeita** das relações

PRESSUPOSTOS PARA O ERRO

- A variável ε é **aleatória** (ou seja, os erros são aleatórios)
- A **média** da variável ε na população é **zero** (ou seja, a média dos erros é zero)
- A **variância** da variável ε é **constante** (ou seja, a variância dos erros é constante) = **homocedasticidade**
- A variável ε tem **distribuição normal**
- Inexistência de **autocorrelação** dos erros

PRESSUPOSTOS PARA O ERRO

MÉDIA ZERO DO ERRO ALEATÓRIO ε

- Refere-se à equação de regressão **populacional**, que é desconhecida

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

- Assim, existe uma relação linear entre a **média de Y e X**

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- Na equação de regressão **amostral**, a média dos resíduos é zero (efeito dos MQO)

PRESSUPOSTOS PARA O ERRO MÉDIA ZERO DO ERRO ALEATÓRIO ε

Para assegurar esse pressuposto é necessário:

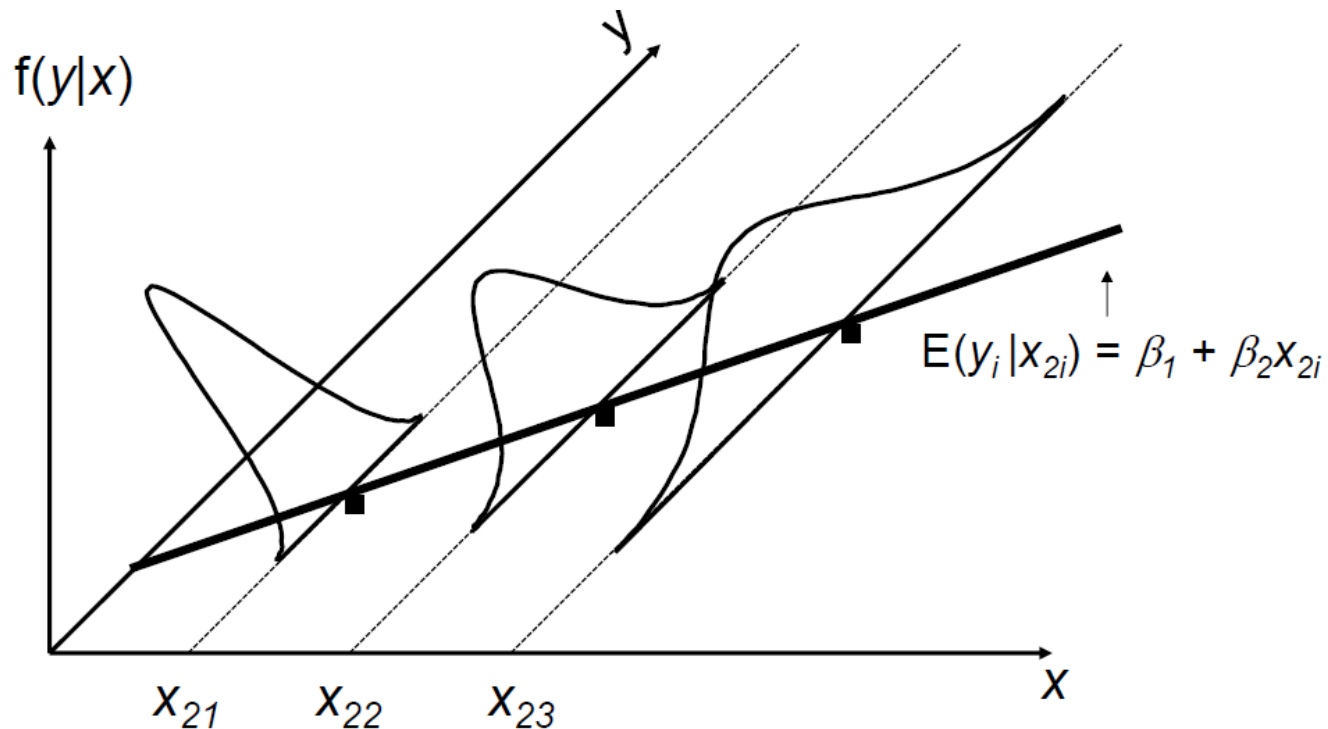
- Incluir no modelo todas as **variáveis independentes importantes** (especificar bem o modelo)
- Não ocorrer **erro de medida** sistemático na **variável dependente**

PRESSUPOSTOS PARA O ERRO HOMOCEDASTICIDADE

- A variância dos erros é constante = **homocedasticidade**
- A variação de Y em torno da função teórica de regressão pode ser descrita por um desvio-padrão dos erros **constante**, ou seja, que não depende dos valores de X

PRESSUPOSTOS PARA O ERRO HOMOCEDASTICIDADE

Caso de heterocedasticidade:



PRESSUPOSTOS PARA O ERRO HOMOCEDASTICIDADE

- **Transgressão** (heterocedasticidade):
 - Erros-padrão dos coeficientes estimados são **tendenciosos**
 - Os intervalos de confiança e os testes t e F ficam prejudicados
 - Na regressão simples:

$$\hat{y}_k \pm t_{n-2; \alpha/2} \cdot \sigma_R \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

constante

PRESSUPOSTOS PARA O ERRO HOMOCEDASTICIDADE

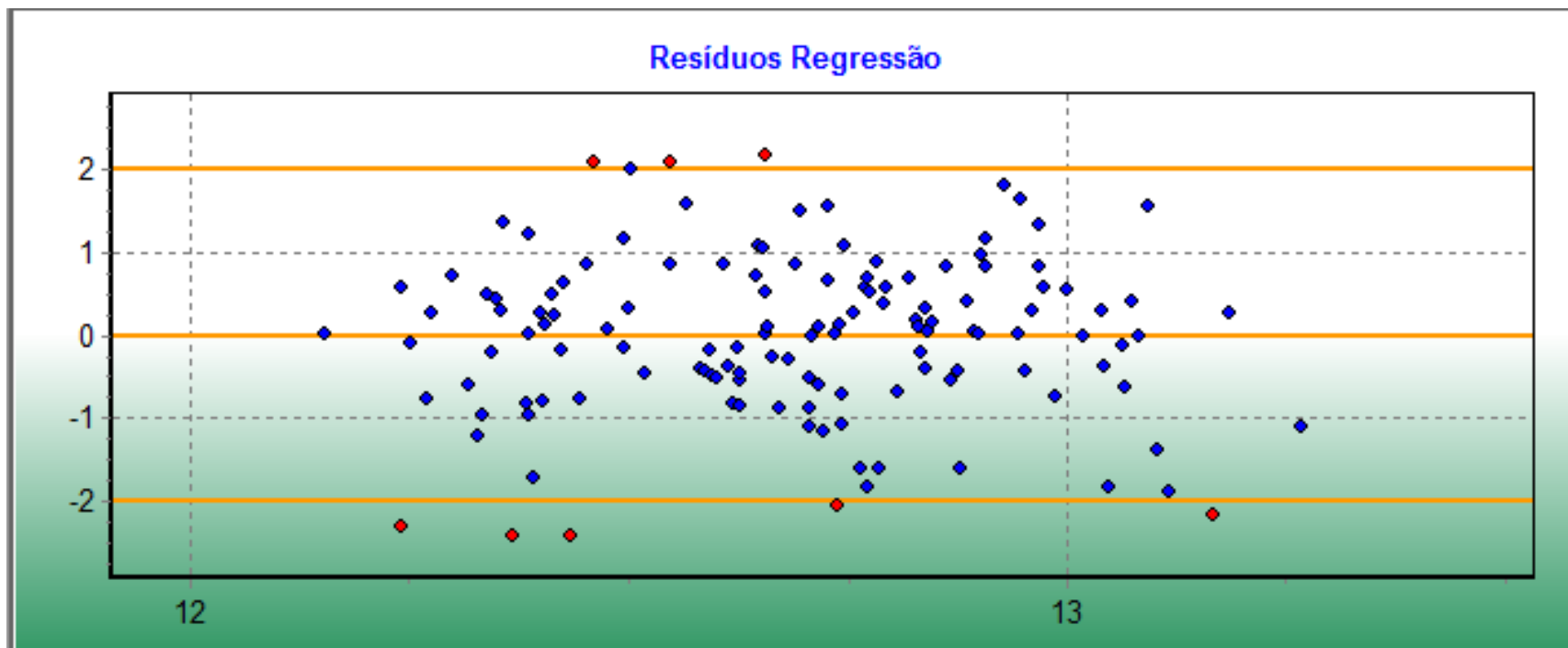
NBR 14.653-2

A.2.1.3 Homocedasticidade

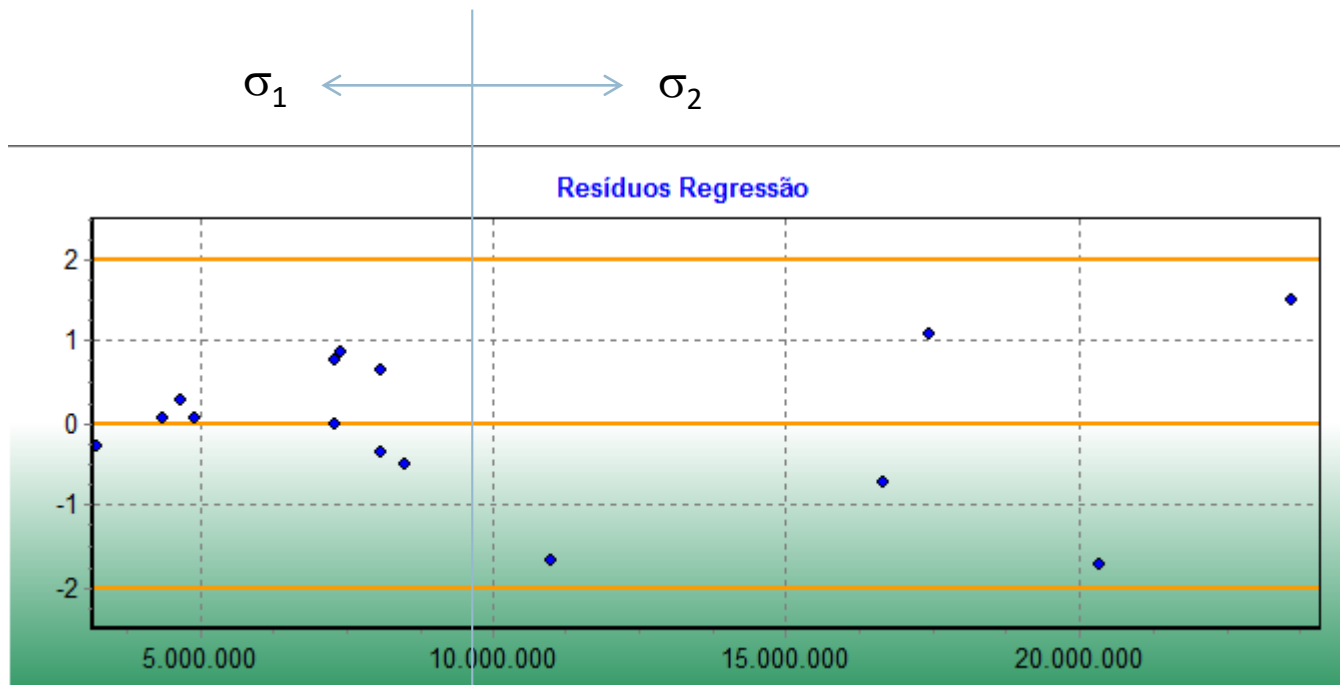
A verificação da homocedasticidade pode ser feita, entre outros, por meio dos seguintes processos:

- a) análise gráfica dos resíduos *versus* valores ajustados, que devem apresentar pontos dispostos aleatoriamente, sem nenhum padrão definido;
- b) pelos testes de Park e de White.

PRESSUPOSTOS PARA O ERRO HOMOCEDASTICIDADE



PRESSUPOSTOS PARA O ERRO HOMOCEDASTICIDADE



Heterocedasticidade
Casas em Irajá

HOMOCEDASTICIDADE

Exemplos 2 e 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

PRESSUPOSTOS PARA O ERRO

NORMALIDADE

- Quando todas as variáveis independentes **importantes** foram incluídas no modelo, é mais provável que ocorram erros pequenos do que erros grandes, sugerindo uma distribuição normal
- Quando falta alguma variável independente importante no modelo (por **má especificação**), os valores do termo aleatório **podem apresentar outra distribuição** que não seja a normal

PRESSUPOSTOS PARA O ERRO NORMALIDADE

- **Transgressão:**
 - Estimativas pontuais continuam não-tendenciosas e consistentes
 - Testes de significância dos parâmetros não são confiáveis
 - Intervalos de confiança não são confiáveis
 - Na regressão simples

$$\hat{y}_k \pm t_{n-2; \alpha/2} \cdot \sigma_R \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Normal ou Student

PRESSUPOSTOS PARA O ERRO NORMALIDADE

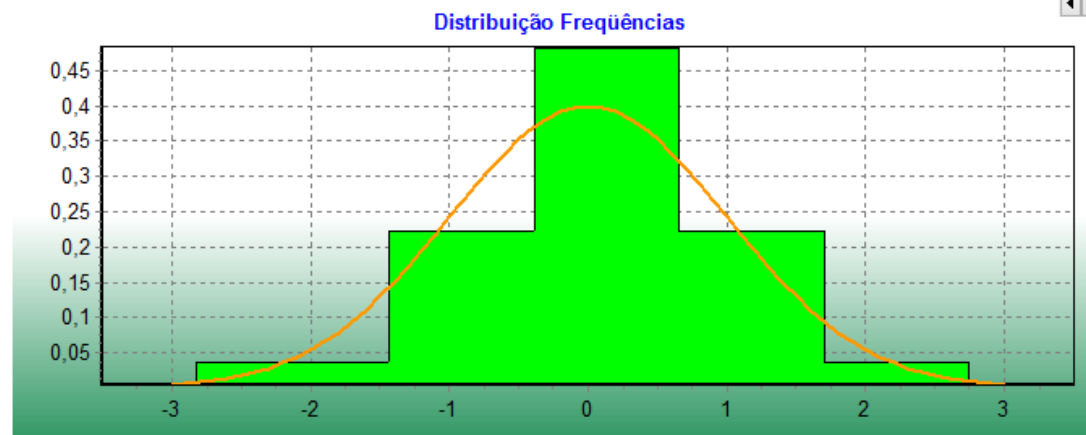
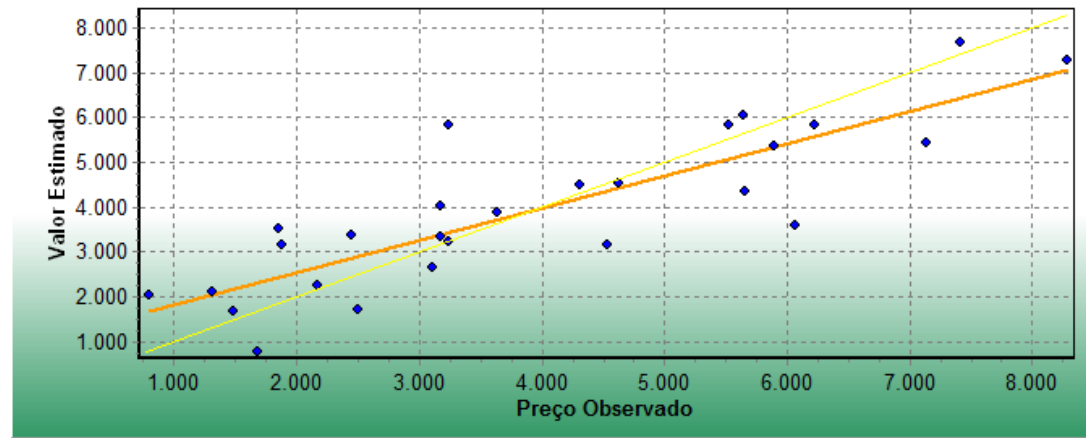
NBR 14.653-2

A.2.1.2 Normalidade

A verificação da normalidade pode ser realizada, entre outras, por uma das seguintes formas:

- a) pelo exame de histograma dos resíduos amostrais padronizados, com o objetivo de verificar se sua forma guarda semelhança com a da curva normal;
- b) pela análise do gráfico de resíduos padronizados *versus* valores ajustados, que deve apresentar pontos dispostos aleatoriamente, com a grande maioria situados no intervalo $[-2; +2]$.
- c) pela comparação da frequência relativa dos resíduos amostrais padronizados nos intervalos de $[-1; +1]$, $[-1,64; +1,64]$ e $[-1,96; +1,96]$, com as probabilidades da distribuição normal padrão nos mesmos intervalos, ou seja, 68 %, 90 % e 95 %;
- d) pelo exame do gráfico dos resíduos ordenados padronizados *versus* quantis da distribuição normal padronizada, que deve se aproximar da bissetriz do primeiro quadrante;
- e) pelos testes de aderência não-paramétricos, como, por exemplo, o qui-quadrado, o de Kolmogorov-Smirnov ajustado por Stephens e o de Jarque-Bera.

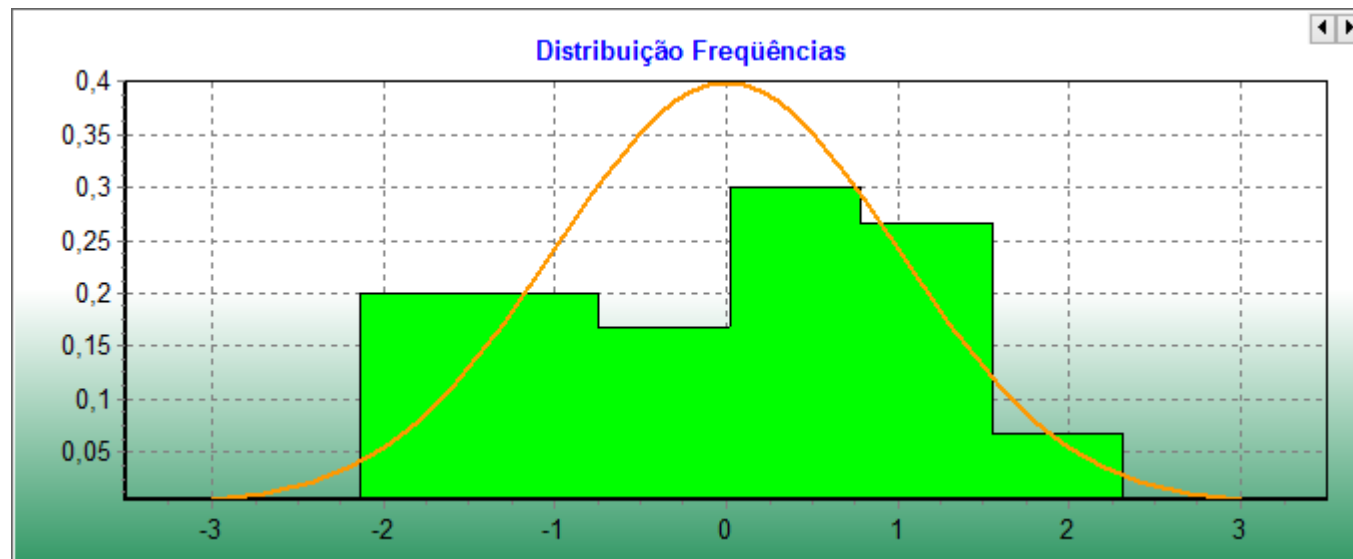
PRESSUPOSTOS PARA O ERRO NORMALIDADE



Valores residuais para terrenos no Centro do Rio de Janeiro

PRESSUPOSTOS PARA O ERRO

NORMALIDADE



Valores para apartamentos em Piratininga

PRESSUPOSTOS PARA O ERRO

NORMALIDADE

Variáveis			Dados	
Total	Consideradas		Total	Considerados
6	4		38	38
Graus de Liberdade			34	
Determinação				Correlação
Linear	Não Linear		Ajustado	Linear
0,861038	0,853178		0,848777	0,927922
Fisher-Snedecor			Desvio Padrão	
F calculado	Significância		Linear	Não Linear
70,22	0,01		0,07	488,47
Normalidade dos Resíduos			D-Watson (Dependente)	
-1 a +1	-1,64+1,64	-1,96+1,96	D Calculado	2,02
73 %	92 %	97 %	Não auto-regressão 90%	
Cálculo			Outliers	
Tipo	Critério	Equação	Linear	Não Linear
Geral	Linear	12	1	3

68%; 90%; 95%

Valores para apartamentos na Vila Pan

NORMALIDADE

Exemplos 2 e 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

PRESSUPOSTOS PARA O ERRO AUTOCORRELAÇÃO

- Não existência de autocorrelação dos erros significa que os termos de erro são **independentes** entre si
- É importante quando existe a inclusão da **variável tempo** no modelo: observação medida em t_1 não deve influir no momento posterior t_2

PRESSUPOSTOS PARA O ERRO AUTOCORRELAÇÃO

- **Transgressão:** os coeficientes permanecem não-viesados, mas não se pode confiar nos testes de significância e nos intervalos de confiança

PRESSUPOSTOS PARA O ERRO AUTOCORRELAÇÃO

NBR 14.653-2

A.2.1.4 Verificação da autocorrelação

O exame da autocorrelação deve ser precedido pelo pré-ordenamento dos elementos amostrais, em relação aos valores ajustados e, se for o caso, às variáveis independentes possivelmente causadoras do problema.

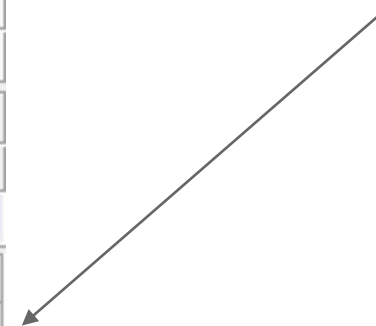
Sua verificação pode ser feita, entre outros procedimentos, pela análise do gráfico dos resíduos cotejados com os valores ajustados, que deve apresentar pontos dispersos aleatoriamente, sem nenhum padrão definido.

PRESSUPOSTOS PARA O ERRO AUTOCORRELAÇÃO

Redes Neurais Artificiais	
Erro Quadrático	
Erro Quadrático Validação	
Interações	
Tolerância	
Taxa de Aprendizagem	
Coeficientes	
Correlação	1 - 0,8014482 / 0,7766129
Determinação	1 - 0,6423192 / 0,6031275
R2 Ajustado	1 - 0,6226045 / 0,5812527
Testes de Hipóteses	
F Calculado	32,58
Significância do Modelo	0,01
Durbin Watson	Não auto-regressão 90%
d Calculado	2,22
Durbin Watson	TEMPO
Normalidade dos Resíduos	
-1 e +1 desvios padrões	72%
-1,64 e +1,64 desvios padrões	89%
-1,96 e +1,96 desvios padrões	93%

PRESSUPOSTOS PARA O ERRO AUTOCORRELAÇÃO

Variáveis			Dados	
Total	Consideradas		Total	Considerados
6	4		38	38
Graus de Liberdade			34	
Determinação				Correlação
Linear	Não Linear		Ajustado	Linear
0,861038	0,853178		0,848777	0,927922
Fisher-Snedecor			Desvio Padrão	
F calculado	Significância		Linear	Não Linear
70,22	0,01		0,07	488,47
Normalidade dos Resíduos			D-Watson (Dependente)	
-1 a +1	-1,64+1,64	-1,96+1,96	D Calculado	2,02
73 %	92 %	97 %	Não auto-regressão 90%	
Cálculo			Outliers	
Tipo	Critério	Equação	Linear	Não Linear
Geral	Linear	12	1	3

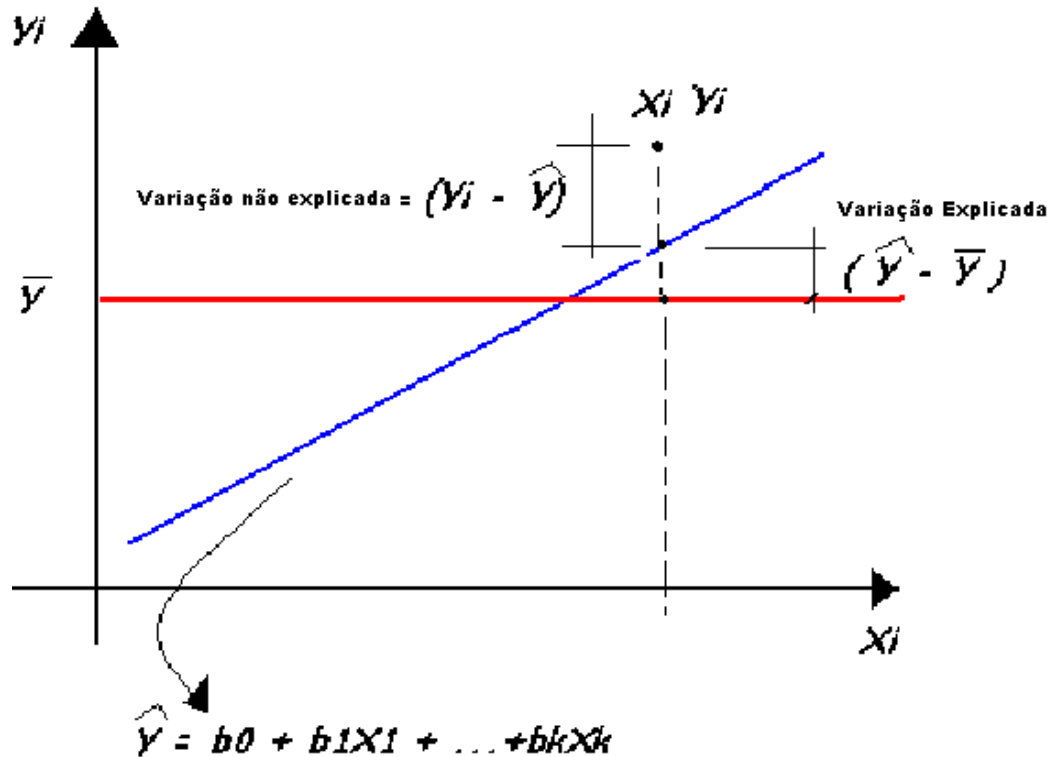


AUTOCORRELAÇÃO

Exemplos 2 e 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

PRESSUPOSTO DA NÃO MULTICOLINEARIDADE



Coeficiente de
determinação

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

$$R = \sqrt{\frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}}$$

Coeficiente de
correlação

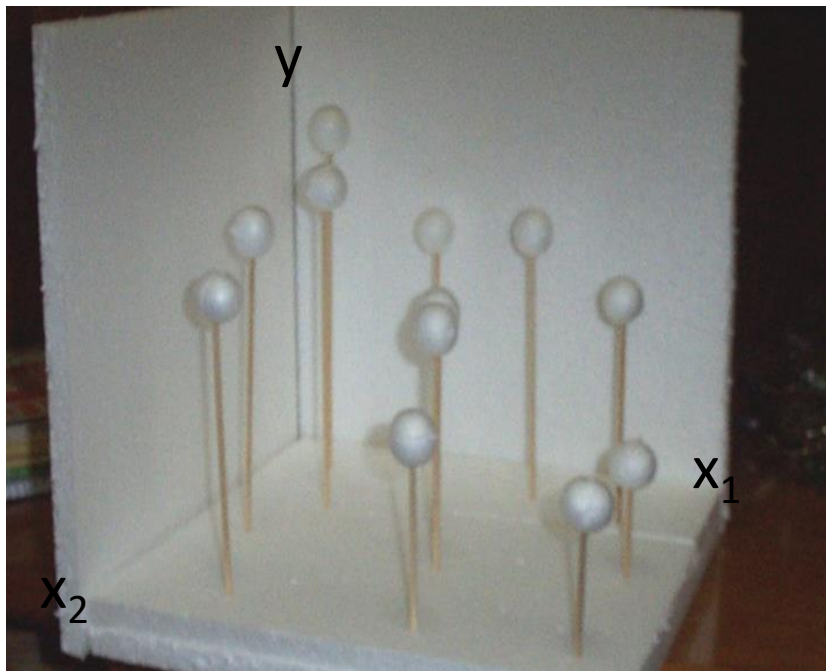
PRESSUPOSTO DA NÃO MULTICOLINEARIDADE

- Não deve existir nenhuma relação exata entre quaisquer variáveis independentes
- Quanto maior for a correlação entre as variáveis independentes, menos informação haverá para a estimativa dos coeficientes das variáveis independentes correlacionadas

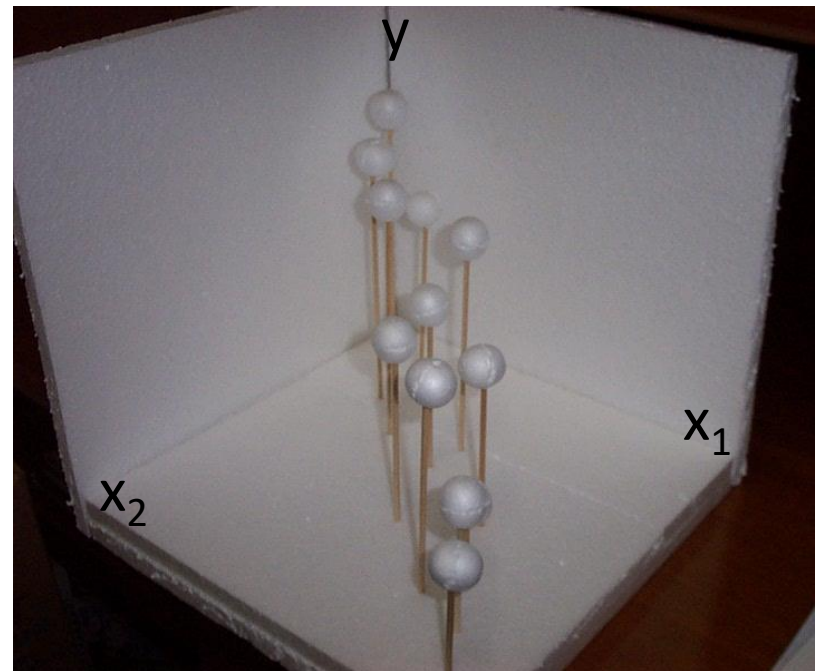
PRESSUPOSTO DA NÃO MULTICOLINEARIDADE

- Uma forte correlação (acima de 80%) vai limitar o uso da equação a apenas avaliar imóveis que possuam a mesma correlação inferida

PRESSUPOSTO DA NÃO MULTICOLINEARIDADE



Não colinearidade entre x_1 e x_2



Colinearidade entre x_1 e x_2

PRESSUPOSTO DA NÃO MULTICOLINEARIDADE

NBR 14.653-2

A.2.1.5 Colinearidade ou multicolinearidade

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.

PRESSUPOSTO DA NÃO MULTICOLINEARIDADE

	For...	Variável	TEMPO	ÁREA	PD PREDIO	EST APTO	EVENTO	RENDIA MD	Elevador	VALOR TOTAL
1	x^2	TEMPO		22	3	-13	-6	10	7	16
2	$1/x^{1/2}$	ÁREA	22		0	-1	-2	-7	45	-45
3	$\ln(x)$	PD PREDIO	3	0		-23	1	-17	30	27
4	$1/x^2$	EST APTO	-13	-1	-23		-19	-18	-15	-45
5	x	EVENTO	-6	-2	1	-19		-7	12	19
6	x^2	RENDIA MD	10	-7	-17	-18	-7		-12	35
7	x	Elevador	7	45	30	-15	12	-12		13
8	$\ln(x)$	VALOR TOTAL	16	-45	27	-45	19	35	13	



Tabela de Correlações (valores em percentua

← Variável	Forma Lin	Nº de que	Andar	Área (m²)	Valor Unit
Nº de quartos	$\ln(x)$		9	94	32
Andar	$1/x^2$	-2		18	30
Área (m²)	$1/x$	-99	2		59
Valor Unit (R\$/m²)	$\ln(y)$	-88	-9	91	

PRESSUPOSTO DA NÃO MULTICOLINEARIDADE

- **Correlação simples (isolada) de x com y**
 r_{xy}
- **Correlação parcial de x com y , dado z**
 - Exemplo do modelo com três variáveis
 - Exclui a influência de Z sobre X e y
 - Correlação dos resíduos de XZ com os resíduos de YZ
 - A influência de Z sobre X e y é retirada

$$r_{xy/z} = \frac{r_{xy} - r_{yz}r_{xz}}{\sqrt{(1-r_{xz}^2)} \cdot \sqrt{(1-r_{yz}^2)}}$$

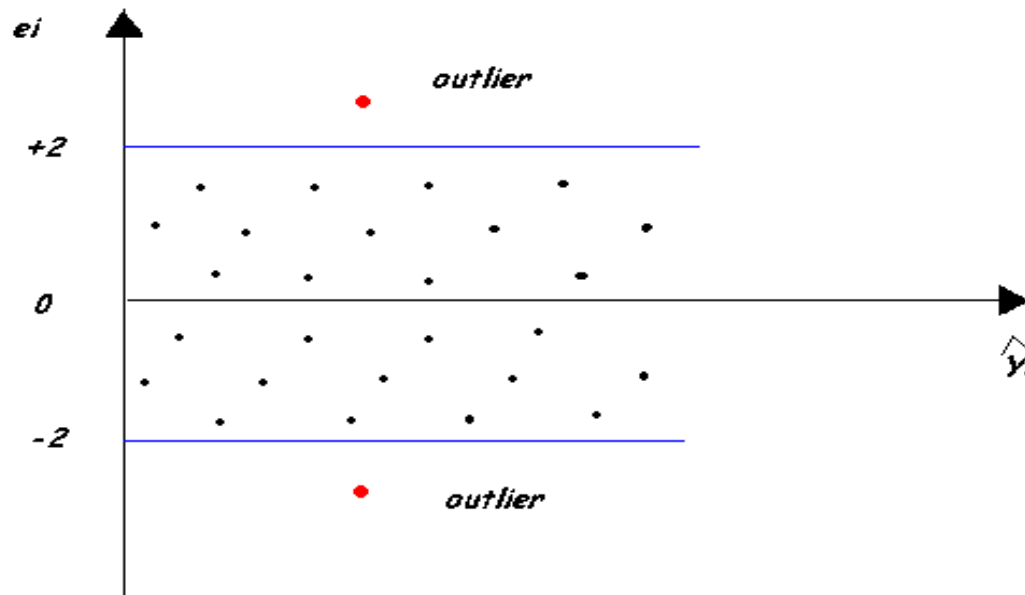
MULTICOLINEARIDADE

Exemplo 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

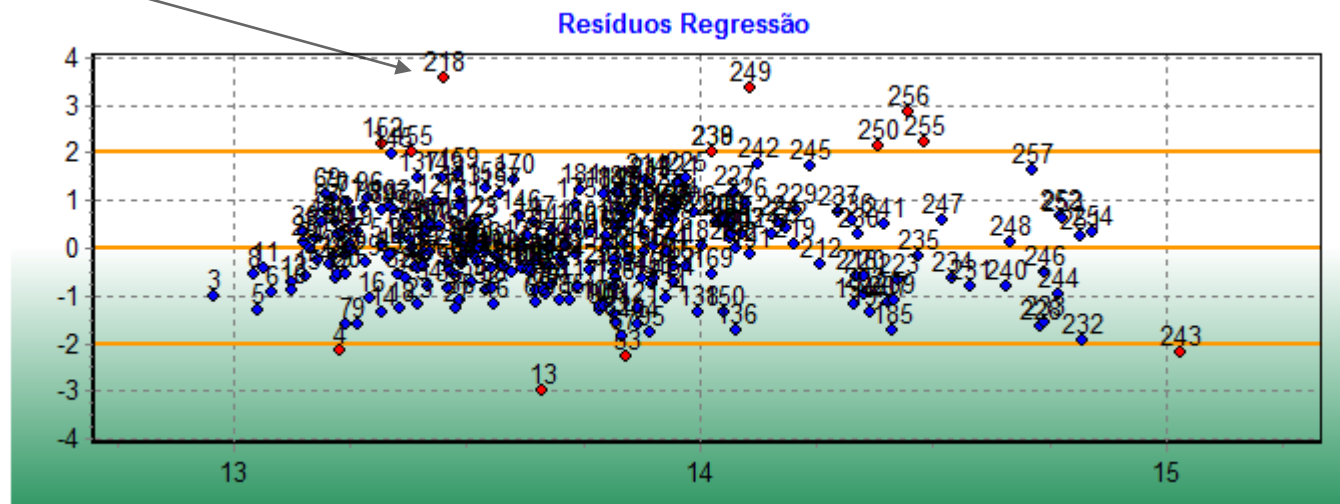
OUTLIER

- Ponto atípico, identificado como estranho à massa de dados, que, ao ser retirado, melhora a qualidade de ajustamento do modelo analisado



OUTLIER

- 255 dados de mercado
- Além de $\pm 2\sigma$ na Curva Normal: 4,56% $\approx$ 5%
- Cerca de 12 casos estariam normalmente nessa situação (x 13 casos, efetivamente)



Apartamentos na Zona Sul do Rio de Janeiro

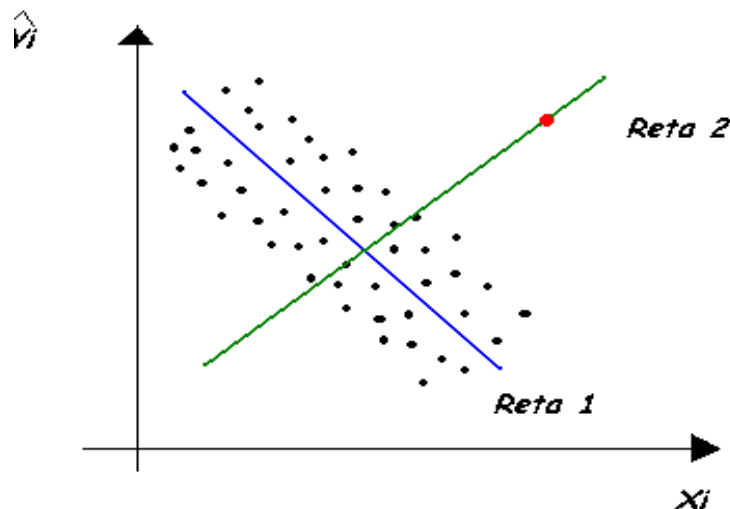
OUTLIER

- Examinar caso a caso
- Elemento 218 é sala comercial (*Quartier Ipanema*)

Da...	Endereço	Complemento	Bairro	Informante	Telefone	IDADE	* TEMPO	ÁREA	VA
211	RUA BENTO LISBOA,120 - BL.3 APTO 802	2014.10.1.D.604	CATETE	KASANOVA	2303-1703	5	178	90,00	
212	AVENIDA ATAULFO DE PAIVA,765 - BL.- APTO 304	2014.9.1.L.282	LEBLON	ANDRE	986.940.517	44	177	121,00	
213	RUA BARÃO DE ICARAI,29 - BL.- APTO 704	2014.10.1.L.584	FLAMENGO	PAULO AN...	99604-8716	35	178	104,00	
214	RUA NINA RODRIGUES,57 - BL.- APTO 102	2014.10.1.D.290	JARDIM B...	MARCIA	2295-1039	76	178	107,00	
215	RUA BENTO LISBOA,120 - BL.3 APTO 802	2014.9.1.D.406	CATETE	KASANOVA	2303-1703	5	177	90,00	
216	RUA PINHEIRO MACHADO,22 - BL.2 APTO 505	2014.9.1.L.654	LARANJEIR...	PAULA	96802-3033	1	177	125,00	
217	RUA SANTA CLARA,357 - BL.- APTO 301	2014.10.1.D.559	COPACAB...	FRANQUIA...	3546-2006	3	178	110,00	
218	RUA VISCONDE DE PIRAJA,414 - BL.- APTO 0	2014.10.1.D.710	IPANEMA	FABIO	99646-2467	0	178	41,50	
219	RUA PINHEIRO GUIMARAES,67 - BL.- APTO 806	2014.11.1.D.518	BOTAFOGO	LEONARDO	96423-3156	1	179	103,00	
220	RUA LEITE REAL, 135/AP. 103 BL. 2	2014.07.1.L.88	LARANJEIR...	CLAUDIA	99966-1424	25	175	122,00	
221	RUA PINHEIRO GUIMARÃES,67 - BL.- APTO 806	2014.9.1.L.304	BOTAFOGO	LEONARD...	964233156	1	177	96,00	
222	RUA PROFESSOR ALFREDO GOMES,7 - BL.- APTO 101	2014.11.1.D.608	BOTAFOGO	ROQUE	0	10	179	109,00	

PONTO INFLUENCIANTE

- Ponto atípico que, quando retirado da amostra, altera significativamente os parâmetros estimados ou a estrutura linear do modelo



PONTO INFLUENCIANTE

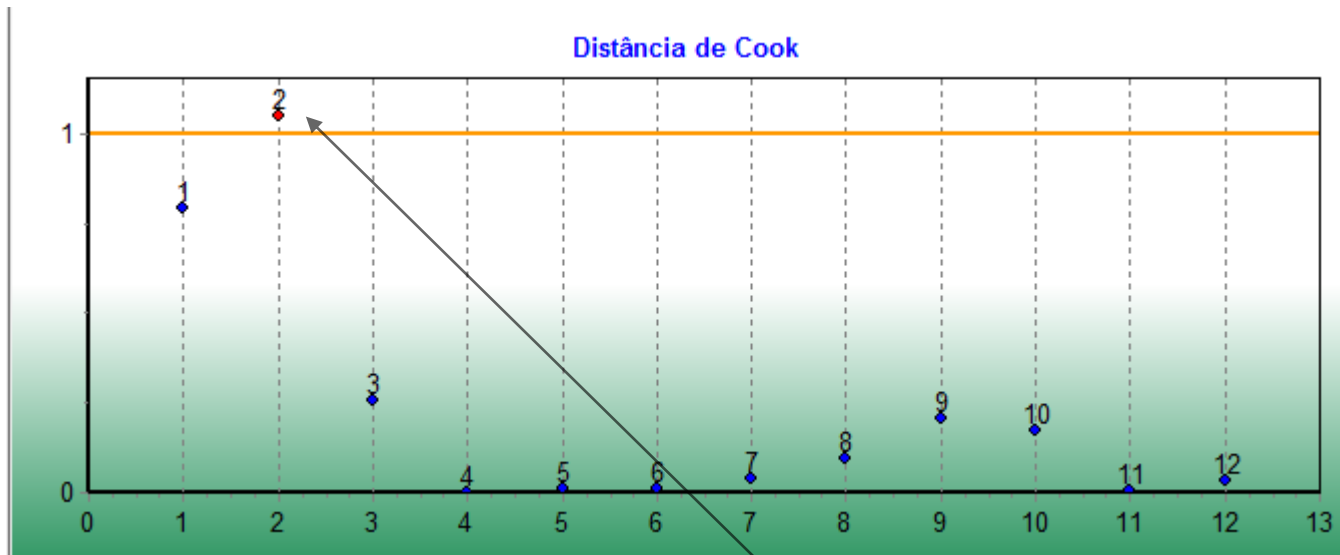
- Distância de Cook (testando o ponto k)

$$D_i = \frac{n \cdot \sum_{i=1}^n (\hat{y}_i - \hat{y}_{i(k)})^2}{p \cdot \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

$D_i > 1 \rightarrow$ ponto influenciante

PONTO INFLUENCIANTE

- Distância de Cook



Casas no Jardim Guanabara

ponto influenciante

PONTOS INFLUENCIANTES E *OUTLIERS*

Exemplos 2 e 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

TESTE DE HIPÓTESE DO MODELO

- **Hipótese Nula** – nenhuma variável escolhida é importante para explicar a variabilidade dos preços

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

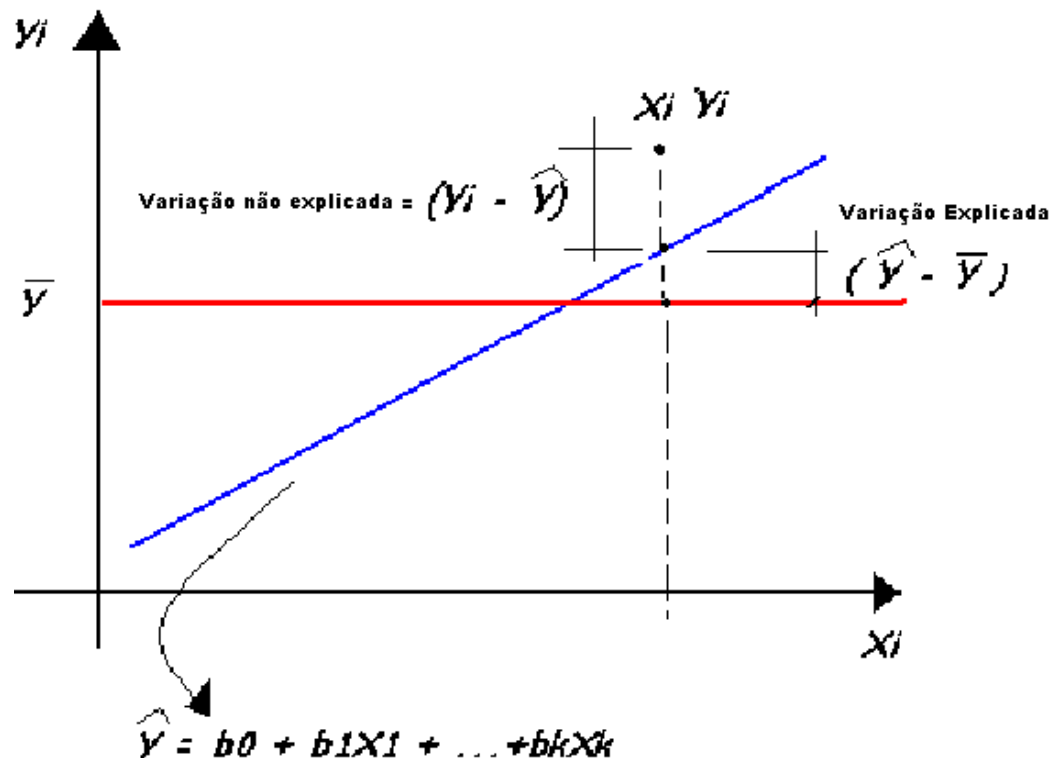
- **Hipótese Alternativa** – pelo menos uma variável escolhida é importante para explicar a variabilidade dos preços

$$H_1 : \beta_j (j = 1 \dots k) \neq 0$$

TESTE DE HIPÓTESE DO MODELO

Variação explicada x variação total

Decomposição da variação



TESTE DE HIPÓTESE DO MODELO

Análise de variância

- Tabela ANOVA

Fonte Variação	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio
Regressão (Explicada)	$SQR = \sum (\hat{Y}_i - \bar{Y})^2$	k	$MQR = SQR/k$
Erro (Não Explicada)	$SQE = \sum (Y_i - \hat{Y})^2$	n - (k + 1)	$EMQ = SQE/(n-k-1)$
Total	$SQT = \sum (Y_i - \bar{Y})^2$	n - 1	

TESTE DE HIPÓTESE DO MODELO

Teste F

- Estatística de Snedecor e Fisher
- Calcula-se a razão entre as variâncias explicada e não explicada e compara-se com F tabelado por Fisher

$$F_c = \frac{MQR}{EMQ} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \hat{Y}_i)^2} \cdot \frac{(n - k - 1)}{k} > F_{tab}$$

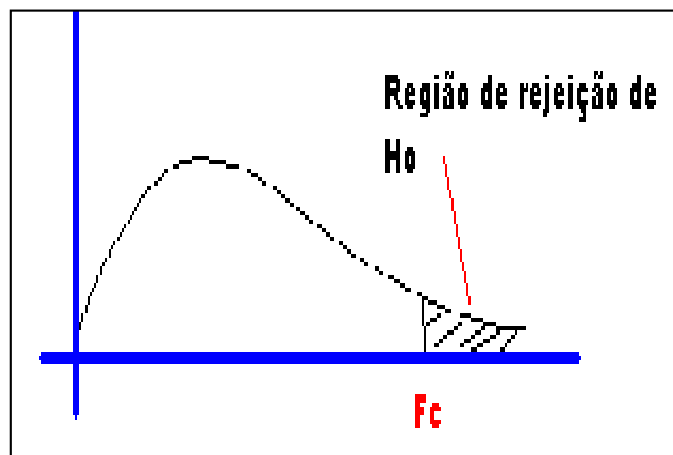
TESTE DE HIPÓTESE DO MODELO

Teste F

$$F_c = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \hat{Y}_i)^2} \cdot \frac{(n - k - 1)}{k} > F_{\alpha; k; n - k - 1}$$

Graus de liberdade do denominador

Graus de liberdade do numerador



TESTE DE HIPÓTESE DO MODELO

Teste F

$$F_c = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \hat{Y}_i)^2} \cdot \frac{(n - k - 1)}{k} > F_{\alpha; k; n - k - 1}$$

Graus de liberdade do numerador

Graus de liberdade do denominador

Rejeita-se a **hipótese nula** (H_0) e aceita-se a **hipótese alternativa** (H_1) de que existe pelo menos uma variável importante para explicar a variabilidade dos preços

TESTE DE HIPÓTESE DO MODELO

NBR 14.653-2

Item	Descrição	Grau		
		III	II	I
6	Nível de significância máximo admitido para a rejeição da hipótese nula do modelo através do teste F de Snedecor	1 %	2 %	5 %

COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO

R^2

- Poder de explicação do modelo
- Calcula a razão entre a variação explicada e a variação total

$$R^2 = \frac{SQR}{SQT} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$$

COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO AJUSTADO

R^2_a

- Compensa a entrada de uma nova variável no modelo pela perda de um grau de liberdade no denominador

$$R^2_a = 1 - \frac{n-1}{n-k} (1 - R^2)$$

COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO AJUSTADO

14.653-2

A.4 Poder de explicação

Em uma mesma amostra, a explicação do modelo pode ser aferida pelo seu coeficiente de determinação. Devido ao fato de que este coeficiente sempre cresce com o aumento do número de variáveis independentes e não leva em conta o número de graus de liberdade perdidos a cada parâmetro estimado, deve-se considerar o coeficiente de determinação ajustado.

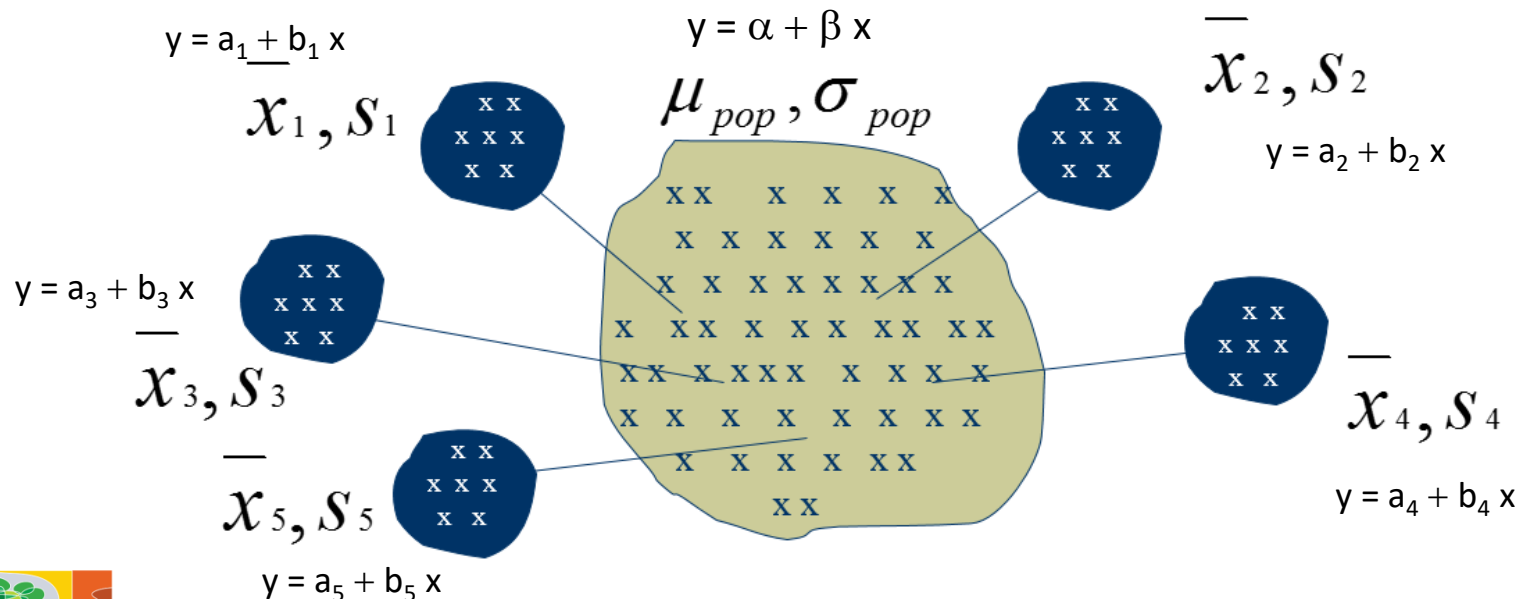
TESTE DE HIPÓTESE DO MODELO

Exemplo 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

TESTE DE HIPÓTESE DO REGRESSOR

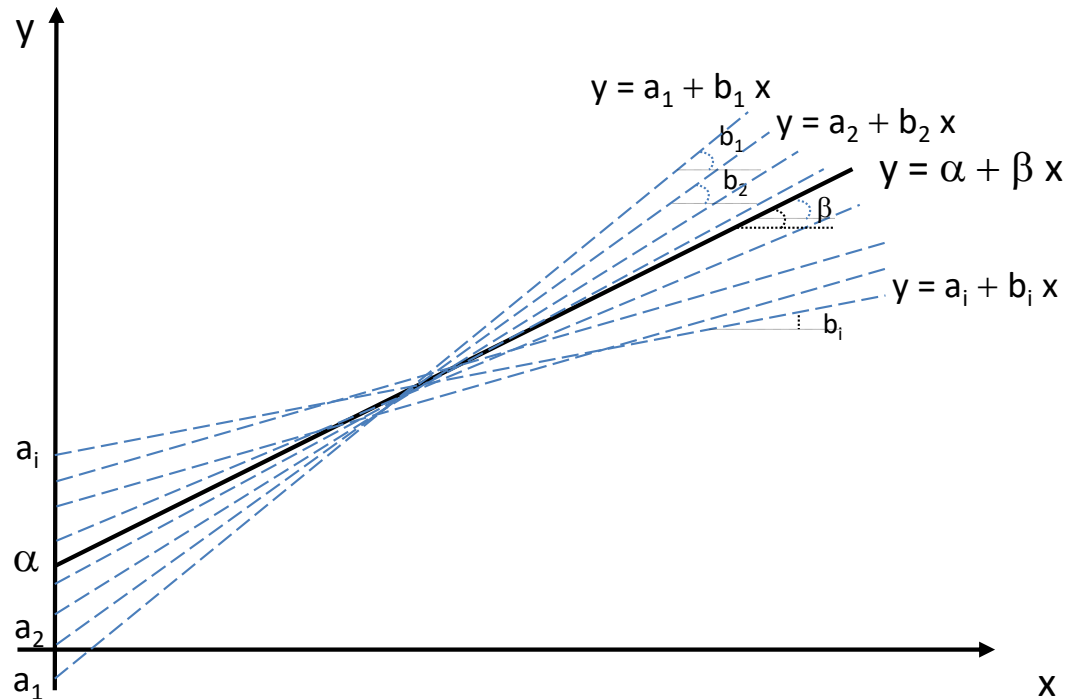
- Inferir o valor dos **parâmetros** da reta de regressão da **população** ($\alpha, \beta_1, \dots, \beta_n$), a partir de **estatísticas** ($a, b_1, \dots, b_n$) da reta de regressão **amostral**



TESTE DE HIPÓTESE DO REGRESSOR

Reta de regressão da população x

Retas de regressão amostrais (regressão simples)



TESTE DE HIPÓTESE DO REGRESSOR

- **Hipótese nula** – a variável não é importante para explicar a variabilidade dos preços

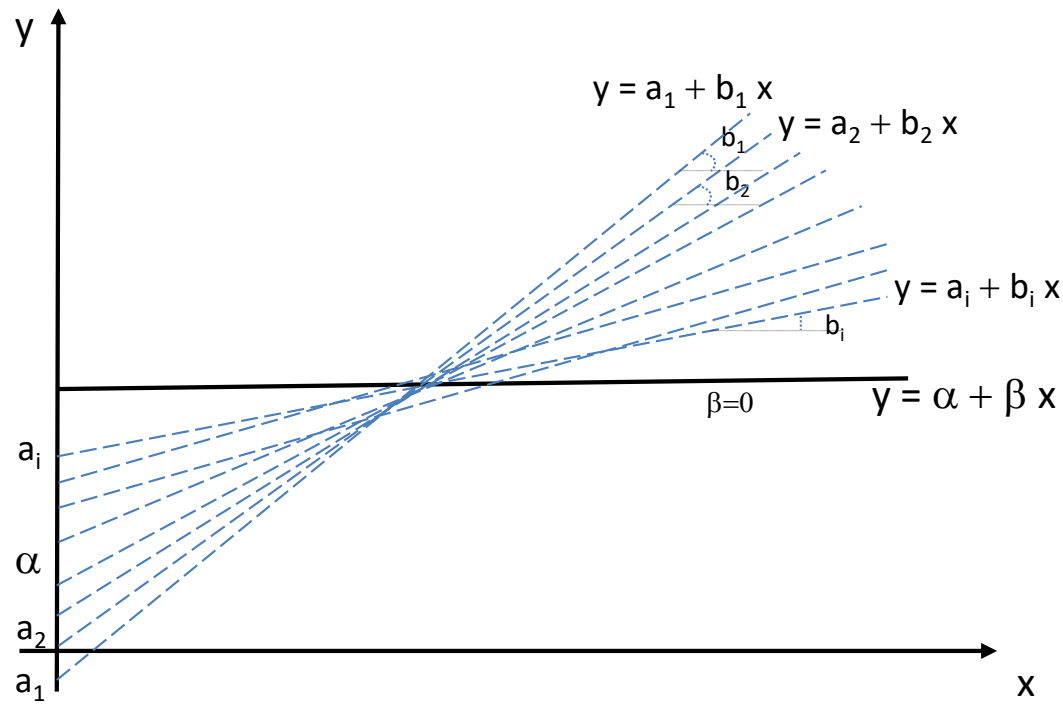
$$\beta_j = 0$$

- **Hipótese alternativa** – contrária à hipótese nula

$$\beta_j \neq 0$$

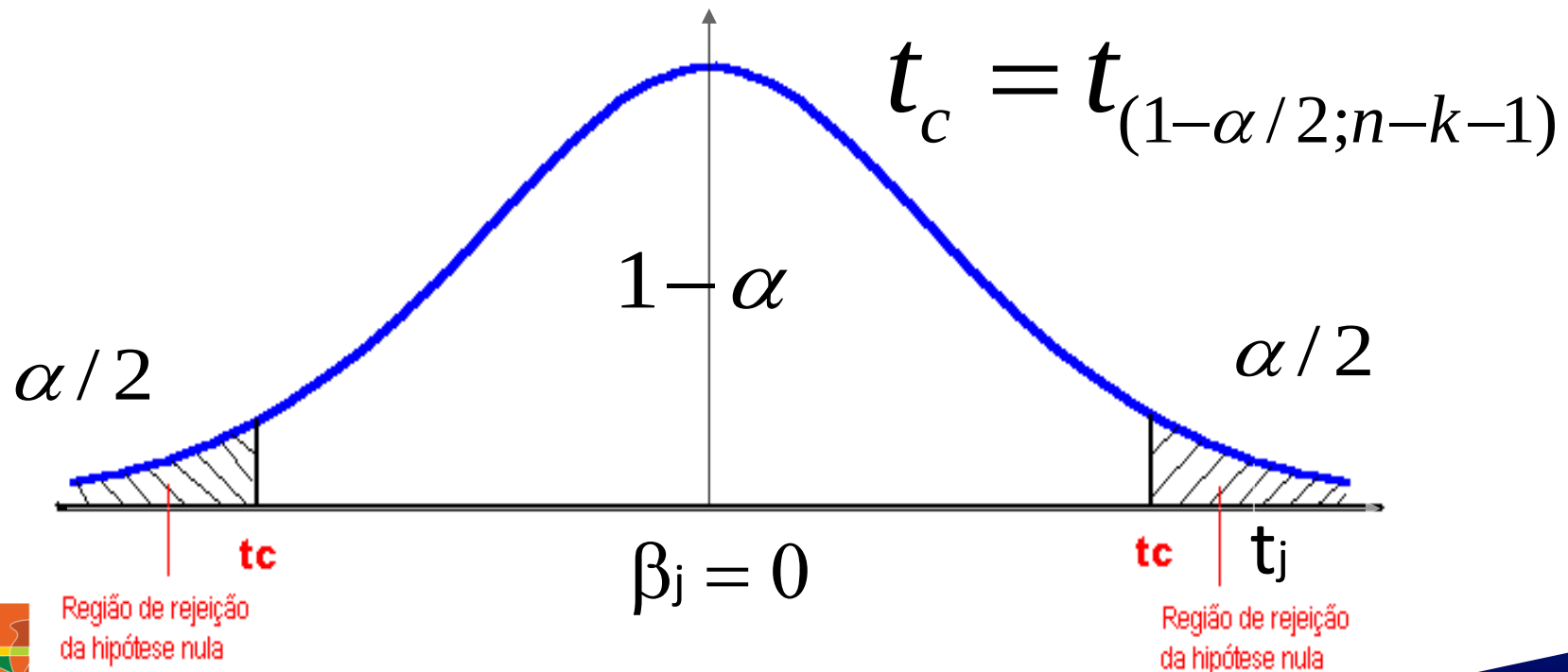
TESTE DE HIPÓTESE DO REGRESSOR

A hipótese nula: $\beta = 0$ (regressão simples)



TESTE DE HIPÓTESE DO REGRESSOR

- Significância individual do regressor
- Nível de significância α (bilateral) = valor de prova
- Distribuição t de Student



TESTE DE HIPÓTESE DO REGRESSOR

$$t_j = \left| \frac{b_j - \beta_j}{s(b_j)} \right| = \left| \frac{b_j}{s(b_j)} \right|$$

b_j é o estimador de β_j

$s(b_j)$ é o desvio padrão estimado do parâmetro

TESTE DE HIPÓTESE DO REGRESSOR

Para a regressão simples

$$t_j = \left| \frac{b - \beta}{s(b)} \right| = \left| \frac{b}{s(b)} \right|$$

$$S(b) = s_e \cdot \frac{1}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}}$$

TESTE DE HIPÓTESE DE CADA REGRESSOR

$$s(b_j) = \sqrt{s_e^2 \cdot a_{ii}}$$

s_e^2 é a variância residual

a_{ii} é o elemento correspondente à variável X_i na diagonal principal da matriz $(X'X)^{-1}$ (matriz de variâncias e covariâncias)

$s(b_j)$ é o desvio padrão estimado do parâmetro

TESTE DE HIPÓTESE DE CADA REGRESSOR

$$t_j > t_{(1-\alpha/2; n-k-1)}$$

Conclui-se que a variável é importante para explicar a variabilidade dos preços e rejeita-se H_0

TESTES DE HIPÓTESE DOS REGRESSORES

NBR 14.653-2

Item	Item	Grau III	Grau II	Grau I
5	Nível de significância α (somatório do valor das duas caudas) máximo para a rejeição da hipótese nula de cada regressor (teste bicaudal)	10 %	20 %	30 %

TESTES DE HIPÓTESE DOS REGRESSORES

Exemplo 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana

INTERVALO DE CONFIANÇA PARA A MÉDIA

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

- **estimativa pontual:** substituir na equação as características do bem avaliando ***h*** ($X_{h1}, X_{h2}, \dots, X_{hk}$)

$$\hat{Y}_h = b_0 + b_1 X_{h1} + b_2 X_{h2} + \dots + b_k X_{hk}$$

INTERVALOS

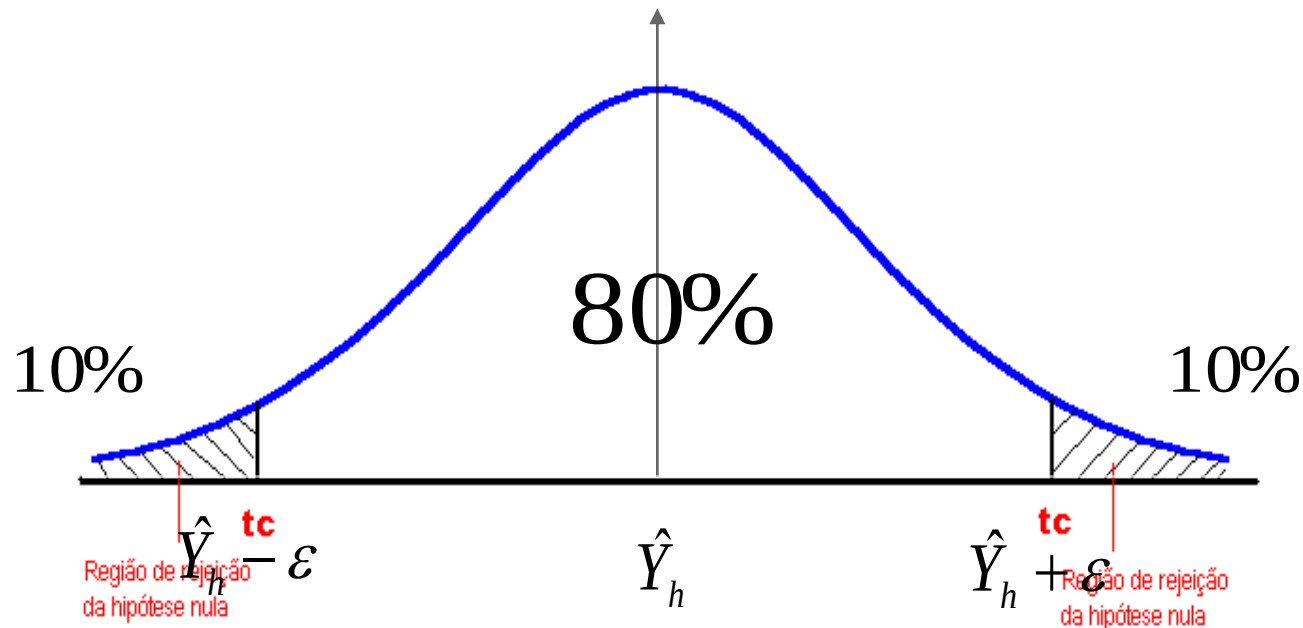
- **Intervalo de confiança** (para a média)
 - Intervalo para a **média** das observações
 - Média: valor mais provável
 - Identificação de um intervalo para o **valor de mercado**, com um dado nível de confiança (80%)
- **Intervalo de predição** (para uma observação singular)
 - Intervalo para **observações individuais**
 - Identificação de **preços** possíveis, com um dado nível de confiança (80%)

INTERVALOS

- **Intervalo de confiança** (para a média)
 - Considera a aleatoriedade do valor ajustado na regressão **amostral** (média) em relação ao valor ajustado na regressão da **população** (média)
- **Intervalo de predição** (para uma observação singular)
 - Considera a aleatoriedade do valor ajustado na regressão **amostral** (média) em relação ao valor ajustado na regressão da **população** (média)
 - Considera a aleatoriedade de uma **observação singular** em relação à **média** das observações na regressão populacional

INTERVALO DE CONFIANÇA PARA A MÉDIA

- Intervalo de confiança ao nível de 80%
- Utilizado no grau de precisão da estimativa



INTERVALO DE CONFIANÇA PARA A MÉDIA

- Intervalo de confiança para $(n-k-1)$ graus de liberdade a um nível de confiança $(1-\alpha)$ de 80%

$$I = \hat{Y}_h \pm t_{(1-\alpha/2; n-k-1)} \cdot s(\hat{Y}_h)$$

$$s(\hat{Y}_h) = s_e \cdot \sqrt{\frac{1}{n} + (X_h - \bar{X})'(X'X)^{-1}(X_h - \bar{X})}$$

$$s_e = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-k-1}}$$

INTERVALO DE CONFIANÇA PARA A MÉDIA

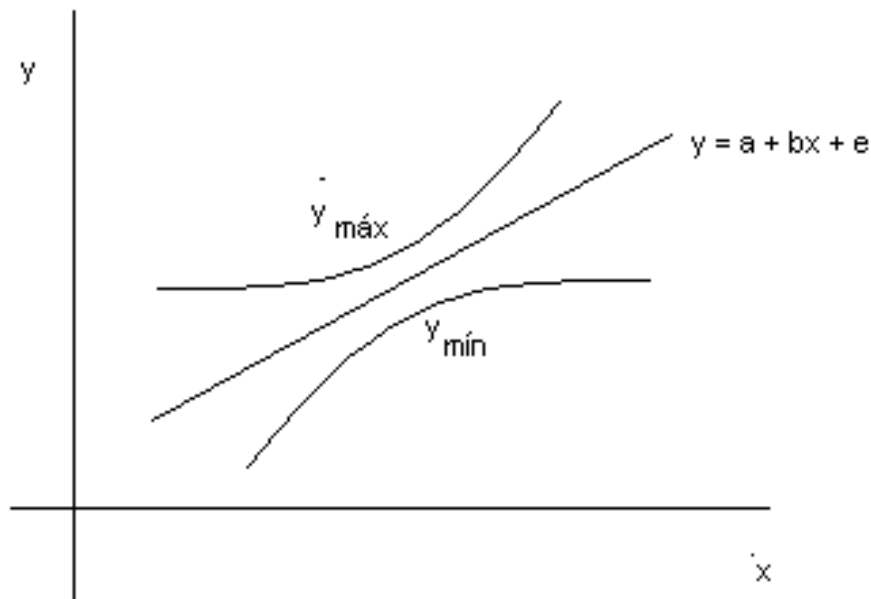
- Intervalo de confiança para a **regressão simples**

$$I = \hat{y}_h \pm t_{(1-\alpha/2; n-2)} \cdot s_e \cdot \sqrt{\left(\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)}$$

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}}$$

INTERVALO DE CONFIANÇA PARA A MÉDIA

- Quando as características do imóvel a avaliar estão próximas às médias das características dos dados de mercado, a amplitude do intervalo é menor



INTERVALO DE CONFIANÇA PARA A MÉDIA

- Grau de precisão da estimativa

NBR 14.653-2

Descrição	Grau		
	III	II	I
Amplitude do intervalo de confiança de 80% em torno do valor central da estimativa	$\leq 30\%$	$\leq 40\%$	$\leq 50\%$
Nota: Observar 9.1 a 9.3 desta Norma.			

INTERVALO DE PREDIÇÃO PARA UMA OBSERVAÇÃO SINGULAR

- Intervalo de predição para a **regressão simples**

$$I = \hat{y}_h \pm t_{1-\alpha/2; (n-2)} \cdot s_e \cdot \sqrt{\left(1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}$$

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}}$$

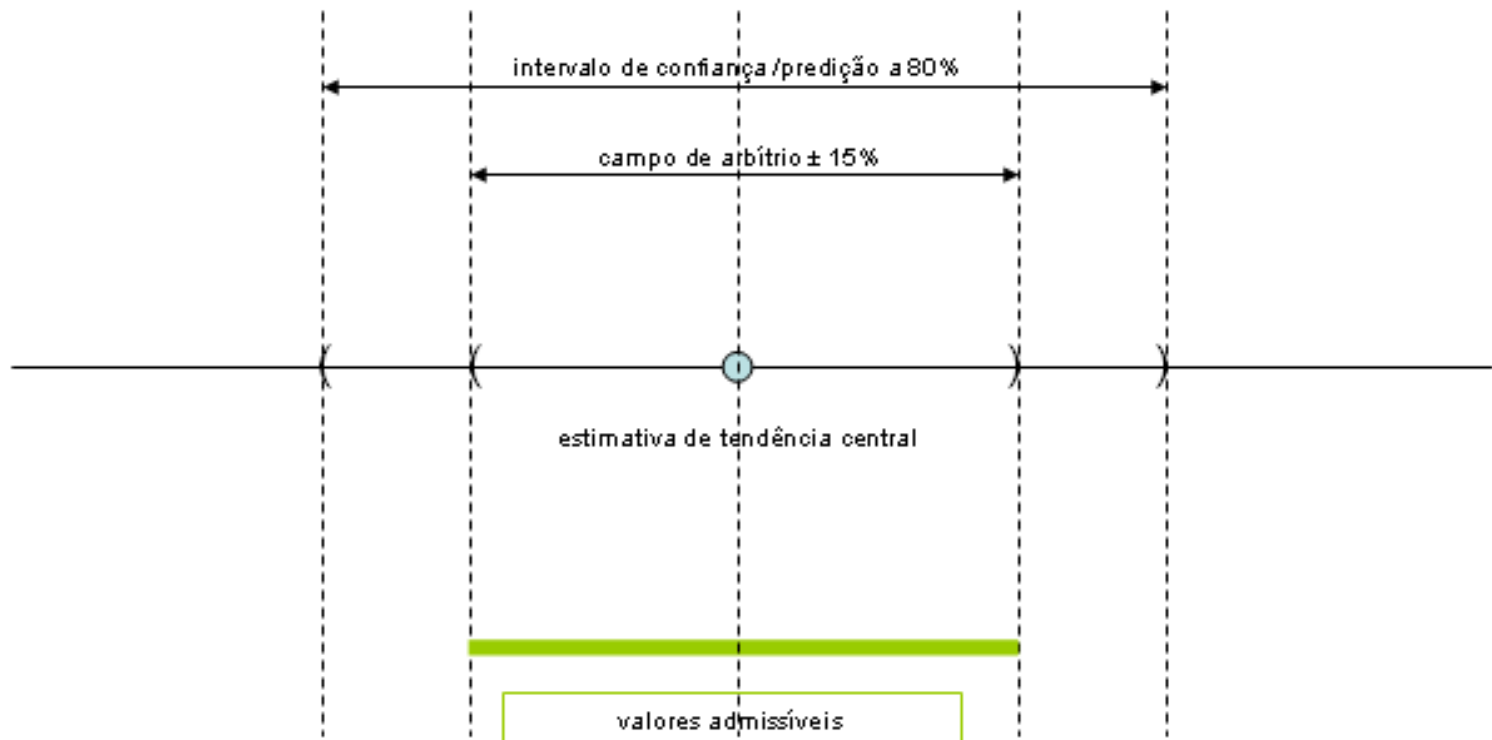
CAMPO DE ARBÍTRIO

NBR 14.653-2

- O **campo de arbítrio** do avaliador não coincide com as fronteiras do **intervalo de confiança**
- Existe para contemplar características do imóvel avaliando que **não foram consideradas no modelo**
- Corresponde a $\pm 15\%$ em torno da **estimativa pontual** adotada
- Se a estimativa pontual não for adotada o avaliador deverá **justificar** sua escolha

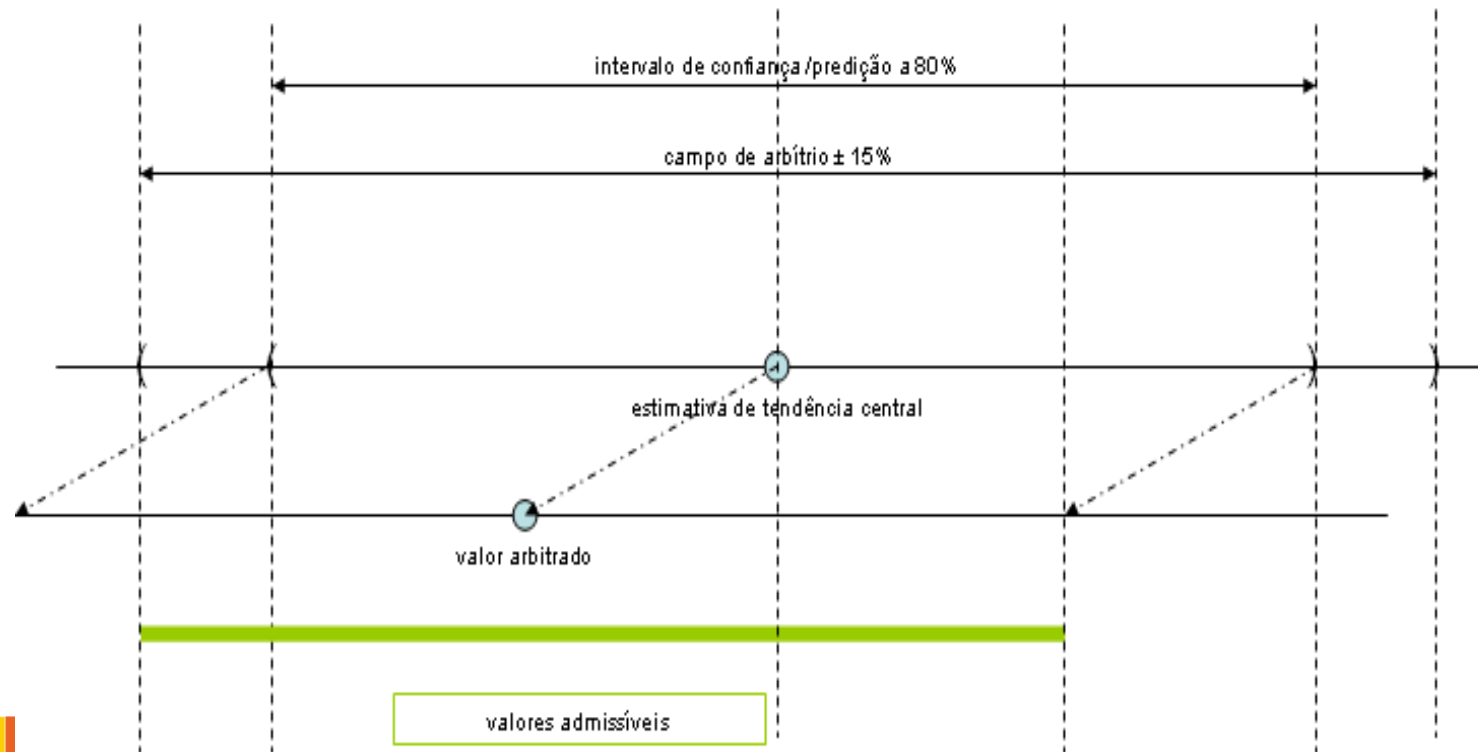
INTERVALO DE VALORES ADMISSÍVEIS

NBR 14.653-2



INTERVALO DE VALORES ADMISSÍVEIS

NBR 14.653-2



INTERVALO DE CONFIANÇA INTERVALO DE PREDIÇÃO CAMPO DE ARBÍTRIO INTERVALO DE VALORES ADMISSÍVEIS

Exemplo 3

Apartamentos de 1, 2, 3 e 4 quartos na
Vila Pan-Americana