

ANALISE ESTATÍSTICA NA REGRESSÃO

CUIDADOS



REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

O modelo de regressão linear simples (RLS) ou múltipla (RLM), aplicado a avaliação de bens imobiliários, tem por objetivo estimar os estimadores do modelo estatístico inferencial e testar hipóteses sobre eles a fim de comprovar se realmente representam os parâmetros da população examinada. Esses estimadores em regra, são os coeficientes, a média e o desvio padrão que após testados podem representar os parâmetros da população ou não. Em caso afirmativo devem ainda atender a algumas premissas quanto ao comportamento da variável em relação à colaboração na formação da variável dependente, isto é, se contribuem positivamente ou negativamente.

O MMQO permite a obtenção do modelo inicial informando à priori que os resíduos ou erros da estimação mostram-se normalmente distribuídos ou não.

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Para que a análise seja completa avança-se ao método clássico da regressão linear (MCRL) onde são analisados os pressupostos estatísticos que conferirão confiabilidade o modelo, dentre os quais se destacam:

- 1 – Preparação da base dados ou amostra e estudo do mercado ao qual o bem se insere
- 2 – Minimização dos resíduos (erros)
- 3 – Normalidade dos resíduos.
- 4 – Outliers e Pontos Influenciadores dos resíduos
- 5 – Independência dos resíduos
- 6 – Linearidade dos resíduos
- 7 – Homocedastidade dos resíduos
- 8 – Multicolinearidade das variáveis

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

A elaboração eficiente e consistente da análise de uma Regressão Linear Múltipla (RLM) pelo Método dos Mínimos Quadrados Ordinários (MMQO) exige que os erros residuais sejam minimizados por métodos algébricos em relação a uma tendência média, com o objetivo de obter estimativas tanto quanto precisas necessárias para o atendimento de regras de precisão e especificações de confiabilidade e erro.

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Minimização dos resíduos (erros)

A minimização dos resíduos dos dados pertencentes a uma amostra em análise estatística consiste em aproximar tanto quanto possível os pontos dispersos em um espaço xy de uma reta de tendência como mostrado na figura 1 seguinte:

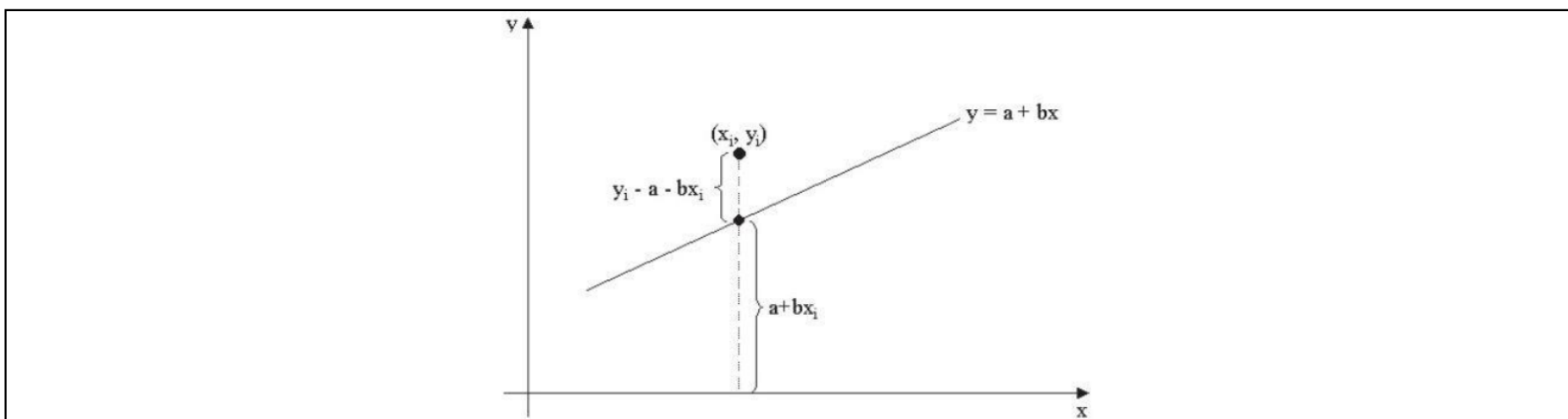


Figura 1 – Ajuste dos resíduos à reta de regressão

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Para que a equação de regressão se mostre adequada segundo as exigências normativas, é necessário elaborar um conjunto de testes e ensaios na amostra, comprovando através dos resultados obtidos que a função matemática obtida representa adequadamente uma média estimativa para o conjunto de valores da amostra, considerando as diferenças de características entre os dados da mesma.

Normalidade dos resíduos

A normalidade dos resíduos é o pressuposto fundamental para a análise correta da regressão linear simples (RLS) ou múltipla (RLM) pelo Método dos Mínimos Quadrados Ordinários (MMQO). Além de ser um pressuposto da própria RLS ou RLM, ele também é um pressuposto para a aplicação para a maioria dos testes formais em complemento às diversas verificações possíveis informais via gráficos de dispersão disponíveis nos softwares customizados, regularmente utilizados.

REALIZAÇÃO

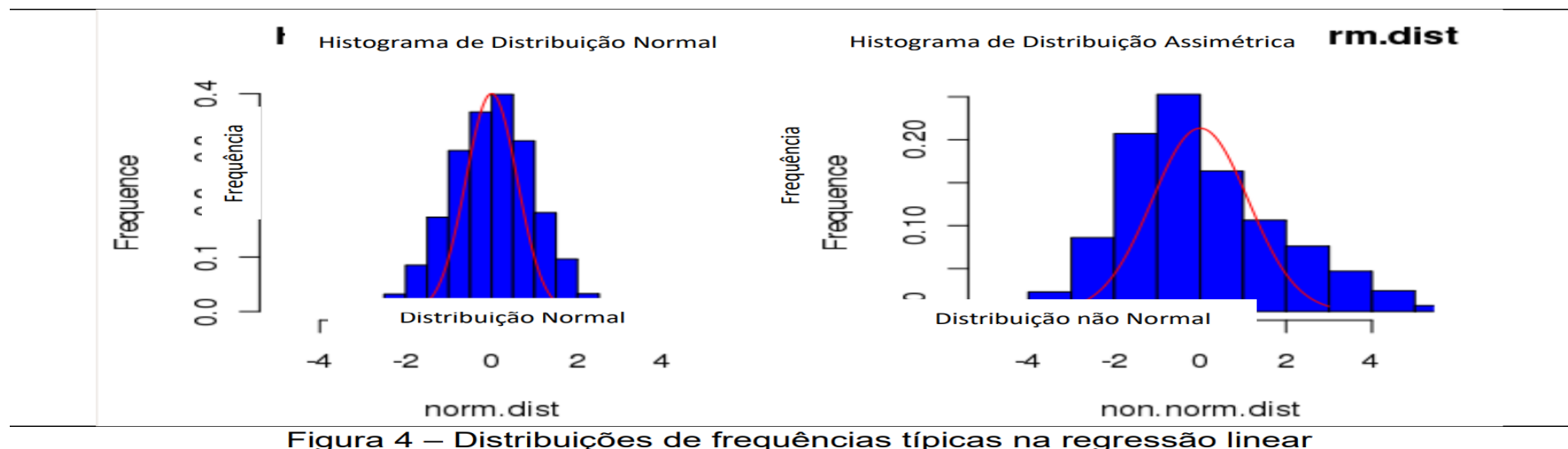


PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Na prática das avaliações o avaliador verifica a normalidade comparando os percentuais distribuídos entre ± 1 desvio padrão, ± 2 desvios padrão e ± 3 desvios padrão e aceita quando as variações entre os percentuais calculados e os percentuais 68%, 90% e 99% estão próximos até algo em torno de $\pm 5\%$ em cada intervalo. Ocorre que mesmo isso acontecendo, nem sempre o histograma de frequências dos dados aparenta normalidade como se observa na figura seguinte em que esses histogramas apresentam a forma normal e não normal de frequências.



ANALISE ESTATÍSTICA NA REGRESSÃO

Essa análise informal pode ser realizada utilizando-se a técnica do Papel (ou gráfico) de Probabilidade Normal em que é possível obter um conjunto de dados de entrada com duas series, uma com os resíduos em análise ordenados e outra com as probabilidades acumuladas padronizadas desses resíduos ordenados pela ordem de coleta.

Não obstante, outros padrões podem se apresentar na análise e sugerir que mais investigações devem ser implementadas para a busca da normalidade dos resíduos, fundamental para a análise da regressão linear com sucesso.

Para ensaio, podem ser tomados como parâmetros os seguintes intervalos de aceitação:

de $[-1,00 \text{ a } +1,00]$ s – Ideal 68% - aceitável: de 64% a 75%;
de $[-1,64 \text{ a } +1,64]$ s – Ideal 90% - aceitável: de 88% a 95%;
de $[-1,96 \text{ a } +1,96]$ s – Ideal 95% - aceitável: de 95% a 100%;

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Dois caminhos podem ser seguidos:

1 – Utilizando-se de um software de planilhas, por exemplo, o EXCEL, preparar um gráfico com as probabilidades acumuladas padronizadas (eixo y) x resíduos ordenados (eixo x) e a partir daí obter a reta de tendência dos resíduos plotados e o seu coeficiente de determinação R^2 .

Esse coeficiente calculado estando bastante próximo de 1, significa que os dados analisados têm uma distribuição normal.

A Figura 2 apresentada a seguir, demonstra algumas configurações das violações da normalidade possíveis de ser identificadas pela análise informal dos Gráficos de Papel de Probabilidade:

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Figura 2a - Curva alongada

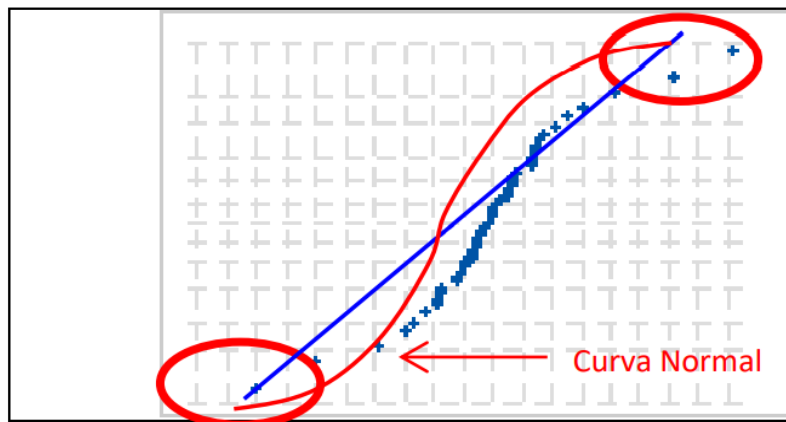


Figura 2b – Curva invertida

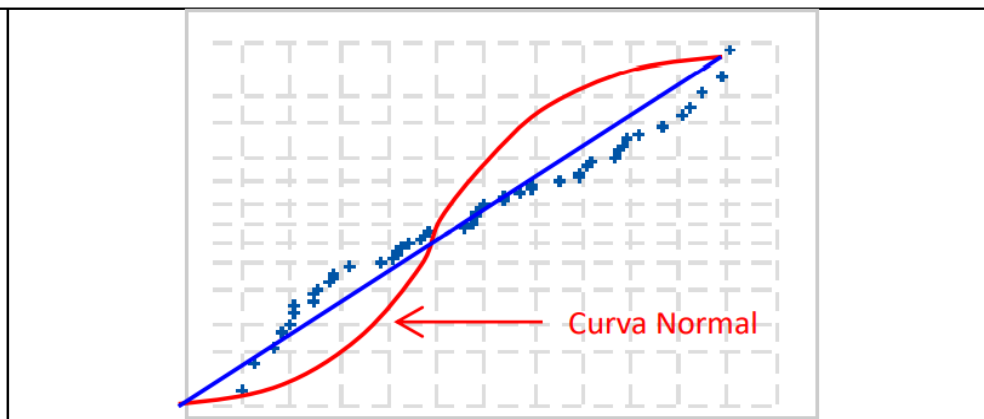


Figura 2c – Curva descendente

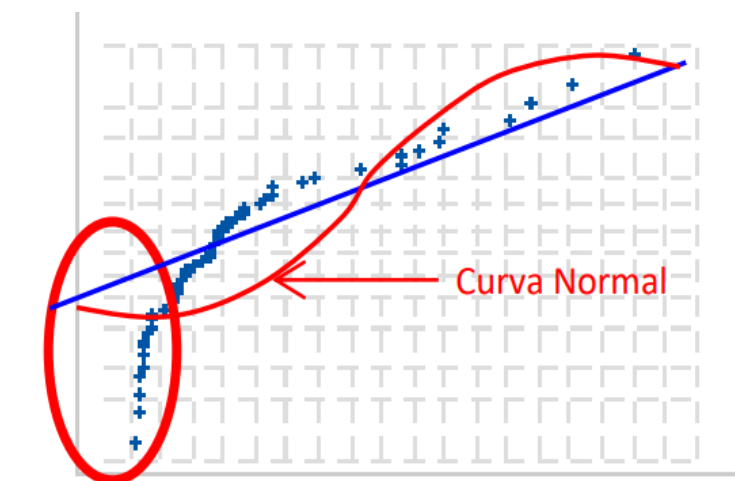
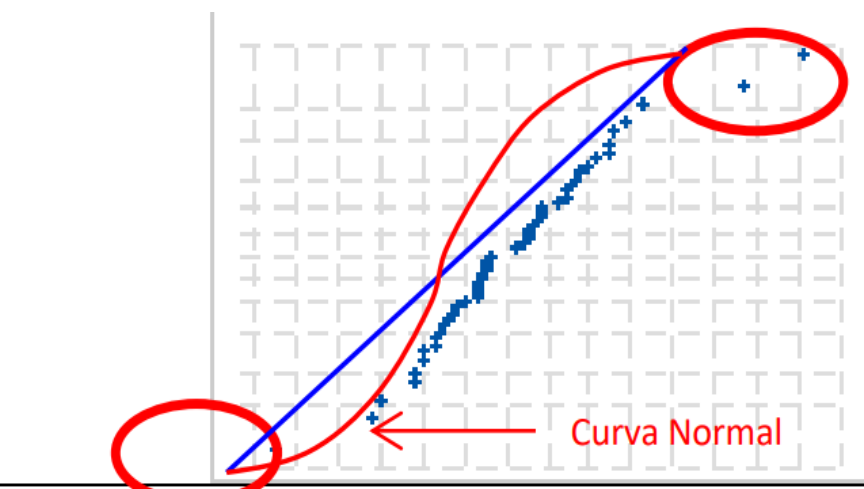


Figura 2d – Pontos distantes da tendência



REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

2 - Regredindo as Probabilidades Acumuladas Padronizadas (eixo y) x Resíduos Ordenados (eixo x), ainda usando o EXCEL, e analisando-se os resultados do valor-p dos coeficientes, define-se quantitativamente a confiabilidade da normalidade dos resíduos para uma significância especificada previamente, finalizando a análise. Uma análise formal, alternativa à análise informal anterior, pode ser feita utilizando-se do software R / R-Studio. Por esse caminho procede-se como indicado no quadro a seguir, aplicando-se os comandos seguintes finalizando com a conclusão e análise:

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

```
## Normalidade dos resíduos:

## Teste de Shapiro-Wilk

## Teste p-value para a distribuição normal dos resíduos

shapiro.test(mod$residuals)

## Nível de significância = 0,05

## Hipóteses

## H0: Distribuição dos resíduos = Normal => p > 0,05
## H0: Distribuição dos resíduos ? Normal => p = 0,05

## No Console do R-Studio deve aparecer a seguinte resposta:
>
≤

## data: mod$residuals
## w = 0,99555, p-value = 0,8269

## Análise e conclusão:

## Como p-value = 0,08269 (calculado) > p-value 0,05 (crítico)
## A distribuição dos resíduos é normal
```

Teste de Normalidade no R / R-Studio

Observa-se que neste exemplo didático o p-value calculado foi superior ao pvalue crítico, sugerindo que há evidências que a distribuição dos resíduos siga uma distribuição normal.

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Outliers e Pontos Influenciadores dos resíduos Os outliers, definidos na ABNT NBR 14.653-2 / 2011, item 3.48, são pontos atípicos e como afirma a norma, são estranhos à massa de dados formadora da amostra que pretende sintetizar o mercado colhida para análise e conclusão quanto ao valor de avaliação dentro de um intervalo de confiança. Não menos danosos, são os pontos influenciadores, definidos na ABNT NBR 14.653-2 / 2011, item 3.56 também como pontos atípicos que quando retirados da amostra, alteram significativamente os parâmetros estimados ou a estrutura do modelo. A ABNT NBR 14.653-2 / 2011 não explicita o motivo, mas cabe aqui informar que alteram em favor da qualidade e melhora da precisão do modelo. Os outliers, ainda que muitas vezes tolerados em modelos de regressão simples como os pontos influenciadores não devem ser admitidos conforme especifica a 14.653-2 / 2011 no item A.2 do Anexo 2. Esses pontos comprometem seriamente os resultados estimados pela da regressão linear, seja, RLS ou RLM pelo Método dos Mínimos Quadrados (MQO). As figuras com os gráficos dos resíduos a seguir demonstram graficamente como ocorre esse comprometimento pela mudança da posição da reta de tendência da regressão:

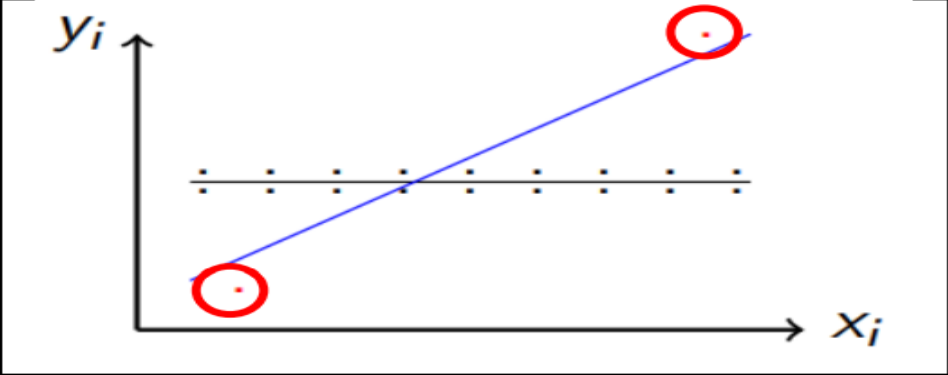
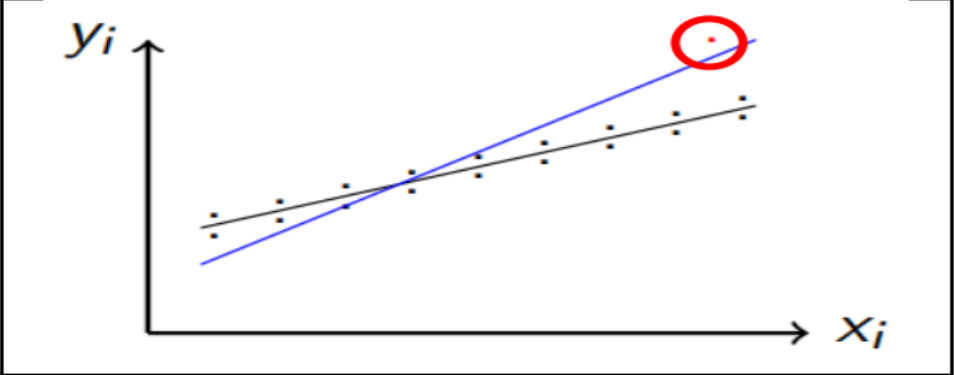
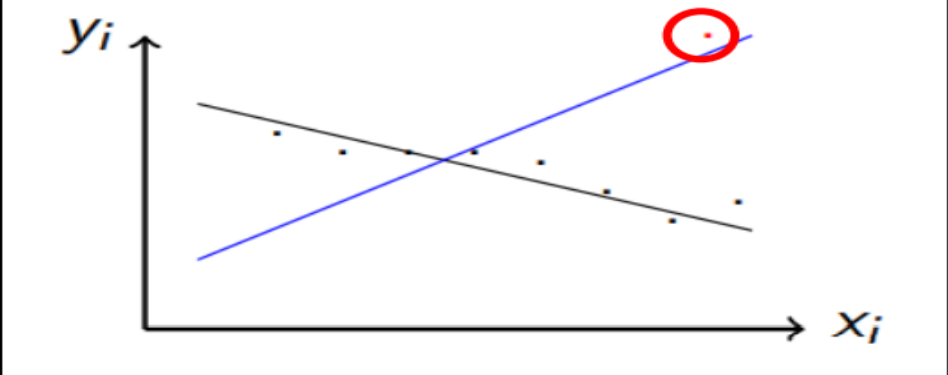
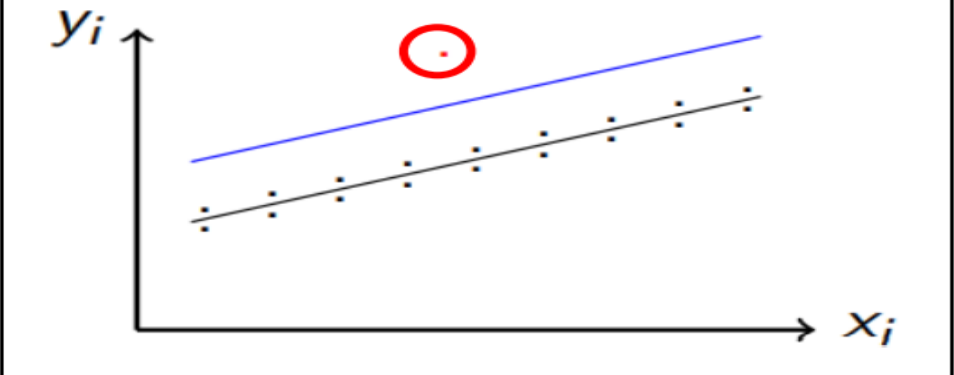
REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

	
O outliers provocam a inclinação	O outlier muda o intercepto
	
O outlier muda a inclinação	O outlier muda o intercepto
Fonte: Profª Silvia Nagib Elia / IME / USP	

REALIZAÇÃO

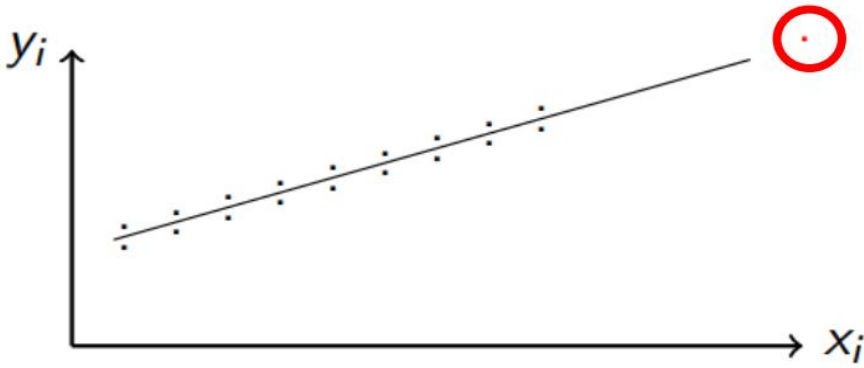
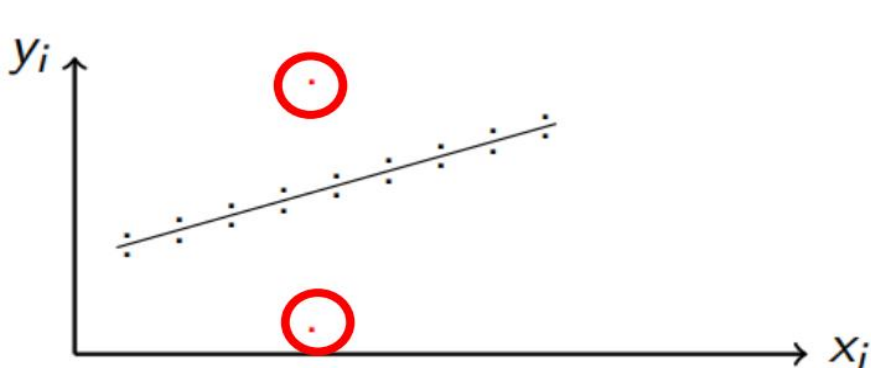


PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Não obstante, desde que cuidados adicionais sejam empreendidos existem pontos atípicos que não alteram a reta de regressão e após análise circunstanciada do modelo resultante podendo até ser tolerados, como demonstrado a seguir, desde que não se constituam em pontos Influenciantes no resultado, normalmente identificados pela Distância de COOK.

	
Outlier alinhado com a tendência	Outliers simétricos com a tendência
Fonte: Profª Silvia Nagib Elia / IME / USP	

ANALISE ESTATÍSTICA NA REGRESSÃO

Uma análise formal, alternativa às análises informais, que normalmente usamos pode ser feita utilizando-se do software R / R-Studio, análises estas além da anterior, tais como:

Independência dos resíduos

A independência dos resíduos não é comentada na ABNT NBR 14.653-2 / 2011 e constitui-se em fator importante para a obtenção de modelos confiáveis. Deve ser verificada sempre, pois que resíduos se auto correlacionando dentro da estrutura de dados da amostra não contribui para a normalidade dos mesmos, invalidando ou comprometendo seriamente a obtenção de resultados confiáveis, pois, quando não independentes, ainda que os coeficientes do modelo permaneçam não viesados ou não tendenciosos e consistentes.

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Linearidade dos resíduos

A linearidade dos resíduos é um pressuposto importante e indispensável até mesmo pelos próprios objetivos dos Método RLS ou RLM, lineares por definição. Quando esse pressuposto é violado, isso implica que a estimativa do erro aumentará, isto é, o intervalo de confiança será maior, porque nesta situação os valores observados na amostra não se aproximarão dos valores calculados ou preditos pela RLS ou RLM (local onde passará a reta). Portanto trata-se de um pressuposto fundamental já que as equações de regressão RLS ou RLM são modelos lineares por definição.

A não linearidade dos resíduos sugere que os resíduos tendem a seguir um padrão diferente da distribuição linear representativa de uma distribuição aleatória. Deve ser observada junto com a variância dos resíduos que devem se manter constante sem tendências ascendentes ou descendentes com as variáveis independentes.

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Homoscedasticidade dos resíduos

A homoscedasticidade é a consideração exigência para aceitação do modelo que a variância dos resíduos seja constante. Os resíduos não se comportando assim, são ditos heterocedásticos ou com variâncias variáveis. Essa variância não constante pode ser atribuída à diversas fontes, todas elas já examinadas nos tópicos anteriores quais sejam, a presença da não normalidade dos resíduos, a existência de outliers e/ou pontos influenciantes nos resíduos, a não independência dos resíduos e a não linearidade dos resíduos. Observe-se que não está listada a auto correlação das variáveis que será abordada a seguir e não influencia na variância do modelo.

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Existe vários testes formais para a verificação da heterocedasticidade, tais como:

Teste de Park

Teste de Glejser

Teste de Goldfeld-Quandt

Teste de Breusch-Pagan-Godfrey

Teste geral de heterocedasticidade de White

Teste de Park

Todos estes testes, sofrem influência da normalidade dos resíduos e temos um bem simples, que independe desta normalidade que é o teste de Koenker-Bassett (KB), o teste KB baseia-se nos resíduos elevados ao quadrado, regredidos em função dos valores estimados do modelo elevados ao quadrado.

ANALISE ESTATÍSTICA NA REGRESSÃO

RES^2	V Calc^2
19018,65	27262684,9
100944,3	23077672,5
23857,54	27430241,9
94940,5	16468371,8
386824,1	15396534,9
4509,278	12481020,9
15829,64	24837605,7
5866,406	18254577
522,8501	29998997,8
189190,1	13562206,7
31449,59	18685386,8
192720	7766316,14
15805,43	8261487,54
7220,189	6682078,67
1681,714	7797058,24
1621,177	9774415,46
314742,8	8103868,22

ANOVA									
	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>p</i>	<i>significação</i>			
Regressão	1	1,47E+10	1,47E+10	1,03816	0,324403				
Resíduo	15	2,12E+11	1,42E+10						
Total	16	2,27E+11							
	<i>Coeficiente</i>	<i>erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>% inferior</i>	<i>% superior</i>	<i>inferior 95,0%</i>	<i>superior 95,0%</i>	
Interseção	145101,2	67655,3	2,144714	0,048761	897,3782	289305,1	897,3782	289305,1	
V Calc^2	-0,00384	0,003771	-1,0189	0,324403	-0,01188	0,004196	-0,01188	0,004196	
				-1,34061		-1,75305			

A hipótese nula é de que $\alpha^2 = 0$. Se esta não for rejeitada, então se poderia concluir que não há heterocedasticidade. A hipótese nula pode ser testada pelo teste usual t ou pelo teste F.

REALIZAÇÃO

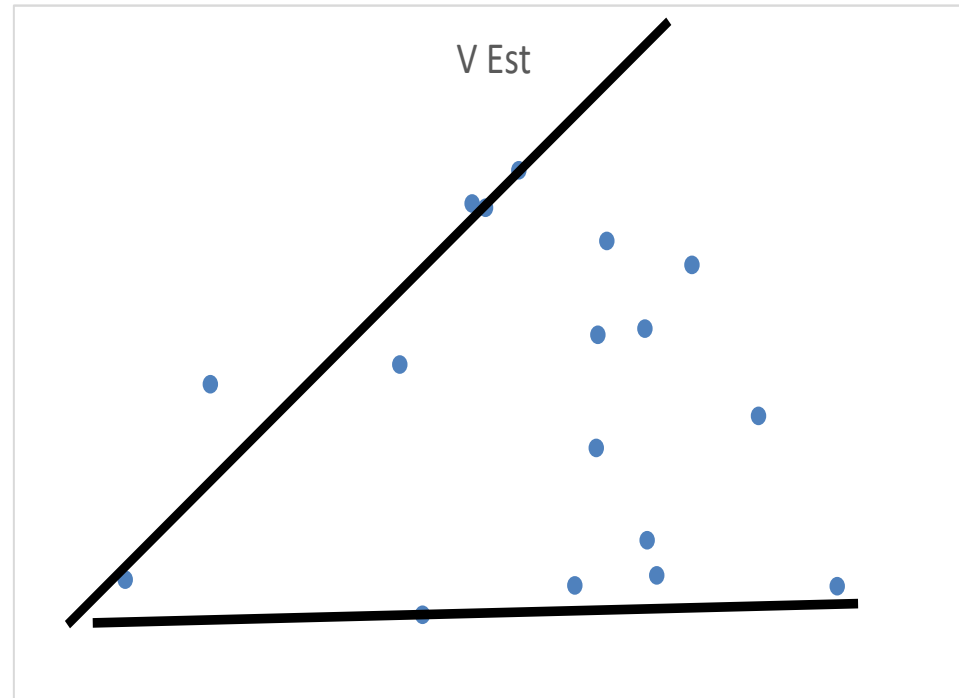
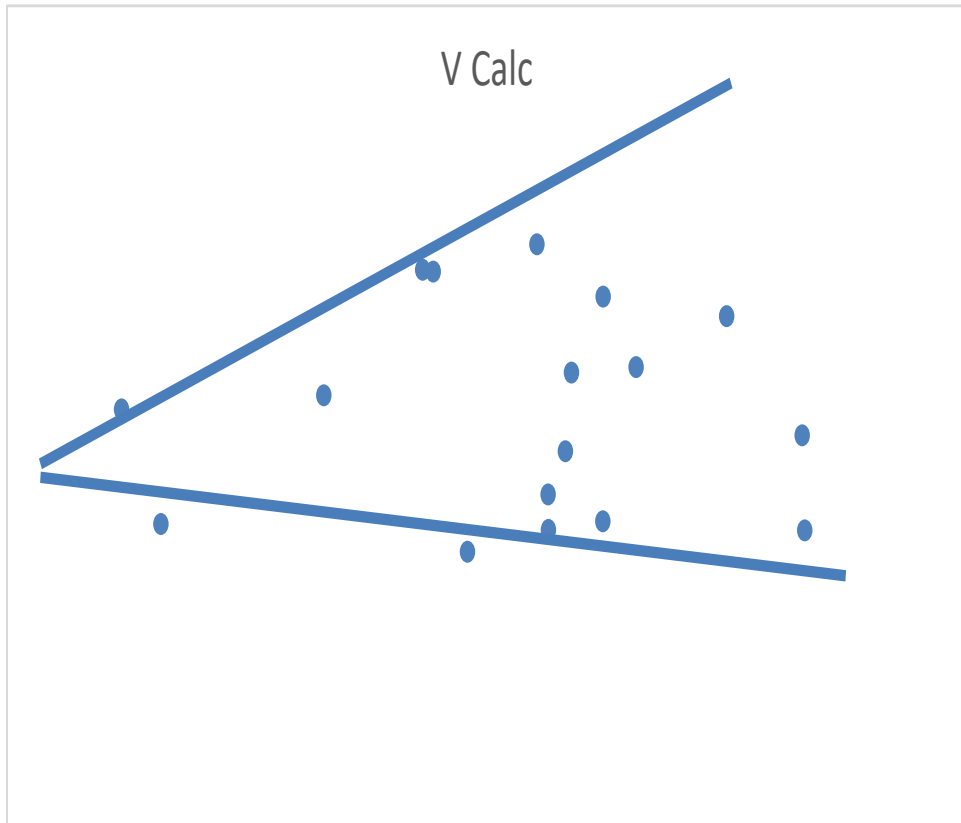


PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Vamos fazer então os gráficos da equação e da transformada dela.



Comprova a heterocedasticidade do modelo

REALIZAÇÃO



PATROCÍNIO



ANALISE ESTATÍSTICA NA REGRESSÃO

Um desafio que deixo para ser examinado e estudado, é que acredito que este teste serve também para escolhermos modelos, pois sabemos que o **valor-p**, nos mostra a probabilidade de erro das variáveis e no presente caso podemos observar que é grande a probabilidade de erro da estimativa.

	Coeficiente	erro padrão	Stat t	valor-P	% inferior	% superior	inferior 95,0%	superior 95,0%
Interseção	145101,2	67655,3	2,144714	0,048760789	897,3782	289305,1	897,3782	289305,1
V Calc^2	-0,00384	0,003771	-1,0189	0,324403229	-0,01188	0,004196	-0,01188	0,004196

Estes cuidados referente a resíduos, tenho observado em vários trabalhos que os mesmos tem sido negligenciados.

REALIZAÇÃO



PATROCÍNIO



OBRIGADO PELA ATENÇÃO DE TODOS

Radegaz Nasser Júnior

(+ 55 27) 999.810.178

radegaz@uol.com.br

REALIZAÇÃO



PATROCÍNIO

