

Módulo 16- Análise de Regressão

Situação Problema

Um grupo de investidores estrangeiros deseja aumentar suas atividades no Brasil. Considerando a conjuntura econômica de moeda desvalorizada que favorece a estrutura exportadora e recuperação econômica o grupo estuda a possibilidade de iniciar uma transportadora que atuaria mais no setor exportador.

Antes de decidir por realizar ou não o investimento, o grupo contratou uma empresa de consultoria para um estudo de mercado em que o ponto central é a mensuração da demanda para a atividade de transporte de cargas de forma geral no Brasil.

Este tutorial apresenta como analisar as tabelas na análise de regressão múltipla.

[Regressão Linear Múltipla](#)

Parte 2- Análise dos resultados

Como escolhemos o método de regressão “enter”, a tabela 1 nos mostra que todas as variáveis independentes selecionadas foram consideradas no estudo, nenhuma foi removida. Neste caso, como se trata de uma regressão múltipla com três variáveis independentes e todas elas foram consideradas no estudo.

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	variação no PIB real, produção de papel e papelão, valor exportado em reais		Enter

- a. All requested variables entered.
 b. Dependent Variable: quantidade de toneladas de carga transportadas

Tabela 1: Variáveis independentes incluídas na análise.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,987 ^a	,974	,970	14789,01

- a. Predictors: (Constant), variação no PIB real, produção de papel e papelão, valor exportado em reais
 b. Dependent Variable: quantidade de toneladas de carga transportadas

Tabela 2: Coeficientes de determinação.

R: é o coeficiente de correlação, ou seja, mostra a relação entre a variável dependente e a variável independente. É também chamado de coeficiente linear da regressão e neste caso é de 0,987;

R-square: é o coeficiente de determinação, que diz quanto da variável dependente é explicada pelo modelo de regressão, neste caso 97,40% da variabilidade na quantidade de cargas transportadas é explicada pelo modelo de regressão, isto é, pelos variáveis independentes. Observe como este valor foi maior que o obtido na regressão linear simples (com apenas uma variável);

Adjusted R-square: é o coeficiente de determinação ajustado. É analisado nas regressões múltiplas, pois faz o ajuste pela inclusão de variáveis independentes. Seu valor deve ser próximo do valor de R^2 , caso contrário é possível que variáveis independentes inadequadas tenham sido colocadas no estudo (variáveis que não exercem influência forte no estudo). Neste caso os valores são bem próximos, então, não parece haver variáveis inadequadas neste estudo;

Std. Error of the estimate: é o erro padrão da estimativa.

A tabela 3 mostra uma análise de variância.

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,62E+11	3	5,39E+10	246,534	,000 ^a
	Residual	4,37E+09	20	2,19E+08		
	Total	1,66E+11	23			

- a. Predictors: (Constant), variação no PIB real, produção de papel e papelão, valor exportado em reais
 b. Dependent Variable: quantidade de toneladas de carga transpostadas

Tabela 3: Análise de variância.

O objetivo de se realizar esta análise é de saber se a variação explicada (variação na variável dependente explicada pela variável independente) é suficientemente forte para provocar uma regressão. Compara a variação explicada com a variação residual (não explicada).

A coluna “mean square” mostra estas variações: *variação explicada* = $5,39 \times 10^{10}$ e *variação não – explicada* = $2,19 \times 10^8$.

Temos um teste de hipóteses:

$H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ (não existe regressão);

$H_1: \beta_1 \neq 0$ ou $\beta_2 \neq 0$ ou $\beta_3 \neq 0$ (existe regressão para pelo menos uma das variáveis consideradas no modelo).

Considerando um nível de significância $\alpha = 5\%$, temos o nível de significância observado $sig = 0\%$. Como $sig = 0\% < \alpha = 5\%$ devemos rejeitar a hipótese nula, ou seja, neste caso admitimos que existe regressão para pelo menos uma das variáveis independentes.

F é o valor da estatística “F” observado. Neste caso temos $F = 246,534$, isto indica a variação explicada é 246,534 vezes maior que a variação não-explicada.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	16992,930	29050,775		,585	,565	-43605,925	77591,785
	valor exportado em reais	-3,74E-04	,000	-,320	-2,842	,010	-,001	,000
	produção de papel e papelão	249,688	27,279	,650	9,153	,000	192,785	306,591
	variação no PIB real	1,056E-04	,000	,666	5,146	,000	,000	,000

a. Dependent Variable: quantidade de toneladas de carga transpostadas

Tabela 4: Resultados da análise.

A tabela 4 nos dá os coeficientes da equação de regressão.

Temos:

B: mostra os coeficientes da equação de regressão. A constante e os coeficientes das variáveis independentes:

$$Y = a_1x_1 + a_2x_2 + a_3x_3 + b = -0,000374x_1 + 249,688x_2 + 0,000374x_3 + 16992,930 ;$$

t: estatística t observada para cada variável;

sig: é o P-valor, ou seja, o nível de significância observado. Testa se a variáveis independentes influencia a variável dependente. O teste é realizado para cada variável separadamente:

a) Valor exportado em reais:

H₀: $\beta_1 = 0$ (a variável independente não influencia a variável dependente);

H₁: $\beta_1 \neq 0$ (a variável independente influencia a variável dependente).

Como o sig para a variável independente é menor que o nível de significância $\alpha = 5\%$ rejeitamos a hipótese nula, ou seja, as exportações influenciam a quantidade de carga transportada nas rodovias, considerando nível de significância de 5%.

b) Produção de papel e papelão:

H₀: $\beta_2 = 0$ (a variável independente não influencia a variável dependente);

H₁: $\beta_2 \neq 0$ (a variável independente influencia a variável dependente).

Como o sig para a variável independente é menor que o nível de significância $\alpha = 5\%$ rejeitamos a hipótese nula, ou seja, a produção de papelão influencia a quantidade de carga transportada nas rodovias, considerando nível de significância de 5%.

c) Variação do PIB real:

H₀: $\beta_3 = 0$ (a variável independente não influencia a variável dependente);

H₁: $\beta_3 \neq 0$ (a variável independente influencia a variável dependente).

Como o sig para a variável independente é menor que o nível de significância $\alpha = 5\%$ rejeitamos a hipótese nula, ou seja, a variação do PIB influencia a quantidade de carga transportada nas rodovias, considerando nível de significância de 5%.

A tabela 5 mostra estatísticas descritivas dos resíduos.

Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	219929,70	477783,94	342257,63	83863,74	24
Residual	-34685,92	21235,80	2,43E-12	13790,82	24
Std. Predicted Value	-1,459	1,616	,000	1,000	24
Std. Residual	-2,345	1,436	,000	,933	24

a. Dependent Variable: quantidade de toneladas de carga transportadas

Tabela 5: Análise dos resíduos.

Os gráficos a seguir servem para a análise de resíduos.

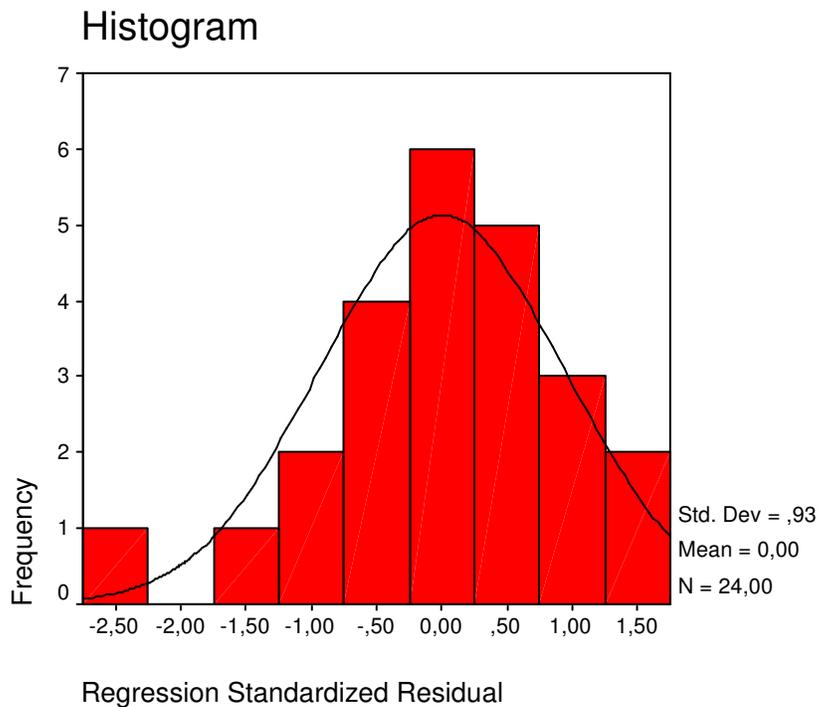


Gráfico 1: Histograma dos resíduos.

O gráfico 1 é um histograma dos resíduos padronizados. Este gráfico deve seguir uma distribuição normal. Neste caso temos uma distribuição que relativamente se aproxima de uma normal.

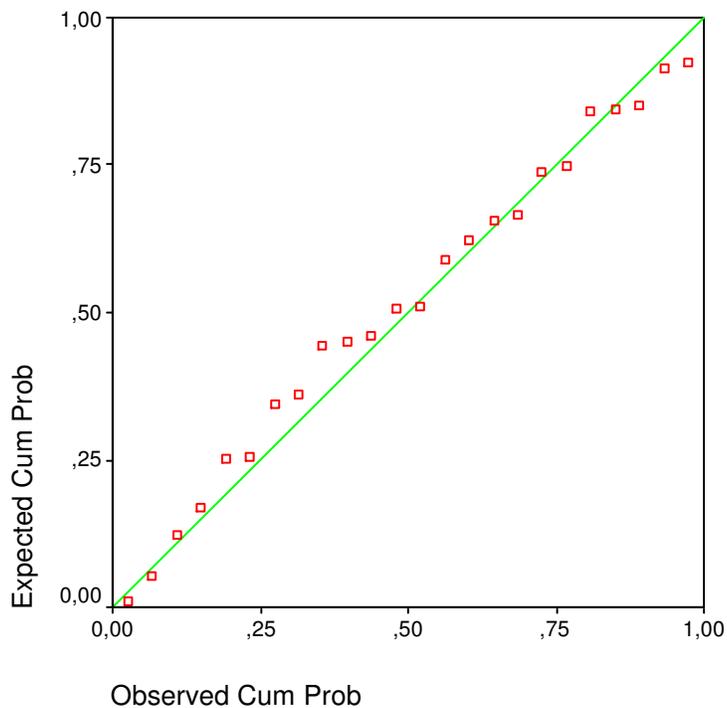


Gráfico 2: Plotagem de resíduos padronizados.

O gráfico 2 mostra a distribuição das probabilidades dos resíduos. Mostra a probabilidade de cada ponto (erro) ocorrer. Eles devem estar o mais próximo possível da reta sem se aglomerar em só lugar da reta.

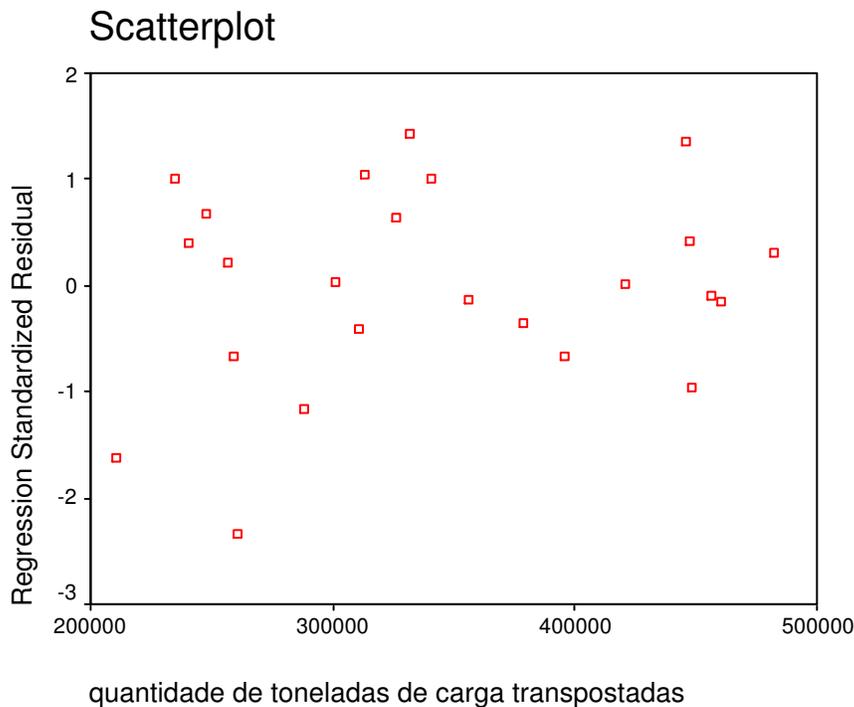


Gráfico 3: Gráfico de dispersão dos resíduos.

O gráfico 3 é um diagrama de dispersão dos erros. Idealmente os pontos deveriam situar-se próximos de zero formando uma figura parecida com uma nuvem. Neste caso percebemos que os pontos situam-se relativamente próximos de zero, o que indica não ter havido muita subestimação ou superestimação das previsões.

Uma ressalva faz-se necessária! Geralmente não se deve utilizar variáveis altamente correlacionadas como variáveis independentes na regressão ao mesmo tempo, podendo gerar o problema da multicolinearidade. Neste exemplo, em especial, para fins didáticos, foram utilizadas as mesmas variáveis do módulo 15- Análise de Correlação, mesmo que com correlação alta entre as variáveis independentes.