

UNIVERSIDADE DE PASSO FUNDO FACULDADE DE CIÊNCIAS ECONÔMICAS, ADMINISTRATIVAS E CONTÁBEIS CENTRO DE PESQUISA E EXTENSÃO DA FEAC (www.upf.br/cepeac)

Texto para discussão

Texto para discussão Nº 06/2021

Roteiro RLS (Regressão Linear Simples) e RLM (Regressão Linear

Múltipla) no Software Statistica 7.0

Andre da Silva Pereira Karolynne Sousa Gomes Leticia Eduarda Bender Valeska Morgana Corrêa Batistela

Passo Fundo - RS - Brasil

UNIVERSIDADE DE PASSO FUNDO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE ALIMENTOS

Roteiro RLS (Regressão Linear Simples) e RLM (Regressão Linear Múltipla) no Software Statistica 7.0

> André da Silva Pereira Karolynne Sousa Gomes Leticia Eduarda Bender Valeska Morgana Corrêa Batistela

> > Passo Fundo-RS 2021

LISTA DE ILUSTRAÇÕES

Figura 1: Gráfico de regressão linear simples	2
Quadro 1- Dados contendo variáveis dependente e independente	3
Figura 2: Importação da tabela no software Statistica 7.0	4
Figura 3: Passo 1	5
Figura 4: Passo 2	6
Figura 5: Passo 3	6
Figura 6: Passo 4	7
Figura 7: Resultado do exemplo gerado para regressão linear simples	7
Figura 8: Passo 1 (gráfico)	8
Figura 9: Passo 2 (gráfico)	9
Figura 10: Passo 3 (gráfico)	9
Figura 11: Geração do gráfico de dispersão entre as variáveis tempo e distância	10
Figura 12: Selecionar a opção para regressão linear múltipla	11
Figura 13: Resultados gerados para a regressão múltipla	11
Figura 14: Tabela para a geração do gráfico (Passo 2).	12
Figura 15: Selecionar as varáveis (Passo 3).	12
Figura 6: Gráfico de dispersão RLM	13

SUMÁRIO

INTI	RODUÇÃO	1
1.	ROTEIRO DE REGRESSÃO LINEAR SIMPLES (RLS) NO STATISTICA 7.0	2
2.	ROTEIRO DE REGRESSÃO LINEAR MÚLTIPLA (RLM) NO STATISTICA 7.0	3
3.	DIAGRAMA DE DISPERSÃO	3
4.	EXEMPLO DE REGRESSÃO SIMPLES NO STATISTICA 7.0	3
5.	EXEMPLO DE ANÁLISE DE REGRESSÃO MÚLTIPLA NO STATISTICA 7.0	10
REF	ERÊNCIAS	15

APOSTILA MÉTODOS QUANTITATIVOS

INTRODUÇÃO

A análise de regressão linear simples e múltipla é amplamente utilizada em diversas áreas do conhecimento, para investigar relações entre variáveis e construção de um modelo (equação).

• **Regressão Linear Simples:** definida como uma relação linear entre a variável dependente e uma variável independente.

• **Regressão Linear Múltipla:** relação linear entre várias variáveis independentes e uma variável dependente.

Entre os resultados que são obtidos a partir da análise de correlação e regressão, estão:

• R múltiplo: expressa o nível de correlação entre a variável dependente e independente, quanto maior o valor melhor o modelo;

• R²: representa o quanto os dados estão próximos da regressão ajustada, representa o quanto o modelo criado representa os dados;

• R² ajustado: é uma versão modificada do R², ajustada para o número de preditores no modelo;

• F: utilizado para determinar se a variabilidade entre as médias do grupo é maior que a das observações dentro dos grupos;

• p: é a probabilidade de se obter teste igual ou muito próximo do observado em uma amostra, sob hipótese nula. Muitas vezes é utilizada significância de 5 % em que pode ser rejeitada a hipótese nula caso o valor-p seja menor que 5 %;

• Erro padrão da estimativa: média de variação de uma média amostral em relação à média da população.

Entre os softwares utilizados para aplicação da regressão linear estão o Excel, a linguagem de programação R, o SPSS e o Statistica 7. Nesta apostila será (1) introduzido conceitos de regressão simples e múltipla, (2) descrito como aplicar os dados e variáveis no software Statistica 7.0, (3) realizada análise dos dados obtidos através dos exemplos.

1

1. ROTEIRO DE REGRESSÃO LINEAR SIMPLES (RLS) NO STATISTICA 7.0

- Envolve variáveis dependentes (VD) ou independentes (VI);
- VD é explicada ou saída;
- VI é explicativa ou preditora;

A regressão linear apresenta forma linear e pode ser definida através da Equação 1.

$$Yi = \alpha + \beta * Xi$$
 [Equação 1]

Onde:

Y = valor previsto da variável dependente;

i, $\alpha \in \beta$ = parâmetros estimados do intercepto e inclinação do modelo;



Figura 1: Gráfico de regressão linear simples.

Χ

Fonte: FÁVERO; BELFIORE, 2017.

2. ROTEIRO DE REGRESSÃO LINEAR MÚLTIPLA (RLM) NO STATISTICA 7.0

 $Yi = a + b1 * X1i + b2 * X2i + \dots + bk * Xki + ui$ [Equação 2]

Onde:

Y = Fenômeno sendo estudado (variável dependente quantitativa);

a = intercepto (constante ou coeficiente linear);

bj (j = 1, 2, ..., k) = coeficiente angular;

Xj = variável explicativa;

u = termo do erro (diferença entre Y real e calculado);

i = observações das análises.

Para a interpretação dos resultados das análises de regressão gera-se um diagrama de dispersão, que permite decidir experimentalmente se uma relação linear entre as variáveis deve ser estabelecida.

3. DIAGRAMA DE DISPERSÃO

Por interpretação do diagrama ou gráfico de dispersão, definimos o grau de relação linear entre as variáveis, se é uma relação forte ou fraca. Conforme o modo como se situam os pontos no diagrama ao redor de uma reta que passa entre os pontos. Considera-se a correlação forte quando os pontos se concentram mais, com pequenos desvios, em relação a reta.

4. EXEMPLO DE REGRESSÃO SIMPLES NO STATISTICA 7.0

A seguir, será representado em um exemplo como fazer análise de regressão linear simples e múltipla utilizando o software Statistica 7.0. Os dados que serão utilizados para exemplificar estão apresentados no Quadro 1.

	Zuuulo I Duuos eo	intendo variaveis de	pendente e indepen	dente
Tempo (min) (Y)	Distância (km) (X1)	Quantidade de Cruzamentos (X2)	Quantidade de Cruzamentos (X2)	Quantidade de Cruzamentos (X2)

Quadro 1- Dados contendo variáveis dependente e independente

15	8	16	32	12
20	6	12	24	20
20	15	30	60	25
40	20	40	39	37
50	25	50	100	32
25	11	22	44	17
10	5	10	20	9
55	32	64	128	60
35	28	56	112	12
30	20	40	80	17

Os dados da tabela do exemplo devem ser importados para a tabela disponível no software Statistica 7.0, como mostra a Figura 2.

Figura 2: Importação da tabela no software Statistica 7.0

STATISTICA -	[Data: Sprea	dsheet1* (1	00v by 100c			-	,												0	×
🗅 😂 🔛 🤞	≣ 🖪 🐇	🗈 🛍 😒	\$ n c	曲 15 /	dd to Work	book • A	dd to Repor	t - 🤌 🅅	?.											
Arial		• 10 •	BI	ī ≣≣	≡ 🗗 .	<u>A</u> - <u>A</u> -	a • 🇞 🛛	1.78 +98	1 H 🖬	🗑 💩 👌	E=? X1=?	Vars - Case	s • 🖾 –	∞						
Eile Edit	View Inser	t F <u>o</u> rmat	Statistics	<u>G</u> raphs <u>T</u> o	ols <u>D</u> ata	Window	<u>H</u> elp												_ 6	×
	1	L	uantidade	uantidade	obebitneu	6	7	8	9	10	11	12	13	14	15	16	17	18	19	-
	Tempo Y	Distância	de	de	de	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	
1	15	8	16	32	12															-
2	20	6	12	24	20															
3	20	15	30	60	25															
4	40	20	40	39	37															
5	50	25	50	100	32		_	_												
6	25	11	22	44	17			1												-
7	10	5	10	20	9															
8	55	32	64	128	60		_													-
9	35	28	56	112	12		-													
10	30	20	40	80	1/		-													
11																				
12																				-
13																				
14							-			-										-
16							-													
17																				-
18																				
19																				
20																				
21																				
22																				
23																				
24																				
25																				
26																				-
1																				١
M																				

Fonte: AUTOR, 2021

Após todos os dados estarem devidamente descritos no programa, devemos abrir a aba "Statistics" e procurar pela função "Multiple Regression".

Desta forma, será aberta uma nova janela no programa para selecionarmos as variáveis dependentes e a variável independente. Assim, será gerada tabela em que estarão o valor de R múltiplo, R2 múltiplo, R2 ajustado, F, p, e o erro estimado.

Para realização da análise de regressão linear simples os passos seguintes deverão ser realizados.

1. "Statistics" - Selecionar a opção "Multiple Regression"

2."Variables"

- 3. Selecionar as variáveis dependentes e independentes
- 4. Selecionar a opção "summary Regression Results".

Os passos citados acima estão exemplificados nas figuras abaixo.

		Fig	gura	a 3: Passo 1	
Eile Edit	View Inser	t Format	Sta	tistics Graphs Jools Data Window Hel	р
			48	Resume Ctrl+R	
	1 Tempo Y	Distância	2	ByGroup Analysis	7 Var7
1	15	8	四	Basic Statistics/Tables	
2	20	6	1	Multiple Regression	
3	20	15	1.1	ANOVA	
4	40	20	1 AL	Nanavametrica	
5	50	25	Lan	Nonparametrics	
6	25	11	2	Distribution Fitting	
7	10	5	12	Advanced Linear/Nonlinear Models	
8	55	32	R	Multivariate Exploratory Techniques	
9	35	28		Industrial Statistics & Six Siama	
10	30	20	111	Deves Analysis & Six Signa	
11			-	Power Analysis	
12			16	Neural Networks	
13			10	Data-Mining	
14			2	QC Data Mining & Root Cause Analysis +	
15			TUST.	Text & Document Mining, Web Crawling +	
16			-		
17				Statistics of Block Data	
18			\$	STATISTICA Visual Basic	
19			-		
20			2	Probability Calculator	
21					
22					
23					
24					
25					
26					

Fonte: AUTOR, 2021

Figura 4: Passo 2	
Z Multiple Linear Regression: Spreadsheet1	>
Quick Advanced	CK OK
✓ariables	Cancel
Dependent none Independent: none	Doptions •
	🗁 Ogen Data
	Weighted moments
	GW-1 CN-1
	MD deletion
	Casewise C Pairwise
See also the General Regression Models (GRM) module	C Mean substitution

Fonte: AUTOR, 2021

Figura 5: Passo 3

1-Tempo Y 2. Distância V	1-Tempo Y 2 Distância X	OK.
2-Distancia A 3-Quantidade de cruzamentos 4-Quantidade de cruzamentos	3-Quantidade de cruzamentos 4-Quantidade de cruzamentos	Cancel
5-Quantidade de Cruzamentos 6-Var6 8-Var8 8-Var8 9-Var9 10-Var10	5-Quantidade de Cruzamentos 6-Var6 7-Var7 8-Var8 9-Var9 10-Var10	Use the "Show appropriate variables only" option to pre-screen variable lists and
< >	< >	show categorical and continuous
Select All Spread Zoom	Select All Spread Zoom	F1 for more
Dependent var. (or list for batch):	Independent variable list:	information.
1	2	-

Fonte: AUTOR, 2021

Figura 6: Passo 4

_ Multiple Regression Results. Spread	asheet I	
Multiple Regression Result	:s	
Dependent: Tempo Y	Multiple R = ,90522134 F = 36, R ² = ,81942568 df = 1	30309 ,8
No. of cases: 10	adjusted R*= ,79685389 p = ,00	00314
Intercept: 5,878378378 St	d.Error: 4,532328 t(8) = 1,2970	p = ,230
(significant betas are high	lighted)	<u> </u>
(significant betas are high phaforhighting effects: 1.05	lighted)	
(significant betas are high phaforhighting effects: 0.5 😭 Quick Advanced Residuals/assumpt	ilighted)	Cancel
(significant betas are high phaforhighfing effects: 0.5 🔵 Quick Advanced Residuals/assumpt Bar Summary: Regression results	tions/prediction	Cancel
(significant betas are high phafor highlighting effects: 0.5 () Quick Advanced Residuals/assumpt Summary: Regression results () () () () () () () () () () () () ()	tions/prediction	Cancel
(significant betas are high phafor highlighting effects: 0.5 🚔 Quick Advanced Residuals/assumpt Summary: Regression result ANOVA (Overall goodness of fit) Cogatiance of coefficients	tions/prediction Image: Partial corretations Image: Partial corretations	Cancel

Fonte: AUTOR, 2021

Figura 7: Resultado do exemplo gerado para regressão linear simples

Summary
Value
0,90522
0,81943
0,79685
36,30309
0,00031
6,71890

Fonte: AUTOR, 2021

O "Multiple R" significa a correlação entre as variáveis tempo e distância, como o resultado obtido no exemplo é próximo de 1 (quanto mais próximo de 1 mais forte a correlação) maior significa que a correlação é positiva.

O "Multiple R²" é a explicação do modelo utilizado, portanto em 81,94% dos casos a distância explica chegar atrasado.

O "Adjusted R²" é a confiabilidade do modelo

O "F" é para a comparação dos modelos lineares

O "p" está relacionado com o teste de hipótese, se for igual a 0 significa que o ajuste do modelo não tem efeito, se for menor que 0,05 indica que se pode rejeitar a hipótese nula, ou seja provavelmente será uma adição significativa ao modelo

O "Err. Of Estimate" é a distância das médias em relação a linha de regressão

Após a realização da análise um gráfico de dispersão deve ser gerado. Para a interpretação dos resultados.

O gráfico será gerado seguindo os passos abaixo.

- 1. "Graphs" selecione a opção "scatterplots"
- 2. Selecione as variáveis- "Regular"
- 3. Selecionar as variáveis dependente e independente (apenas uma por vez).

Multiple R 0.90 Multiple R 0.91 Multiple R 0.92 Multiple R 0.92 Adjusted R ² 0.75 Scatterplots Scatterplots p 0.00 Std.Err. of Estimate 6.71 D Graphs 3D Sequential Graphs Matrix Plots Matrix Plots Matrix Plots Matrix Plots Graphs of Block Data Matrix Plots				(preadsheet1)
Multiple R 0.90 Multiple R ² 0.81 Adjusted R ² 0.75 Sum Scatterplots F(1.8) 36,30 P 0.00 Std.Err. of Estimate 6,71 Image: Stress of the stres	tistic	Valı 🖓	ByGroup Analysis	Series and the series of the
Multiple R ^a 0,81 Image: Histograms Adjusted R ^a 0,75 Scatterplots F(1,8) 36,30 Means w/Error Plots p 0,00 Surface Plots Std.Err. of Estimate 6,71 2D Graphs Std.Err. of Estimate 6,71 Image: Surface Plots Matrix Plots SD Sequential Graphs Matrix Plots Matrix Plots Matrix Plots Stage: Surface Plots Graphs of Block Data Graphs of Block Data Stage: Surface Plots	tiple R	0,90	11 de la compañía de	
Adjusted R ^a 0,75 Scatterplots F(1,8) 36,30 Means w/Error Plots p 0,00 Surface Plots Std.Err. of Estimate 6,71 Image: Std.Err. of Estimate 7,72 Image: Std.Er	tiple R ²	0,81	Histograms	the second se
F(1,8) 36,3C LE ^N Means <u>w</u> /Error Plots p 0,0C Le ^N Surface Plots Std.Err. of Estimate 6,71 Image: Constraint of Estimate 9	usted R ²	0,79	Scatterplots	
p 0,00 less Surface Plots Std.Err. of Estimate 6,71 Image: Std.Err. of Estimate 3D Sequential Graphs Image: Std.Err. of Estimate Matrix Plots Image: Std.Err. of Estimate 1 Con Plots Image: Std.Err. of Estimate 1 Std.Err. of Estimate Image: Std.Err. of Estimate 1 Std.Err. of Estimate <t< td=""><td>(8)</td><td>36,30</td><td>Means w/Error Plots</td><td></td></t<>	(8)	36,30	Means w/Error Plots	
Std.Err. of Estimate 6,71 Image: Constraint of Estimate Image: Constraint of Estimate Image: Constraint of Estimate Image: Constraint of Estimate <td></td> <td>0,00 😂</td> <td>Surface Plots</td> <td></td>		0,00 😂	Surface Plots	
Bo Graphs Bo Graphs So Sequential Graphs So Sequential Graphs So Sequential Graphs So Sequential Graphs Matrix Plots So Icon Plots Categorized Graphs User-defined Graphs So User-defined Graphs	Err. of Estimate	6,71	2D Graphs	
So Sequencial Graphs So Sequencial Graphs So Sequencial Graphs Matrix Plots Categorized Graphs User-defined Graphs So Sequencial Graphs Graphs of Block Data			2D Sequential Graphs	States in the second
		i i i i i i i i i i i i i i i i i i i	3D VVZ Create	
Matrix Plots Con Plots Categorized Graphs User-defined Graphs Graphs of Block Data		S	SD XYZ Graphs	-
Categorized Graphs User-defined Graphs Graphs of Block Data			Matrix Plots	
User-defined Graphs		25	Icon Plots	
User-defined Graphs			Categorized Graphs	•
Graphs of Block Data			User-defined Graphs	•
		5	Graphs of Block Data	
Graphs of Input Data		1	Graphs of Input Data	•
Multiple Graph Layouts		di s	Multiple Graph Layouts	

Fonte: AUTOR, 2021

2D Scatterplots		×
Quick Advanced	Appearance Categorized Options 1 Options 2	
X: no	ine ine	
Graph type: Regular Multiple	Regression bands Off level: Confidence .95	
Fit type: Linear		
	Doptions -	OK Cancelar

Figura 9: Passo 2 (gráfico)

Fonte: AUTOR, 2021

Figura	10:	Passo	3	(gráfico)
--------	-----	-------	---	-----------

1-Tempo Y 2 Distância Y	1-Tempo Y	OK
2-Distancia A 3-Quantidade de cruzamentos 4-Quantidade de cruzamentos 5-Quantidade de Cruzamentos 6-Var6 7-Var6 8-Var8 9-Var9 10-Var10	3-Quantidade de cruzamentos 4-Quantidade de cruzamentos 5-Quantidade de Cruzamentos 6-Var6 7-Var7 8-Var8 9-Var9 10-Var10	Cancel
< >	< >	
Select All Spread Zoom	Select All Spread Zoom	
×	Y:	
1	2	

Fonte: AUTOR, 2021



Figura 11: Geração do gráfico de dispersão entre as variáveis tempo e distância.

Fonte: AUTOR, 2021

O gráfico indica uma correlação linear positiva forte, pois como pode se observar na Figura 10, a linha segue em sentido crescente indicando assim que a correlação apresenta valor próximo de 1.

5. EXEMPLO DE ANÁLISE DE REGRESSÃO MÚLTIPLA NO STATISTICA 7.0

Para realização da análise de regressão múltipla segue-se basicamente os mesmos passos, porém com algumas alterações como descrito a seguir:

> 1. "Statistics" -Selecionar a opção "Multiple Regression"

2. "Variables" - Selecionar a variável dependente e as variáveis independentes.

Observação: Na regressão múltipla as variáveis são selecionadas todas de uma vez diferente de como ocorre na regressão linear simples onde as variáveis independentes são selecionadas uma por vez.

3. Na tela inicial da definição de variáveis deve-se ir na opção "method" e selecionar a opção "forward stepwise".

4. E então apertar em "OK" e a tabela de regressão será gerada.

Model Definition: Spreadsheet1	×
Quick Advanced Stepwise Descriptives	Cancel

Figura 12: Selecionar a opção para regressão linear múltipla.

Fonte: AUTOR, 2021

	Summary
Statistic	Value
Multiple R	0,94947
Multiple R ²	0,90150
Adjusted R ²	0,87335
F(2,7)	32,03212
P	0,00030
Std.Err. of Estimate	5,30505

Figura 13: Resultados gerados para a regressão múltipla.

A interpretação dos dados apresentados segue os mesmos padrões de interpretação da análise de regressão linear simples. Porém para o diagrama de dispersão na RLM (regressão linear múltipla) é necessário seguir os passos abaixo.

Geração do gráfico de dispersão RLM

- 1. "Graphs" selecione a opção "scatterplots"
- 2. Selecione no painel inicial das variáveis a opção "multiple"
- 3. Apertar na opção "variáveis" e selecionar as variáveis para geração do

gráfico.

Observação: Aqui todas as variáveis independentes serão selecionadas de uma só vez.

4. Aperta em "OK" e o gráfico será gerado.

Fonte: AUTOR, 2021

2D Scatterplots		×	14	15	16	17
Quick Advanced Appearance Categorize	d Options 1 Options 2		Var14	Var15	Var16	Var1/
<u>V</u> ariables: X: Tempo Y Y: Distância X						
	Select Variables for Scatterplot				?	X
Graph type: Regular Multiple Fit type: ✓ Linear Graph type: ✓ Prediction	1-Tempo Y 2-Distância X 3-Quantidade de cruzamentos 5-Quantidade de Cruzamentos 6-Var6 7-Var7 8-Var8 9-Var9 10-Var10 < Select All Spread Zoom 	1-Tempo Y 2-Distánci 3-Quantid 4-Quantid 5-Quantid 6-Var6 7-Var7 8-Var8 9-Var9 10-Var10 < Select All	a X ade de cruza ade de Cruza ade de Cruza	amentos amentos amentos d Zoor	2 2	ncel
Deptions -	Show appropriate variables only OK Ca	ncelar				

Figura 14: Tabela para a geração do gráfico (Passo 2).

Fonte: AUTOR, 2021



1-Tempo Y 2-Distência X	1-Tempo Y 2.Distância X	0K		
3-Quantidade de cruzamentos 4-Quantidade de cruzamentos	3-Quantidade de cruzamentos 4-Quantidade de cruzamentos	Cancel		
5-Quantidade de Uruzamentos 6-Var6 7-Var7 8-Var8 9-Var9 10-Var10	6-Vushidade de Lruzamentos 6-Var6 7-Var7 8-Var8 9-Var9 10-Var10	Use the "Show appropriate variables only" option to pre-screen variable lists an		
<	< >	show categorical and continuous		
Select All Spread Zoom	Select All Spread Zoom	F1 for more information.		
1	2-5			

Fonte: AUTOR, 2021





Fonte: AUTOR, 2021

O tempo possui forte correlação com todas as variáveis independentes e que a hipótese do modelo utilizado não é nula.

O software Statistica 7.0 pode ser utilizado para as análises de regressão linear, tanto simples quanto múltiplas. Apresentando as ferramentas necessárias para que os dados de regressão linear sejam interpretados corretamente, como por exemplo o diagrama de dispersão.

REFERÊNCIAS

HENRIQUES, Carla. Análise de regressão linear simples e múltipla. **Departamento de Matemática. Escola Superior de Tecnologia de Viseu**. Portugal, 2011.

FÁVERO, L. P.; BELFIORE, P. Manual de análise de dados: Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®. 1. ed. Rio de Janeiro: Elsevier, 2017, p. 1219.