

Aula 10: Que factores explicam a variação nos salários na organização? (II)

Validação e Interpretação de Modelos de Regressão Linear

Docente: Amílcar Moreira Data & Hora: 15/12/2020, 20:30-22:30 Local: FRANCESINHAS 2, Sala 101



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Na Aula Anterior
 - Aprendemos a implementar e interpretar o resultado de um modelo de regressão linear
- Objetivos da Aula
 - Parte Teórica
 - Perceber qual a necessidade de fazermos diagnósticos aos pressupostos do nosso modelo de regressão
 - Saber quais são os pressupostos do modelo de regressão linear
 - Saber como, com a ajuda de gráficos e testes estatísticos, podemos conferir se os pressupostos do modelo estão a ser cumpridos
 - Parte Prática
 - Saber implementar os diagnósticos de um modelo no SPSS



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Já sabemos implementar um modelo de regressão linear...

Já sabemos interpreter os resultados do modelo de regressão linear...

O que é que nos falta?

📄 *Saíd	a1 [Docu	mento1] - Visua	alizador	do IBM SF	SS Statis	stics										-		\times
rquivo	<u>E</u> ditar	<u>V</u> isualizar	Dados	Trans	formar	Inserir	F <u>o</u> rmatar	Analisar	<u>G</u> ráfic	os <u>U</u> tilit	ários	Extensões	Janela	Ajuda				
						~	i				1							
∎ <mark>₪</mark> S [aída P Log P Log		•	Regr	essão)												*
Ġ (E Regre	ssão ítulo		`	/ariáve	eis Inse	ridas/Rem	ovidas ^a										
		bservações			Va	riáveis	Variáve	is										
		ariáveis Inseri esumo do mo	da de	Modelo) in: evalu	seridas	removi	las M	erir									
		NOVA			expe	rience		. 113										
		o en ciente o		a. V	ariável D	ependen	te: y_wage2											
				b. T	odas as	variáveis	solicitadas ir	iseridas.										
				Bosumo do modelo														
							esumo ao	R quadra	do E	ro padrão								
				Modelo) F	R	quadrado	ajustad	o da	estimativa	1							
				1	9.	998 ^a	.997		997	473.1935	6							
				a. P	reditores	: (Consta	ante), evaluati	on, experie	nce									
									, a									
								ANOV	A									
				Modelo)		Soma dos Quadrados	df	Qua	idrado édio	z	2	Sig.					
				1	Regre	ssão	3.202E+11		2 1.	01E+11	71490	3.694	.000 ^b					
					Residu	uo	1087093458	485	5 223	912.144								
				a V	Total ariável D	enenden	3.212E+11	485	/									
				b.P	reditores	: (Consta	ante), evaluati	on, experie	nce									
								Coefici	entes ^a									
							Confisiont	a fa nadro-	izadas	Coeficien	tes							
				Model			B	Erro	Erro	padroniza Beta	uus	t	Sig.					
			•	modelle								-						•
										0	process	ador do IBN	I SPSS Statis	stics está	pronto	U	nicode:C	IN



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Saber se o modelo cumpre os pressupostos que legitimam que os seus resultados podem ser considerados estatísticamente significativos

Quais são estes presupostos?



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Pressupostos do modelo de regressão

	Pressupostos	Definição	Forma de Validação
I	Linearidade	O efeito das variáveis independentes na variável dependente é linear e aditivo.	VER ACIMA
II	Normalidade da Distribuição dos Erros	Os erros seguem uma distribuição normal.	Análise de ResíduosGráfico de Q-Q
111	Média Condicional Zero dos Erros	O termo de erro aleatório tem valor esperado igual a zero.	Análise de Resíduos
IV	Homocedasticidade (ou Igual Variância)	A distribuição dos erros apresenta uma variância constante (hipótese da homocedasticidade).	Análise de Resíduos
V	Independência dos Erros	Os erros não estão correlacionados, i.e., o valor de um erro não depende de qualquer outro erro.	Dublin-Watson
VI	Ausência de multicolinearidade perfeita	As variáveis independents não estão perfeitamente correlacionadas entre si.	 Diagnósticos de Colinearidade
VII	Ausência de Observações Influentes	Não existem observações que tenham uma influência anormal nos resultados do modelo.	Cook's Distance



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

E qual é o problema se estes pressupostos não se verificarem?

 Os intervalos de confiança ou os p-values podem estar a ser subestimados (i.e. mais pequenos do que na realidade são) ...

ou seja: estamos a atribuir significância estatística a uma estimativa que na realidade não a terá!



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Como podemos saber se estes pressupostos estão a ser cumpridos?



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

- 1. Estimar o modelo de regressão com os diagnósticos
- 2. Avaliação do Pressuposto II: Normalidade da Distribuição dos Erros
- 3. Avaliação do Pressuposto III: Média Condicional Zero dos Erros
- 4. Avaliação do Pressuposto IV: Homocedasticidade
- 4. Avaliação do Pressuposto V: Independência dos Erros
- 5. Avaliação do Pressuposto VI: Ausência de Multicolinearidade Perfeita 6. Avaliação do Pressuposto VII: Ausência de Observações Influentes



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

1. Estimar o modelo de regressão com os diagnósticos



(A)

(B)

(C)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Dependente'
 - **Exercício:** Colocar as variáveis 'sex', 'experience' e 'evaluation' na caixa 'Independente(s)'





(A)

B

(C**)**

(D)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Dependente'
 - Exercício: Colocar as variáveis 'sex', 'experience' e 'evaluation' na caixa 'Independente(s)'
- Selecionar botão 'Estatísticas'

					Analisar	Grancos		Extensoe	s Janeia	Ajuda			
			• -	Kegressao Li	near					>	6	Visível	24 de 24 varis
		🐓 year	🗞 sex	🖋 year	4	•	Dependente:			E <u>s</u> tatísticas Gráficos	artme ∢	hrs_week	learing
1		2018.00	Fema	📣 age		BIOCO	1 de 1			Salvar	arketing	38.85	N
2		2018.00	Fema	education		Ante	rior		Próximo	Oncões	nancial	39.18	N
3		2018.00	Fema	Start_yr			Independent	e(s):			gistics	40.24	N
4		2018.00	Fema	<pre>stan_age experience</pre>			💑 sex			Estijo	duction	38.62	N
5		2018.00	Ma	y_wage		•	education	۱ م		<u>B</u> ootstrap	ogistics	38.86	N
6		2018.00	Fema	뤚 departmen	t		- experience				Admin	41.00	N
7		2018.00	Ma	hrs_week			Mét	odo: Inseri	ir 🔻		duction	38.90	N
8		2018.00	Ma	skin			Variával da ao	10000			Sales	41.70	N
9		2018.00	Ma	a back		•	variav <u>e</u> i de se	ieçao.	Regra		duction	40.95	N
10		2018.00	Fema	💰 muscle_up	per		Dátulos do co	0.01			nancial	41.07	N
11		2018.00	Fema	Ruscle_lo	wer	-	Rotulos de <u>c</u> a	50.			gistics	40.61	N
12		2018.00	Fema	neadache			Pandaração V	WI C:		1	duction	41.16	N
13		2018.00	Fema	anxiety	-	•	Fonderação V	VL3.			duction	39.90	N
14		2018.00	Ma								Admin	40.59	N
15	4	2018-00	Fema		к	Colar	Reconfigur	ar Cano	elar	Ajuda	prietice	40.05	N
Visualiza	ção de	dados Vis	ualização de	variável			***				_		



(A)

B

(C**)**

(D)

(E)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Regressão' / 'Linear'
- Selecionar a variável 'y_wage2'
- Colocar na caixa 'Dependente'
 - Exercício: Colocar as variáveis 'sex', 'experience' e 'evaluation' na caixa 'Independente(s)'
- Selecionar botão 'Estatísticas'
- Selecionar 'Estimativas'

			Regress	ão Linear			ı 🔤	1		
					adapta:	_			Visível:	24 de 24 va
	🛷 year	🗞 sex	🛷 year	ressão Linear: estatísticas	idenie.	×	ísticas ficos	partme nt	🔗 hrs_week	🗞 hearii
1	2018.00	Fema	📣 age	Coeficientes de regressão —	Ajuste do <u>m</u> odelo		Ivar	irketing	38.85	
2	2018.00	Fema	edur	✓ Estimativas	📃 Alteração de R quadra	do	tões	nancial	39.18	
3	2018.00	Fema	start	Intervalos de confiança	Descritivos		tile	gistics	40.24	
4	2018.00	Fema	expe	Nível (%): 95	📃 Correlações <u>p</u> arciais e	de parte	uio	duction	38.62	
5	2018.00	Ma	✓ y_wa	Matriz de covariâncias	🗹 Diagnósticos de co <u>l</u> ine	aridade	strap	gistics	38.86	
6	2018.00	Fema	💑 depa	r Residuais				Admin	41.00	
7	2018.00	Ma	hrs_	Durbin Wataan				duction	38.90	
8	2018.00	Ma	a skin	Diagnástico por coco				Sales	41.70	
9	2018.00	Ma	and back	Valores discrepantes no l	ado de fora: a desv	os nadrão		duction	40.95	
10	2018.00	Fema	🚴 mus	Tedas as ssess		ios padrao		nancial	41.07	
11	2018.00	Fema	💑 mus					gistics	40.61	
12	2018.00	Fema	a head	Continuar	Cancelar			duction	41.16	
13	2018.00	Fema	anxier		Allocial	_		duction	39.90	
14	2018.00	Ma	C					Admin	40.59	
15	2018.00	Fema		OK Colar <u>R</u> e	configurar Cancelar	Ajuda		adistics	40.05	
15	1	Fema						Interiore	40.05	



(F)

G

(H)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Ajuste do modelo'
- Selecionar 'Descritivos'
- Selecionar 'Diagósticos de colinearidade'
- Selecionar 'Dublin-Watson'
- Selecionar 'Continuar'

				Regressão I	inear						×			
					2	_	Denendente	a.		6			Visível: 2	24 de 24 variá
		🔗 year	🗞 sex	🛷 year 🐚	essão Li	near: estatís	ticas				ticas	partme	🔗 hrs_week	🗞 hearing
1		2018.00	Fema	age C	oeficientes o	de regressã	io — 🔽 A	juste do <u>m</u> ode	lo		ar	rketing	38.85	N
2		2018.00	Fema	educ	🗸 <u>E</u> stimativ	/as		Iteração de R o	uadrado		00	nancial	39.18	N
3		2018.00	Fema	start	Intervalos	s de confian	nça 🔽 🖸	escritivos			es	gistics	40.24	N
4		2018.00	Fema	expe	Nivel (%):	95	🗖 C	orrelações <u>p</u> a	ciais e de p	arte	0	duction	38.62	N
5		2018.00	Ma	y_w	🔲 Matriz de	covariância	as 🔽 🖸	iagnósticos de	colinearida	ade S	trap	gistics	38.86	N
6		2018.00	Fema	💑 dep:	esiduais -							Admin	41.00	N
7		2018.00	Ma	hre	Durbie We							duction	38.90	N
8		2018.00	Ma		Diagnósti	alson						Sales	41.70	N
9		2018.00	Ma	& back		discrepante	e no lado de	fora: 2	desvios n	adrão		duction	40.95	N
10		2018.00	Fema	💫 mus	Todas a	uiscrepante	55 110 1800 U	s 101a. 3	desvios p	aurao		nancial	41.07	N
11		2018.00	Fema	💑 mus	@ 10005 0	IS Casus						gistics	40.61	N
12		2018.00	Fema	a head	\frown	Continu	ar Cancel	ar Aiuda		_		duction	41.16	N
13		2018.00	Fema	anxier-	[H]_	Contanta	Cancer	an Aluga				duction	39.90	N
14		2018.00	Ma									Admin	40.59	N
15		2018 00	Fema		ок	Colar	Reconfig	urar Cance	elar	Ajuda		naistics	40.05	N
	4													



 \bigcirc

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Diagnósticos

• Selecionar botão 'Salvar'

ta ADG	RH_BD	_Aula10.sav [Co	onjuntodeDa	dos1] - Editor de dad	los do IBI	VI SPSS Stat	istics					_		\times
Arquivo	<u>E</u> dita	ar <u>V</u> isualiza	r <u>D</u> ados	<u>T</u> ransformar <u>A</u>	nalisar	<u>G</u> ráficos	<u>U</u> tilitários	E <u>x</u> tensões	Janela	Ajuda				
				Regressão Linea	r	and the second				×				
							Dependente:					Visível:	24 de 24 vari	áveis
		🛷 year	🗞 sex	🛷 year 💰 sex		- Place 1	✓ y_wage2			Estatisticas Gráficos	(\cup)	⁹ hrs_week	🗞 hearing	J
1		2018.00	Fema	💉 age		BIOCO I	del	-		S <u>a</u> lvar	keting	38.85	N	io 📥
2		2018.00	Fema	education		Anteri	or		Próximo	Opcões	nancial	39.18	N	io 🗌
3		2018.00	Fema	start_yr			Independent	e(s):		Estilo	ogistics	40.24	N	lo
4		2018.00	Fema	experience			Sex ducation			Esu <u>i</u> o	duction	38.62	N	lo
5		2018.00	Ma	💉 y_wage		*	education	i ce		Bootstrap	ogistics	38.86	N	lo
6		2018.00	Fema	lepartment		L	· · · · · ·				Admin	41.00	N	lo
7		2018.00	Ma	hrs_week			<u>M</u> ét	odo: Inserir	~		duction	38.90	N	lo
8		2018.00	Ma	al skin			/ariável de se	lecão:			Sales	41.70	N	lo
9		2018.00	Ma	💑 back		🔶 🔶	ranav <u>e</u> r ac se	icçuo.	Regra		duction	40.95	N	lo
10		2018.00	Fema	Ruscle_upper	r 📕		Rótulos de ca	so:			nancial	41.07	N	lo
11		2018.00	Fema	muscle_lower		🔶 🤟 j	<u>.</u> .				ogistics	40.61	N	lo
12		2018.00	Fema	anjury			Ponderação V	VLS:			duction	41.16	N	o
13		2018.00	Fema	💰 anxiety	-	₩ [duction	39.90	N	o
14		2018.00	Ma	OK		Color	Beconfigur	or Concel		iudo	Admin	40.59	N	
15	1	2018 00	Fema			Colai	Reconligur	Cancel	an A	Juua	naistics	40.05	N	
Visualia	zação	le dados Visu	ualização de	variável			***							
								O processad	or do IBM SI	PSS Statistics est	á pronto	Unicode	ON	



 \bigcirc

 (\mathbf{J})

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar botão 'Salvar'
- Selecionar 'Padronizado'

Valores preditos	- Residuais)
Nao padronizado	Nao padronizado	
Padronizado	Padronizado	
Ajustado	Estudentização	
S.E. de predições medias	Excluido	
	Estudentizado excluído	
Distâncias	Estatísticas de influência	٦
Ma <u>h</u> alanobis	Df <u>B</u> eta(s)	
🕅 de Coo <u>k</u>	DfBeta(s) padronizado(s)	
📃 Valores de ponto alavanca	DfFit	
Intervalos de predição	DfFit padronizado	
Média 🔲 Individual	🔲 Razão de co <u>v</u> ariância	
Intervalo de <u>c</u> onfiança: 95 %		
Estatísticas de coeficiente		7
📃 Criar estatísticas de c <u>o</u> eficiente		
Oriar novo conjunto de dados		
Nome do conjunto de dados:		
O Gravar um novo arquivo de dados		
Arquivo		
Exportar informações do modelo para o ar	quivo XML	-
	Navegar	
Incluir a matriz de covariâncias		



 \bigcirc

(J)

(K)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar botão 'Salvar'
- Selecionar 'Padronizado'
- Selecionar 'de Cook' e 'Valores de ponto alavanca'

Valores preditos	Residuais
Não padronizado	Não padronizado
Padronizado	Padronizado
Ajustado	Estudentização
S.E. de predições médias	Excluído
	<u>E</u> studentizado excluído
Distâncias	Estatísticas de influência
Mahalanobis	Df <u>B</u> eta(s)
de Cook	DfBeta(s) padronizado(s)
Valores de ponto alavanca	D <u>f</u> Fit
Intervalos de predição	DfFit padronizado
Média I Individual	Razão de co <u>v</u> ariância
Intervalo de <u>c</u> onfiança: 95 %	
Estatísticas de coeficiente	
Criar estatísticas de coeficiente	
Criar novo conjunto de dados	
Nome do conjunto de dados:	
Gravar um novo arquivo de dados	
Exportar informações do modelo para o a	arquivo XML
	Navegar
Incluir a matriz de covariâncias	



(1)

(к)

(M)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar botão 'Salvar'
- Selecionar 'Padronizado'
- Selecionar 'de Cook' e 'Valores de ponto alavanca'
- Selecionar 'DfBeta(s) padronizado(s)
- Selecionar 'Continuar'/ 'OK'





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Diagnósticos

• Os resultados são publicados no 'Visualizador de Resultados'

te *Saída1 [l	Documento1] - Vis	ualizador do	o IBM SPSS Stat	istics										-		>
Arquivo Ed	ditar <u>V</u> isualizar	<u>D</u> ados	Transformar	Inserir	F <u>o</u> rmatar	Analisa	<u>G</u> ráficos	<u>U</u> tilitários	Extensõ	ies Janela	Ajuda					
😑 🖁		P	🛄 🗠	<u>a</u> [· 📥 =										
	a og Regressão Título Correlações Corre	scritiva ridas/Remo iodelo e colinearid e resíduos or *zpred Gr	ovidas Jade áfico de Dispe	REG /DE /MIS /ST /CF /NC /DE /ME /SC /RE /SA	RESSION SCRIPTIN SSING LIS ATISTICS NTERIA=F DORIGIN PENDENT THOD=EI ATTERP SIDUALS VE LEVEI	/ES MEAN STWISE COEFF C IN(.05) PC T y_wage2 NTER sex LOT=(*ZRE DURBIN R ZRESID	STDDEV C DUTS R ANO DUT(.10) education e ESID ,*ZPRE SDBETA.	CORR SIG N DVA COLLI experience ED)	N TOL							
					E	statística	Descritiva	a								
					_	Média	Erro Desv	rio N								
				y_w	age2 2	25405.6559	8132.604	05 485	8							
				sex		1.50	.5	00 485	8							
				edu	cation	3.00	1.4	11 485	8							
				expe	erience	10.6015	5.717	32 485	8							
							Co	orrelações]				
								y_wage2	sex	education	experience					
				Corr	elação de l	Pearson	y_wage2	1.000	250	.238	.224					
							sex	250	1.000	010	.014					
							education	.238	010	1.000	023					
							experience	.224	.014	023	1.000					
				Sig.	(1 extremic	iade)	y_wage2		.000	.000	.000					
							sex	.000		.245	.170					
							education	.000	.245		.054					
4							experience	.000	.170	.054	4050					
										Opr	ocessador do l	BM SPSS Statistic	cs está pronto	U	nicode:(N



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Diagnósticos

 Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis

ta *ADO	SRH_BD_Aula1	0.sav [Cor	njuntodeDados1] - Eo	ditor de dados do IBM S	PSS Statistics				- 0	×
Arquivo	<u>E</u> ditar <u>V</u>	isualizar	<u>D</u> ados <u>T</u> ransfo	ormar <u>A</u> nalisar <u>G</u> r	ráficos <u>U</u> tilitários I	E <u>x</u> tensões Janela	Ajuda			
	H 🖨	II ,	5	🖹 📥 🗐 🛛	M 🔜		•			
1 : ZRE_	1	-1.138	394522293957						Visível: 31 de 31 v	ariáveis
	🕫 eva	uation 2	SRE_1			SDB0_1	SDB1_1	SDB2_1	SDB3_1	
1		55.17	-1.13895	.00031	.00075	02008	.01582	01082	.02414	-
2		51.75	73316	.00009	.00043	00244	.01050	00751	00841	
3		54.79	.08094	.00000	.00032	.00054	00115	.00082	.00032	2
4		58.30	10802	.00000	.00110	00361	.00153	.00227	.00237	
5		49.89	-1.61943	.00061	.00073	.00170	02284	.03244	01683	3
6		52.74	.26865	.00001	.00061	.00066	00379	.00543	.00045	i
7		47.60	51252	.00005	.00059	.00244	00757	00507	.00848	3
8		52.65	-1.39629	.00044	.00069	00962	02039	.01471	.02727	
9		48.48	09011	.00000	.00036	00012	00130	.00092	.00063	3
10		52.40	09098	.00000	.00061	00032	.00128	00183	.00008	3
11		47.21	93018	.00018	.00064	02087	.01351	.01905	00301	
12		48.39	.34721	.00004	.00096	.00550	00511	00699	.00636	i
13		50.19	1.06687	.00039	.00117	.01455	01579	02136	.02492	2
14		40.86	1.27964	.00034	.00063	02269	.01872	.02621	.00163	3
15		45 36	98336	00013	00033	00845	- 01418	00006	01105	
Visualiz	tação de dado	visual	lização de variável			O proces	sador do IBM SPSS Stat	istics está pronto	Unicode:ON	



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
 - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE_1) para cada observação

tal *ADGRH	*ADGRH_BD_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics — □ × Arquivo Editar Visualizar Dados Transformar Analisar Gráficos Utilitários Extensões Janela Ajuda														
Arquivo E	ditar <u>V</u> isualiza	r <u>D</u> ados <u>T</u> ransfo	rmar <u>A</u> nalisar <u>G</u> i	áficos <u>U</u> tilitários	E <u>x</u> tensões Janela	Ajuda									
🔁 H		F A	🖹 📥 🗐 🛛	H 4		•									
1: ZRE_1	1: ZRE_1 -1.13894522293957 Visível: 30 de 30 variáve														
	evaluation # ZRE_1 # LEV_1 # SDB0_1 # SDB1_1 # SDB2_1 # SDB3_1 var 1 55.17 Name: ZRE_1 5 - 02008 01582 - 01082 02414 #														
1	1 55.17 Nome: ZRE_1 5 02008 .01582 01082 .02414 2 54.75 Rétulo: Standardized Residual 2 00244 0450 02374 00244														
2	2 51.75 Rótulo: Standardized Residual 300244 .010500075100841														
3	3 54.79 Tipo: Numérico Medida: Escala 2 .00054 00115 .00082 .00032														
4	3 34.19 Medida: Escala 2 .00034 00113 .00022 .00032 4 58.30 10802 .00110 00361 .00153 .00227 .00237														
5	49.89	-1.61943	.00073	.00170	02284	.03244	01683								
6	6 52.74 .26865 .00061 .00066 00379 .00543 .00045														
7	0 32.14 2000 .0001 .0000 00515 .00045 .00045 7 47.60 51252 .00059 .00244 00757 00507 .00848														
8	52.65	-1.39629	.00069	00962	02039	.01471	.02727								
9	48.48	09011	.00036	00012	00130	.00092	.00063								
10	52.40	09098	.00061	00032	.00128	00183	.00008								
11	47.21	93018	.00064	02087	.01351	.01905	00301								
12	48.39	.34721	.00096	.00550	00511	00699	.00636								
13	50.19	1.06687	.00117	.01455	01579	02136	.02492								
14	40.86	1.27964	.00063	02269	.01872	.02621	.00163								
15	45 36	98336	00033	00845	- 01418	00006	01105	~							
Visualizaç	ão de dados Visi	ualização de variável													
				0	processador do IBM SF	SS Statistics está pron	to Unicode:ON								



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
 - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE_1) para cada observação
 - Uma variável que mede a distancia de Cook associada a cada observação (COO_1)

ta *ADO	*ADGRH_BD_Aula10.sav [ConjuntodeDados1] - Editor de dados do IBM SPSS Statistics - - × 'quivo Editar Visualizar Dados Iransformar Analisar Gráficos Utilitários Extensões Janela														
Arquivo	<u>E</u> ditar <u>V</u> isualizar	Dados <u>T</u> ransform	nar <u>A</u> nalisar <u>G</u> rá	ficos <u>U</u> tilitários E	<u>x</u> tensões Janela	Ajuda									
	🖨 🛄	ା 🗠 🔁 📱	ă 📥 🗐 🛽	a 🗛 🔚											
1:CO0_	: COO_1 0.00031099841585 Visível: 31 de 31 variáveis														
	Image: state weaking and state														
1	55.17	-1.13895	.00031	.00075	02008	.01582	01082	.02414							
2	51.75	73316	.00009	.00043	00244	.01050	00751	00841							
3	3 54.79														
4	58.30	10802	.00000	.00110	00361	.00153	.00227	.00237							
5	49.89	-1.61943	.00061	.00073	.00170	02284	.03244	01683							
6	52.74	.26865	.00001	.00061	.00066	00379	.00543	.00045							
7	47.60	51252	.00005	.00059	.00244	00757	00507	.00848							
8	52.65	-1.39629	.00044	.00069	00962	02039	.01471	.02727							
9	48.48	09011	.00000	.00036	00012	00130	.00092	.00063							
10	52.40	09098	.00000	.00061	00032	.00128	00183	.00008							
11	47.21	93018	.00018	.00064	02087	.01351	.01905	00301							
12	48.39	.34721	.00004	.00096	.00550	00511	00699	.00636							
13	50.19	1.06687	.00039	.00117	.01455	01579	02136	.02492							
14	40.86	1.27964	.00034	.00063	02269	.01872	.02621	.00163							
15	45 36	98336	00013	00033	00845	- 01418	00006	01105							
Visualiz	Visualização de variável														



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
 - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE_1) para cada observação
 - Uma variável que mede a distancia de Cook associada a cada observação (COO_1)
 - Uma variável que mede influência relativa de cada observação no ajuste do modelo (LEV_1).

ta *ADO	GRH_BD_A	ula10.sav [Cor	njuntodeDados1] - Ed	itor de dados do IBM S	PSS Statistics				- 0	×
Arquivo	<u>E</u> ditar	<u>V</u> isualizar	Dados <u>T</u> ransfo	rmar <u>A</u> nalisar <u>G</u>	ráficos <u>U</u> tilitários	E <u>x</u> tensões Janela	Ajuda			
				ă 📥 🗐 🛛			•			
1:LEV_1	I	.0007	5130556031						Visível: 31 de 31 v	ariáveis
	11	evaluation 2	SRE_1		√ ^P LEV_1			SDB2_1	SDB3_1	
1		55.17	-1.13895	.0003	.00075	02008	.01582	01082	.02414	4
2		51.75	73316	.0000	.00043	00244	.01050	00751	00841	1
3		54.79	.08094	.0000	.00032	.00054	00115	.00082	.00032	2
4		58.30	10802	.0000	.00110	00361	.00153	.00227	.00237	7
5		49.89	-1.61943	.0006	.00073	.00170	02284	.03244	01683	3
6		52.74	.26865	.0000	.00061	.00066	00379	.00543	.00045	5
7		47.60	51252	.0000	.00059	.00244	00757	00507	.00848	3
8		52.65	-1.39629	.00044	.00069	00962	02039	.01471	.02727	7
9		48.48	09011	.0000	.00036	00012	00130	.00092	.00063	3
10		52.40	09098	.0000	.00061	00032	.00128	00183	.0000	3
11		47.21	93018	.00018	.00064	02087	.01351	.01905	00301	1
12		48.39	.34721	.00004	.00096	.00550	00511	00699	.00636	5
13		50.19	1.06687	.0003	.00117	.01455	01579	02136	.02492	2
14		40.86	1.27964	.00034	.00063	02269	.01872	.02621	.00163	3
15	4	45 36	98336	0001	00033	00845	- 01418	00006	01106	5
Visualiz	ação de o	lados Visual	lização de variável			O proces	sador do IBM SPSS Stat	istics está pronto	Linicode:ON	



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Quando instruímos o SPSS para produzir os diagnósticos, é criado um conjunto variáveis
 - Uma variável com os 'Resíduos Padronizados' da variável dependente (ZRE_1) para cada observação
 - Uma variável que mede a distancia de Cook associada a cada observação (COO_1)
 - Uma variável que mede influência relativa de cada observação no ajuste do modelo (LEV_1).
 - Por cada variável independente é criada uma variável com os DFBETA Padronizado, mede a influência de uma dada observação na estimação dos parâmetros.

ta *ADC	GRH_BD_Aula	10.sav [Cor	njuntodeDados1] - Ec	litor de dados do IBM S	PSS Statistics				- 0	×
Arquivo	<u>E</u> ditar <u>V</u>	(isualizar	<u>D</u> ados <u>T</u> ransfo	rmar <u>A</u> nalisar <u>G</u> r	áficos <u>U</u> tilitários E	<u>x</u> tensões Janela	Ajuda			
	88			ř 🏪 🗐 🛽			•			
1:SDB1	_1	.0158	2242634099						Visível: 31 de 31 va	ariáveis
	🔊 eva	luation 2	SRE_1		& LEV_1		SDB1_1	SDB2_1	SDB3_1	
1		55.17	-1.13895	.00031	.00075	02008	.01582	01082	.02414	
2		51.75	73316	.00009	.00043	00244	.01050	00751	00841	
3		54.79	.08094	.00000	.00032	.00054	00115	.00082	.00032	
4		58.30	10802	.00000	.00110	00361	.00153	.00227	.00237	
5		49.89	-1.61943	.00061	.00073	.00170	02284	.03244	01683	
6		52.74	.26865	.00001	.00061	.00066	00379	.00543	.00045	
7		47.60	51252	.00005	.00059	.00244	00757	00507	.00848	
8		52.65	-1.39629	.00044	.00069	00962	02039	.01471	.02727	
9		48.48	09011	.00000	.00036	00012	00130	.00092	.00063	
10		52.40	09098	.00000	.00061	00032	.00128	00183	.00008	
11		47.21	93018	.00018	.00064	02087	.01351	.01905	00301	
12		48.39	.34721	.00004	.00096	.00550	00511	00699	.00636	
13		50.19	1.06687	.00039	.00117	.01455	01579	02136	.02492	
14		40.86	1.27964	.00034	.00063	02269	.01872	.02621	.00163	
15	4	45 36	98336	00013	00033	00845	- 01418	00006	01105	
Visualiz	ação de dad	os Visual	lização de variável			O proces	sador do IBM SPSS Stat	istics está pronto	Unicode:ON	



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

2. Avaliação do Pressuposto II: Normalidade da Distribuição dos Erros



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Para avaliarmos se os erros seguem uma distribuição normal, vamos usar a variável com os 'Resíduos Padronizados' da VD (ZRE_1) que acabamos de criar.
- Vamos então criar usar um gráfico Q-Q para representar a distribuição dos resíduos padronizados

ta *ADG	RH_BD_Aula10	.sav [Cor	njuntodeDa	ados1] - I	Editor de dados do	IBM SPSS	Statistics					—		×
Arquivo	<u>E</u> ditar <u>V</u> is	ualizar	<u>D</u> ados	Transt	formar <u>A</u> nalisa	r <u>G</u> ráfic	os <u>U</u> tilitários	E <u>x</u> tensões	Janela	Ajuda				
					E 🛓	ĸ	H		A 14					
1: ZRE_1		-1.138	394522293	957							١	/isível: 30	de 30 va	riáveis
	evalua 2	ation	🔗 ZRI	E_1	LEV_1		🔗 SDB0_1	💉 S	DB1_1	SDB2_1	🔗 SDI	B3_1	Va	ar
1	5	5.17	1	Nome: Zl	RE_1	5	020	08	.01582	01	082	.02414		
2	5	1.75	1	Rótulo: S	tandardized Resi	tual 3	002	44	.01050	00	751	00841		
3	5	4.79		Medida: E	Escala	2	.000	54	00115	.00	082	.00032	!	
4	5	8.30	L	1080/	<u>د</u> .	0110	003	61	.00153	.00	227	.00237		
5	4	9.89		-1.61943	3	00073	.001	70	02284	.03	244	01683	1	
6	5	2.74		.2686	5	00061	.000	66	00379	.00	543	.00045	i	
7	4	7.60		51252	2 .	00059	.002	44	00757	00	507	.00848	1	
8	5	2.65		-1.3962	9	00069	009	62	02039	.014	471	.02727	·	
9	4	8.48		0901	1 .	00036	000	12	00130	.00	092	.00063)	
10	5	2.40		09098	3	00061	000	32	.00128	00	183	.00008	1	_
11	4	7.21		93018	3	00064	020	87	.01351	.019	905	00301		
12	4	8.39		.3472	1	00096	.005	50	00511	00	599	.00636	ļ	
13	5	0.19		1.0668		00117	.014	55	01579	02	136	.02492	:	_
14	4	0.86		1.27964	4.	00063	022	69	.01872	.02	521	.00163	(
15	4	5 36		9833	5	00033	008	45	- 01418	00	106	01105		
Visualiza	ição de dados	Visual	lização de	variável			***							
								O processa	dor do IBM SF	PSS Statistics está	pronto U	nicode:ON	1	



A

(B)

(C)

(D)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Gráficos Q-Q'
- Selecionar a variável 'ZRE_1'
- Colocar na caixa 'Variáveis'
- Selecionar 'OK'







Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Gráficos Q-Q'
- Selecionar a variável 'ZRE_1'
- Colocar na caixa 'Variáveis'
- Selecionar 'OK'





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Linha diagonal reflecte uma distribuição normal
- Os resíduos sobrepõe-se quase totalmente com a linha de diagonal
- Os resíduos parecem estar normalmente distribuídos
- Neste, caso os as caudas da distribuição dos resíduos afasta-se da diagonal, o que sugere que a distribuição dos erros pode não ser normal





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- O gráfico é publicado no 'Visualizador de Resultados'
- Neste caso podemos concluir que os erros seguem uma distribuição normal!





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

3. Avaliação do Pressuposto III: Média Condicional Zero dos Erros



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Para avaliarmos se o termo de erro aleatório tem valor esperado igual a zero, vamos usar a variável com os 'Resíduos Padronizados' da VD (ZRE_1) que acabamos de criar.
- Mas neste caso, vamos olhar para as estatísticas descritivas desta variável.

				00 000000				
Arquivo E	ditar <u>V</u> isualiza	r <u>D</u> ados <u>T</u> ransfo	ormar <u>A</u> nalisar <u>G</u> rá	ficos <u>U</u> tilitários	E <u>x</u> tensões Janela	Ajuda		
📄 H		r a	🖹 📥 🗐 🛽			•		
1: ZRE_1	-1.1	3894522293957					Visível: 30 d	e 30 variáveis
	evaluation 2	SRE_1	LEV_1		SDB1_1	SDB2_1	SDB3_1	var
1	55.17	Nome: ZR	E_1 5	02008	.01582	01082	.02414	
2	51.75	Rótulo: Sta	andardized Residual 3	00244	.01050	00751	00841	
3	54.79	Medida: E	scala 2	.00054	00115	.00082	.00032	
4	58.30	10802	01100.	00361	.00153	.00227	.00237	
5	49.89	-1.61943	.00073	.00170	02284	.03244	01683	
6	52.74	.26865	.00061	.00066	00379	.00543	.00045	
7	47.60	51252	.00059	.00244	00757	00507	.00848	
8	52.65	-1.39629	.00069	00962	02039	.01471	.02727	
9	48.48	09011	.00036	00012	00130	.00092	.00063	
10	52.40	09098	.00061	00032	.00128	00183	.00008	
11	47.21	93018	.00064	02087	.01351	.01905	00301	
12	48.39	.34721	.00096	.00550	00511	00699	.00636	
13	50.19	1.06687	.00117	.01455	01579	02136	.02492	
14	40.86	1.27964	.00063	02269	.01872	.02621	.00163	
15	45 36	98336	00033	00845	- 01418	00006	01105	v
Visualizaçã	io de dados Visu	ualização de variável			processador do IBM SF	PSS Statistics está pror	nto Unicode:ON	



(A)

(B)

(c)

(D)

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Descritivos'
- Selecionar a variável 'ZRE_1'
- Colocar na caixa 'Variável(is)'
- Selecionar 'OK'

	Editar Visualizar L		Relatórios	Extensoes Janeia	Ajuda	<u>–</u> † (A	
1:ZRE 1	-1.13894	522293957	Estatística Descritiva	Frequên	cias		
		valuation & ZRE	Tabelas	Descritiv	0S)B3 1	
	· · · · · · · · · · · · · · · · · · ·	2	– Co <u>m</u> parar Médias	Explorar.			
1	17529.60	55.17 -1.	Modelo Linear <u>G</u> eral	I abeia d	le reteren <u>c</u> ia cruzada	.024 4	
2	24843.10	51.75 -	Modelos lineares generalizados	Razao		008.1	
3	29874.72	54.79	Modelos mistos	Graticos	<u>E-P</u>	.000 2	
5	10044 01	49.89 -1	Correlacionar	Graticos	<u>Q</u> -Q	- 016 33	
6	32321.53	52.74	<u>R</u> egressão	0379	.00543	93045	
ta *ADG	GRH_BD_Aula10.sav [Co	onjuntodeDados1] - E	ditor de dados do IBM SPSS Stati	stics			— 🗆
Arquivo	<u>E</u> ditar <u>V</u> isualizar	r <u>D</u> ados <u>T</u> ransf	ormar <u>A</u> nalisar <u>G</u> ráficos	Utilitários Exten	ısões Janela Ajud	а	
		 • •				1	
	▤ ▤ Щ) ר או	le 🔚 📰 💵 🕯	n 📰 🎞	1 1 🖌 🕗 💽		
2 : SDB3_	_1008	841231486353					Visível: 31 de 31 va
	evaluation	🔗 ZRE 1 👝	A COO 1	VIEV 1	SDB0_1	SDB1 1	SDB2 1
	evaluation 2	✓ ZRE_1	Descritivas	FV 1	SDB0 1	Ø_SDB1_1 ×	SDB2_1
1	evaluation 2 55.17	✓ ZRE_1 -1.138	Descritivas	PIEV 1	SDB0_1	SDB1_1 × .01582	& SDB2_1 01082
1	evaluation 2 55.17 51.75	√ ² ZRE_1 -1.13 √ 3 √ 3 ✓ 3	Descritivas	Veriável(is):	SDB0 1	SDB1_1 X .01582 .01050	SDB2_10108200751
1 2 3	evaluation 2 55.17 51.75 54.79	-1.138	COO 1 Descritivas y_wage2	Variável(is):	SDB0_1	SDB1_1 X .01582 .01050 00115	SDB2_10108200751 .00082
1 2 3 4	 ✓ evaluation 2 55.17 51.75 54.79 58.30 	-1.13 (B) 100	COO 1 Descritivas y_wage2 Consolout2 Standardized Re.	Variável(is):	SDB0_1	× .01582 .01050 .00115 .00153	SDB2_10108200751 .00082 .00227
1 2 3 4 5	✓ evaluation 2 55.17 51.75 54.79 58.30 49.89	ZRE_1 -1.13 -1.13 -1.13 -1.13 -1.13 -1.01 -1.01 -1.01	Coold	Variável(is):	SDB0_1	× 0.01582 	SDB2_10108200751 .00082 .00227 .03244
1 2 3 4 5 6	evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74	ZRE_1 -1.13 73 8 73 73 1.161 -1.615 .265	Coold	Variável(is):	SDR0_1	× .001582 	SDB2_1 .01082 .00751 .00082 .00227 .03244 .00543
1 2 3 4 5 6 7	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 	ZRE_1 -1.138 731 731 733 -1.616 -1.616512	Control Contr	Variável(is):	SDR0_1	× .001582 	✓ SDB2_1 01082 00751 .00082 .00227 .03244 .00543 00507
1 2 3 4 5 6 7 8	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 	ZRE_1 -1.138 -1.138 -1.6185121.396	Cool Cool Control Cool Cool Cool Cool Cool Cool Cool C	Variável(is):	SDR0_1	× .01582 	
1 2 3 4 5 6 7 8 9	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 	✓ ZRE_1 1.1.13 1.1.13 7.31 0.100 1.1.610 1.610 .1.610	Coo 1 Descritivas y_wage2 Standardized Re Cooke Distance 1 Cooke Distance 1 Cooke Distance 1 Standardized DF	Variável(is):	SDB0_1	 COLORED COLORED	✓ SDB2_1 01082 00751 .00082 .00227 .03244 .00543 00507 .01471 .00092
1 2 3 4 5 6 7 8 9 9	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 52.40 	✓ ZRE_1 1.1.13 1.1.13 7.3 8.1 8.1 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.08 1.09 1.09 1.09 1.09	Coon 1 Descritivas y_wage2 crandonovi2 Standardized Re Centered Levera Standardized DF Standardized DF.	Variável(is):	SDB0_1	× .01582 .01582 .01050 00115 .00153 02284 00379 00757 02039 00130 .00128	✓ SDB2_1 01082 00751 .00082 .00227 .03244 .00543 .00507 .01471 .00092 .00183
1 2 3 4 5 6 7 8 9 9 10 11	✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 52.40 47.21	✓ ZRE_1 1.1.13 1.1.13 7.31 8.1 1.108	Coo 1 Descritivas y_wage2 Standardized Re Contered Levera Standardized DF Standardiz	Variável(is):	SDB0_1	× .01582 .01050 .00153 .00153 .00153 .00284 .00379 .00757 .022039 .00130 .00128 .01351	✓ SDB2_1 01082 00751 .00082 .00227 .03244 .00543 .00507 .01471 .00092 .00183 .01905
1 2 3 4 5 6 7 8 9 10 11 11	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 52.40 47.21 48.39 	✓ ZRE_1 1.1.13 1.1.13 7.3 8.108 1.108 <p< td=""><td>Coo 1 Descritivas V_wage2 Control Control Standardized Re Control Levera Standardized DF Standardized D</td><td>Variável(is):</td><td>SDB0 1 Qpções Estilo. Bootstra Cancelar Ajuda</td><td>× .01582 .01050 .00153 .00153 .00153 .00284 .00379 .00757 .022039 .00130 .00128 .01351 .00511</td><td>✓ SDB2_1 01082 00751 .00082 .00227 .03244 .00543 00507 01471 00092 00183 01905 00699 </td></p<>	Coo 1 Descritivas V_wage2 Control Control Standardized Re Control Levera Standardized DF Standardized D	Variável(is):	SDB0 1 Qpções Estilo. Bootstra Cancelar Ajuda	× .01582 .01050 .00153 .00153 .00153 .00284 .00379 .00757 .022039 .00130 .00128 .01351 .00511	✓ SDB2_1 01082 00751 .00082 .00227 .03244 .00543 00507 01471 00092 00183 01905 00699
1 2 3 4 5 6 7 8 9 10 11 11 12 13	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 52.40 47.21 48.39 50.19 	✓ ZRE_1 1.133 1.133 0.108 1.018	Coold Descritivas y_wage2 cvaloadon2 Standardized Re Coold Distance I Coold Distance I Standardized DF Standardize	Variável(is):	SDB0_1	× .01582 .01050 .00153 .00153 .00284 .00379 .00757 .02284 .00379 .00757 .02039 .00130 .00128 .01351 .00511 .00511 .00511 .01579 .0157 .015 .0157 .0157 .0157 .0157 .0157 .015 .015 .015 .015 .015 .015 .015 .015	✓ SDB2_10108200751 .00082 .00227 .03244 .00543 .00507 .01471 .0009200183019050069902136
1 2 3 4 5 6 7 8 9 10 11 11 12 13 14	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 52.40 47.21 48.39 50.19 40.86 	✓ ZRE_1 -1.138 -1.138 -1.618 -1.618 -1.618612512512512512090090090930 .341 1.066881 1.27964	Coold Descritivas y_wage2 voidoatoriz Standardized Re Coold Distance I Coold Distance I Coold Distance I Standardized DF Standardized DF Standardized DF Standardized DF Standardized DF Coolar OK Coolar OK Coolar 00034 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 0003 000 000 000 00	Variável(is): Variável(is): como variáveis Reconfigurar	SDB0_1	× .001582 .01050 .00153 .00153 .002284 .00379 .00757 .02239 .00130 .00128 .01351 .00511 .00511 .01579 01872	✓ SDB2_10108200751 .00082 .00227 .03244 .00543 .00507 .01471 .00092 .00183 .01905 .00699 .02136 .02621
1 2 3 4 5 6 7 8 9 9 10 11 11 12 13 14	 ✓ evaluation 2 55.17 51.75 54.79 58.30 49.89 52.74 47.60 52.65 48.48 52.40 47.21 48.39 50.19 40.86 45.36 	✓ ZRE_1 -1.138 -1.138 -1.616616512512512090090090090930 .341 1.066857 1.27964 98336	Coold Descritivas	Variável(is): Variável(is): como variáveis Reconfigurar 00117 .00063 00033	✓ SDB0_1	× .01582 .01050 .00153 .00153 .00284 .00379 .00757 .02039 .00130 .00128 .01351 .00511 .00511 .00511 .01579 .01872 .01418	✓ SDB2_10108200751 .00082 .00227 .03244 .00543 .00507 .01471 .00092 .00183 .01905 .00699 .02136 .02621 .00006



(A**)**

B

 \bigcirc

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Analisar' / 'Estatística Descritiva' / 'Descritivos'
- Selecionar a variável 'ZRE_1'
- Colocar na caixa 'Variável(is)'
- Selecionar 'OK'

🔁 la	🖨 🔲	🗠 🔺 🦉	¥ = K	AA 🔚					
2 : SDB3_1	0084	1231486353						Visível: 31 de 31	variá
	evaluation 2	ZRE_1	Scritivas	🔗 FV 1	SDF	30 1 🔗 S X	DB1_1	SDB2_1	•
1	55.17	-1.138					.01582	01082	
2	51.75	731 ør e	valuation	Variável(is	i): Interfiere di Disse i d	Opções	.01050	00751	
3	54.79	.080 🛷 y	_wage2		lardized Resid	Estilo	00115	.00082	
4	58.30	108 🛷 e	valuation2			Bootstrap	.00153	.00227	
5	49.89	-1.61	ook's Distance [(Teoremotern)	02284	.03244	
6	52.74	.268	entered Levera tandardized DF	*			00379	.00543	
7	47.60	512 🛷 s	tandardized DF				00757	00507	
8	52.65	-1.390 🛷 S	tandardized DF				02039	.01471	
9	48.48	090 🛷 S	tandardized DF 🔽				00130	.00092	
10	52.40		ivar . Jores padronizado	os como variáveis			.00128	00183	
11	47.21		01/				.01351	.01905	
12	48.39	.347	OK Colar	Reconfigurar	Cancelar	Ajuda	00511	00699	
13	50.19	1.06687	.00039	.00117	7	.01455	01579	02136	
14	40.86	1.27964	.00034	.00063	3	02269	.01872	.02621	
15	45 36	98336	00013	00033	1	00845	- 01418	00006	



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- O gráfico é publicado no 'Visualizador de Resultados'
- Os 'Resíduos Padronizados' da VD (ZRE_1) tem uma média muito próximo de 0,
- Neste caso podemos concluir que se cumpre o pressuposto da Média Condicional Zero dos Erros.





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

4. Avaliação do Pressuposto IV: Homocedasticidade



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Homocedasticidade (ou Igual Variância)

 A distribuição dos resíduos apresenta uma variância constante ao longo dos valores previstos da variável dependente. Não há indicação de variação nãoconstante.

 Neste, o valor dos resíduos aproxima-se de 0 para os valores mais baixos da predição, mas aumentam à medida que os valores previstos também aumenta.

ou seja, a variação não é constante.



Random Cloud (No Discernible Pattern)

Heteroscedasticity









Fan Shape (Pattern)

Fonte: https://towardsdatascience.com/assumptions-of-linear-regression-algorithm-ed9ea32224e1



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Homocedasticidade

- Para avaliar se se cumpre este pressuposto, temos de olhar para o Gráfico de Dispersão que compara a distribuição dos 'Residuos Padronizados' com os 'Valores Preditos Padronizados'
 que o SPSS produz automaticamente.
- Neste caso, a representação da distribuição parece sugerir que a variação dos resíduos é relativamente constante.
- Ou seja, cumpre-se o pressuposto da Homocedasticidade





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

5. Avaliação do Pressuposto V: Independência dos Erros



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Independência dos Erros

- Para avaliar se se cumpre este pressuposto, temos de olhar para o resultado do teste Durbin-Watson - que pedimos ao SPSS para produzir.
- Interpretação:
- = 2 -> Erros <u>são</u> independentes
- > 2 / <2 -> Erros <u>não são</u> independentes
- Neste caso os erros são independentes





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

6. Avaliação do Pressuposto VI: Ausência de Multicolinearidade Perfeita



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Ausência de Multicolienaridade Perfeita

- Quando há fortes relações lineares entre os preditores numa regressão, a precisão dos coeficientes de regressão diminui em comparação com o que teria sido se os preditores não se correlacionassem entre si
- Um valor de VIF > 3 sugere a existência de colinearidade no modelo
- Deve repensar-se as variáveis a incluir no modelo



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Ausência de Multicolinearidade

- Para testarmos este pressuposto, temos de olhar para a Tabela de Coeficientes - que o SPSS produz automaticamente.
- Interpretação

VIF > 3 -> presença de colinearidade

- Neste caso, não se identifica a presença de colinearidade...
- Portanto, cumpre-se o pressuposto da ausência de Multicolinearidade





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Validação do Modelo de Regressão Linear

7. Avaliação do Pressuposto VII: Ausência de Observações Influentes



Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- A existência de 'Outliers' (valores extremos) não é um problema em si.
- Mas torna-se um problema quando os Outlier têm influência sobre os resultados do modelo
- Nos painéis de baixo, o Outlier é uma 'Observação Influente'





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Ausência de Observações Influentes

 Para testarmos a presença de observações influentes vamos usar a variável com os 'Distância de Cook' (COO_1) que acabamos de criar.

ta *ADG	GRH_BD_Aul	a10.s	av [Conjuntode[Dados1] - Editor de dad	os do IBM SPSS Statis	tics			- 🗆	×
Arquivo	Editar	<u>V</u> isu	alizar <u>D</u> ados	<u>T</u> ransformar <u>A</u> n	alisar <u>G</u> ráficos	<u>U</u> tilitários E <u>x</u> tensões	Janela Ajuda			
			🛄 🗠	∽⊒ 📓 📥		i 📰 📃 👖				
1:COO_	1		.00031099841	585				Visi	ível: 32 de 32	variáveis
	y_wa	ge2	evaluation 2	N ZRE_1	<i>∲</i> COO_1	LEV_1	SDB0_1		🔗 SDI	B2_1
1	1752	9.60	55.17	-1.13895	.0003	.00075	02008	.01582		010
2	2484	3.10	51.75	73316	.0000	.00043	00244	.01050		007
3	2987	4.72	54.79	.08094	.0000	.00032	.00054	00115		.000
4	2099	4.46	58.30	10802	.0000	.00110	00361	.00153		.002
5	1004	4.01	49.89	-1.61943	.0006	.00073	.00170	02284		.032
6	3232	1.53	52.74	.26865	.0000	.00061	.00066	00379		.005
7	1875	7.44	47.60	51252	.0000	.00059	.00244	00757		005
8	910	6.36	52.65	-1.39629	.0004	.00069	00962	02039		.014
9	2043	2.69	48.48	09011	.0000	.00036	00012	00130		.000
10	2932	7.90	52.40	09098	.0000	.00061	00032	.00128		001
11	1822	3.72	47.21	93018	.0001	.00064	02087	.01351		.019
12	2966	8.47	48.39	.34721	.0000	.00096	.00550	00511		006
13	3565	8.63	50.19	1.06687	.0003	.00117	.01455	01579		021
14	3573	3.01	40.86	1.27964	.0003	.00063	02269	.01872		.026
15	3616	0 52	45 36	98336	0001	I <mark>N</mark> 00033	00845	- 01418		
Visualiz	ação de da	dos	Visualização d	e variável						
						O processador o	to IBM SPSS Statistics	esta pronto Unic	ode:ON	



(A)

(в)

(C**)**

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'DispersãoPontos'
- Selecionar 'Dispersão (Simples)'

tadgrh_B	D_Aula10.sav [ConjuntodeDados1] - Ed	litor de dados do	BM SP	SS Statis	tics			
Arquivo <u>E</u> di	itar <u>V</u> isualiza	ar <u>D</u> ados <u>T</u> ransfo	rmar <u>A</u> nalisa	<u>G</u> rá	ficos	<u>U</u> tilitários	E <u>x</u> tensões	Janela	Ajuda
2			🛱 🖶 🚽	1	<u>C</u> onstru	tor de Gráfic	:0		
🔚 💶 💶 💶 🖬 🔛 🔤 🖬 🔤 Seletor de modelo de tabelas de gráficos									
2:SDB3_1	00	0841231486353		Caixas de diálogo legadas				•	
	evaluation 2	SRE_1	<i>∳</i> ∕° COO_1		4	LEV_1	- 	DB0_1	-
1	55.17	-1.13895	l l	0031		.000	75	02008	
2	51.75	73316	.0	0009		000	13	- 00244	
3	54.79	.08094	.0	0000		.000	32	.00054	





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'DispersãoPontos'
- Selecionar 'Dispersão (Simples)'
- Selecionar Variável 'id'
- Colocar no eixo 'x'





G

Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'Gráficos' / 'Construtor de Gráfico'
- Selecionar 'DispersãoPontos'
- Selecionar 'Dispersão (Simples)'
- Selecionar Variável 'id'
- Colocar no eixo 'x'
- Selecionar Variável 'Cooks Distance'
- Colocar no eixo 'x'





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Selecionar 'ID de grupos/ponto'
- Selecionar 'Rótulo da ID do Ponto'
- Selecionar Variável 'id'
- Colocar na caixa 'Variável do rótulo do ponto'
- Selecionar 'OK'





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

- Reparem que o gráfico permite identificar o ID dos outliers
- Interpretação
- CD > 4/n -> Caso Influente
- Neste caso, o valor de corte é 0.008 (4 / 5000)
- Neste caso não há observações acima do valor de corte.
- Cumpre-se o pressuposto da ausência de observações influentes





Aula 10: Validação e Interpretação de Modelos de Regressão Linear

Por hoje é tudo. Até à próxima aula!