

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo

Nem todos os factores explicativos do valor esperado condicionado da variável dependente são de natureza quantitativa. Por vezes é necessário considerar um ou mais **factores qualitativos**;

Exemplo: Quando se pretende explicar o salário uma variável explicativa a considerar é o **género**. Uma outra é a **localização** do emprego. Estas são **variáveis qualitativas**.

A introdução de variáveis qualitativas no modelo passa pela definição das chamadas **variáveis artificiais**.

Variáveis artificiais (“dummies”): são variáveis que apenas podem assumir o valor 0 ou o valor 1

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo – variáveis artificiais

Factor qualitativo com 2 níveis:

Variável artificial (“dummy”) - $d = \begin{cases} 0 & \text{se atributo não está presente} \\ 1 & \text{se atributo está presente} \end{cases}$

Exemplo: a **variável artificial** género será definida como: $d = 1$ se género feminino e $d = 0$ se género masculino.

A variável qualitativa pode ter efeito no termo constante β_0 ou numa ou mais das variáveis explicativas x_i .

Efeito no termo constante - o modelo escreve-se:

$$E[\mathbf{y}_i | x_{i1}, x_{i2}, x_{i3}, d_i] = \beta_0 + \delta d_i + \beta_1 educ_i + \beta_2 exper_i + \beta_3 antig_i$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo – variáveis artificiais

Factor qualitativo com 2 níveis:

A - **Efeito apenas no termo constante** - o modelo escreve-se:

$$E[\mathbf{y}_i | x_{i1}, d_i] = \beta_0 + \delta d_i + \beta_1 educ_i$$

Estima-se o modelo e efectuam-se os testes como se fosse outra variável qualquer. A única diferença reside na interpretação a dar ao parâmetro.

O coeficiente δ representa a diferença, no termo independente, entre o salário médio auferido por mulheres ($d_i = 1$) e por homens ($d_i = 0$).

O valor de δ traduz a variação autónoma de salário das mulheres quando comparado à mesma variação de salário dos homens.

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

A - Efeito apenas no termo constante - o modelo escreve-se:

$$E(\lnsal_i) = \beta_0 + \delta Mulher + \beta_1 educ_i$$

$$Mulher = \begin{cases} 1 & \text{se for sexo feminino} \\ 0 & \text{caso contrario} \end{cases}$$

$$E(\widehat{\lnsal}_i) = 6.2 - 0.19 + 0.0546 educ_i$$

$\delta = -0.19$ diz-nos que, tudo o resto constante o salário médio das mulheres é aproximadamente 19% inferior ao dos homens para o mesmo número de anos de escolaridade.

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

A - Efeito apenas no termo constante - o modelo escreve-se:

Regression Statistics

Multiple R	0.391853207
R Square	0.153548936
Adjusted R Square	0.15185094
Standard Error	0.379833124
Observations	1000

ANOVA

	df	SS	MS	F	Significance F
Regression	2	26.09310638	13.04655	90.4295	8.12E-37
Residual	997	143.8403825	0.144273		
Total	999	169.9334889			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	6.200159574	0.063624299	97.44955	0	6.075307	6.325012
mulher	-0.188915925	0.024734521	-7.63774	5.17E-14	-0.23745	-0.14038
educ	0.054641686	0.00490542	11.13904	3.04E-27	0.045016	0.064268

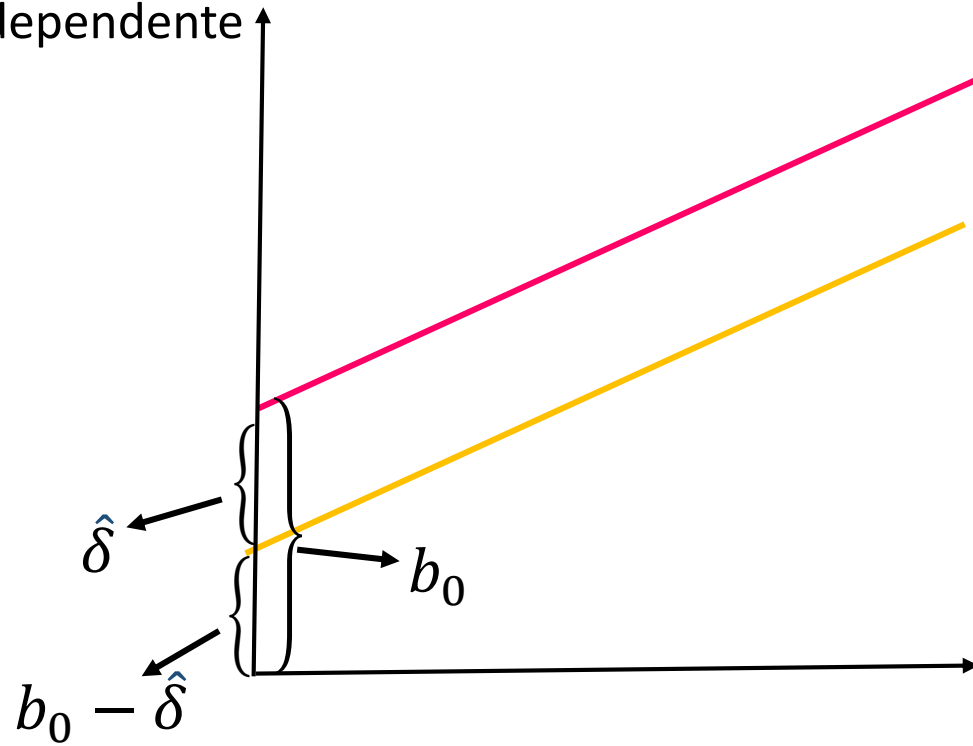
Capítulo 11 – Complementos sobre o Modelo de Regressão linear

A - Efeito apenas no termo constante - o modelo escreve-se:

$$\text{Modelo sem variável artificial} \quad E[\mathbf{Ln\ sal}_i | educ_i] = \beta_0 + \beta_1 educ_i$$

$$\text{Modelo com variável artificial} \quad E[\mathbf{y}_i | educ, mulher] = \beta_0 + \delta mulher_i + \beta_1 educ_i$$

Variável
dependente



$$M = 0 \quad E[\widehat{\mathbf{Ln\ sal}}_i | Educ_i] = 6.2 + 0.054 educ_i$$

$$M = 1$$

$$E[\widehat{\mathbf{Ln\ sal}}_i | Educ_i] = 6.2 - \underbrace{0.19}_{\hat{\delta}} + 0.054 educ_i$$

$\hat{\delta} = -0.19$ diz-nos que, tudo o resto constante o salário médio das mulheres é aproximadamente 19% inferior ao dos homens

Variável
independente

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo – variáveis artificiais

Factor qualitativo com 2 níveis:

B - Efeito apenas no coeficiente de declive - o modelo escreve-se:

$$E[y_i | x_{i1}, d_i] = \beta_0 + \beta_1 x_{i1} + \delta d_i x_{i1}$$

\downarrow
Preço

\nearrow Área (m^2)
 \searrow $d_i = \begin{cases} 0 & \text{se má localização} \\ 1 & \text{se boa localização} \end{cases}$

Estima-se o modelo e efectuam-se os testes como se fosse outra variável qualquer.

β_1 mede o impacto da área no preço esperado para um imóvel situado numa má localização ($d_i = 0$); $\beta_1 + \delta$ mede o mesmo impacto para um imóvel com boa localização ($d_i = 1$).

O valor de δ traduz a variação *da var. dependente* resultante da diferente localização do imóvel

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

B - Efeito apenas no coeficiente de declive - o modelo escreve-se:

$$E(\widehat{\ln sal}_i) = \beta_0 + \beta_1 educ_i + \delta mulher * educ_i$$

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0,389001167					
R Square	0,151321908					
Adjusted R Square	0,149619444					
Standard Error	0,38033247					
Observations	1000					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	2	25,71465973	12,85733	88,88408	3,00872E-36	
Residual	997	144,2188292	0,144653			
Total	999	169,9334889				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	6,128949326	0,063090003	97,14613	0	6,005144897	6,252753755
educ	0,060065189	0,004972262	12,08005	1,88E-31	0,05030789	0,069822489
mulher*educ	-0,014361719	0,001926649	-7,45425	1,96E-13	-0,018142472	-0,010580967

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

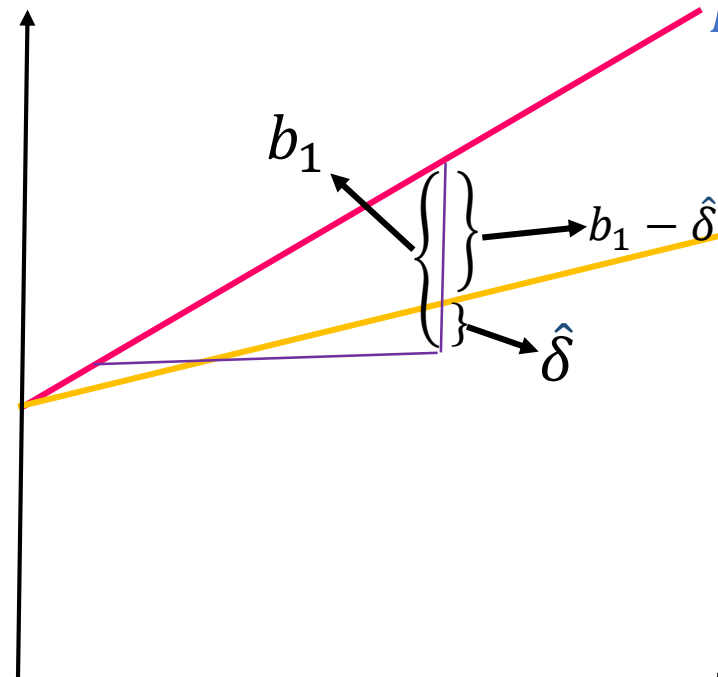
A - Efeito apenas no declive - o modelo escreve-se:

$$\text{Modelo sem variável artificial } E[\mathbf{Ln\ sal}_i | educ_i] = \beta_0 + \beta_1 educ_i$$

Modelo com variável artificial

$$E[\mathbf{y}_i | educ, mulher] = \beta_0 + \beta_1 educ_i + \delta mulher_i * educ_i$$

Variável
dependente



$$M = 0 \quad E[\widehat{\mathbf{Ln\ sal}}_i | Educ_i] = 6.12 + 0.06 educ_i$$

$$M = 1$$

$$E[\widehat{\mathbf{Ln\ sal}}_i | Educ_i] = 6.12 - \left(\underbrace{0.06}_{b_1} - \underbrace{0.014}_{\hat{\delta}} \right) educ_i$$

Variável
independente

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

B - Efeito apenas no coeficiente de declive - o modelo escreve-se:

$$E(\widehat{\ln sal}_i) = 6.129 + 0.06 educ_i - 0.014mulher * educ_i$$

b_1 – tudo o resto constante, um acréscimo de 1 ano escolaridade traduz-se num aumento esperado de 6% no salário dos homens

$\hat{\delta} = -0.014$ diz-nos que, tudo o resto constante, para o mesmo número de anos de escolaridade o salário médio das mulheres é aproximadamente 1,4% inferior ao do homem.

$b_1 + \hat{\delta} = 0.06 - 0.014 = 0.046$ diz-nos que, tudo o resto constante, o valor médio do acréscimo do salário resultante de um ano adicional de escolaridade é para as mulheres apenas de aproximadamente 4.6 % e não de 6%

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo – variáveis artificiais

Factor qualitativo com mais de 2 alternativas:

Definem-se tantas variáveis artificiais quanto o **número de alternativas** do factor **menos uma**.

Exemplo: Considerem-se 3 sectores de actividade para um conjunto de empresas (financeiro, industrial, outros). Como existem **3 categorias** utilizam-se **2 variáveis artificiais**, d_{i1} e d_{i2} .

$$d_{i1} = \begin{cases} 1 & \text{(a empresa pertence ao sector financeiro),} \\ 0 & \text{(caso contrário)} \end{cases}$$

$$d_{i2} = \begin{cases} 1 & \text{(a empresa pertence ao sector industrial),} \\ 0 & \text{(caso contrário)} \end{cases}$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo – variáveis artificiais

Factor qualitativo com mais de 2 alternativas:

Uma vez definidas as variáveis artificiais, segue-se um procedimento em tudo semelhante:

A - **Efeito apenas no termo constante**- o modelo escreve-se:

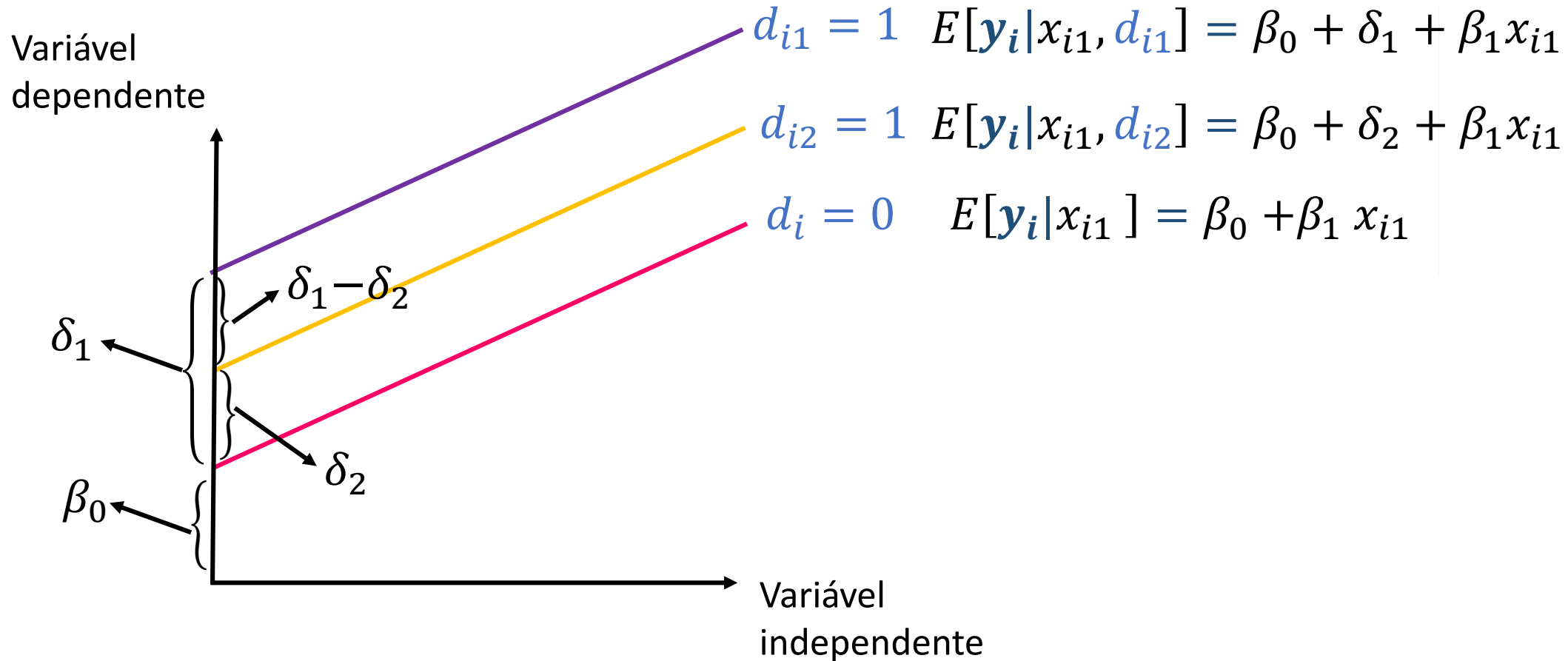
$$E[\mathbf{y}_i | x_{i1}, d_{i1}, d_{i2}] = \beta_0 + \delta_1 d_{i1} + \delta_2 d_{i2} + \beta_1 x_{i1}$$

A correspondência entre sector de actividade e valores das variáveis artificiais é

sector	d_{i1}	d_{i2}
financeiro	1	0
industrial	0	1
Outro	0	0

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Factor qualitativo com mais de 2 alternativas:



Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Introdução de variáveis qualitativas no modelo – variáveis artificiais

Factor qualitativo com mais de 2 alternativas:

Estima-se o modelo e efectua-se os testes como se fossem outras variáveis explicativas quaisquer.

Interpretação:

δ_1 - representa a diferença no termo independente entre o impacto de uma empresa do sector financeiro e outra do sector “outros”, tudo o resto constante.

δ_2 - representa a diferença no termo independente entre o impacto de uma empresa do sector industrial e do sector “outros”.

$\delta_1 - \delta_2$ - a diferença entre o sector financeiro e o sector industrial.

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão

- Uma utilidade do MRL consiste em prever o valor da variável dependente y , conhecidos os valores das variáveis explicativas $x_{01}, x_{02}, \dots, x_{0k}$.

Dois problemas merecem reflexão:

- O que se quer prever?
 - valor de um imóvel concreto
 - valor esperado de certo tipo de imóvel
- Como se quer prever?
 - previsão através de um valor único (previsão pontual);
 - previsão com base num intervalo.

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão Pontual

A melhor solução quer para o valor de um imóvel particular quer para o valor médio de certo tipo de imóvel é dada por aplicação directa do modelo.

$$\widehat{y}_0 = \widehat{E}(y_0) = b_0 + b_1 x_{01} + b_2 x_{02} + \cdots + b_k x_{0k}$$

Se y resulta da transformação de uma variável inicial é por vezes mais adequado aplicar uma correcção quando se procede à transformação inversa.

Exemplo: se $y = \ln(\textit{salario})$

uma solução possível consiste em obter o salário previsto fazendo

$$e^{\widehat{y}_0 + s^2/2}$$

é a estimativa de σ^2

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão pontual

- Exemplo

$$lnsal_i = \beta_0 + \delta mulher + \beta_1 educ_i + \beta_2 exper_i + \beta_3 antig_i + \varepsilon_i$$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.475227522
R Square	0.225841198
Adjusted R Square	0.222729002
Standard Error	0.363615991
Observations	1000

$$mulher = \begin{cases} 1 & \text{género feminino} \\ 0 & \text{género masculino} \end{cases}$$

ANOVA

	df	SS	MS	F	Significance F
Regression	4	38.37798267	9.594496	72.5665	5.58981E-54
Residual	995	131.5555063	0.132217		
Total	999	169.9334889			

	Coefficients	Standard Error	t Stat	P-value
Intercept	5.876184905	0.070086209	83.84224	0
educ	0.055711258	0.004702554	11.84702	2.2E-30
exper	0.023363019	0.002461073	9.49302	1.61E-20
antig	0.004540699	0.002346242	1.935307	0.053236
mulher	-0.19428109	0.023693236	-8.19985	7.37E-16

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão pontual

Exemplo:

$$\lnsal_i = \beta_0 + \beta_1 educ_i + \beta_2 exper_i + \beta_3 antig_i + \varepsilon_i$$

Pretende-se a previsão pontual do salário médio de uma mulher com 10 anos de escolaridade, 15 de experiência e 5 de antiguidade na empresa onde agora se encontra.

$$\begin{aligned}\widehat{\lnsal}_0 &= \hat{E}(\lnsal_0) = 5.8762 - 0.1943 + 0.0557 * 10 + 0.0234 * 15 + 0.0045 * 5 \\ &= 7.0007\end{aligned}$$

⇒ Previsão para o Salário médio desta mulher é
 $\approx e^{\widehat{\lnsal}_0 + s^2/2} = e^{(7 + 0.0364^2/2)} = 1171.74$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão em Média

Preende-se estimar o parâmetro

$$\theta = E(y_t | x_{t1} = c_1, x_{t2} = c_2, \dots, x_{tk} = c_k) = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k$$

Desta equação vem: $\beta_0 = \theta - \beta_1 c_1 - \beta_2 c_2 - \dots - \beta_k c_k$

substituindo no modelo vem:

$$\begin{aligned} y_t &= \theta - \beta_1 c_1 - \beta_2 c_2 - \dots - \beta_k c_k + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i \\ &= \theta + \beta_1 (x_{1i} - c_1) + \beta_2 (x_{2i} - c_2) + \beta_k (x_{ki} - c_k) + u_i \end{aligned}$$

Estimando o modelo com as variáveis transformadas obtêm-se $\hat{\theta}$ e $s_{\hat{\theta}}$

Estimativa para a previsão em média

Desvio padrão da previsão em média

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

1º passo – estimação do modelo

Regression Statistics					
Multiple R	0.7352				
R Square	0.5405				
Adjusted R Square	0.4883				
Standard Error	73.9549				
Observations	50				
ANOVA					
	df	SS	MS	F	Significance F
Regression	5	283103.6251	56620.7250	10.3524	0.0000
Residual	44	240650.1195	5469.3209		
Total	49	523753.7446			
	Coefficients	Standard Error	t Stat	P-value	
Intercept	-124.4477	87.2357	-1.4266	0.1608	
area	2.0177	0.7063	2.8566	0.0065	
area^2	-0.0030	0.0014	-2.0827	0.0431	
na	41.8446	17.9031	2.3373	0.0240	
ncb	46.8517	26.9683	1.7373	0.0893	
dcm	-2.5614	4.1381	-0.6190	0.5391	

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

2º passo – estimação do modelo com variáveis transformadas

Previsão em média quando $area = 200, na = 4, ncb = 2$ e $dcm = 6$

Regression Statistics					
Multiple R	0.7352				
R Square	0.5405				
Adjusted R Square	0.4883				
Standard Error	73.9549	5469.320898			
Observations	50				
ANOVA					
	df	SS	MS	F	Significance F
Regression	5	283103.6251	56620.7250	10.3524	0.0000
Residual	44	240650.1195	5469.3209		
Total	49	523753.7446			
	Coefficients	Standard Error	t Stat	P-value	
Intercept	406.4044	20.4454	19.8775	0.0000	
area-200	2.0177	0.7063	2.8566	0.0065	
area^2-200^2	-0.0030	0.0014	-2.0827	0.0431	
na - 4	41.8446	17.9031	2.3373	0.0240	
ncb - 2	46.8517	26.9683	1.7373	0.0893	
dcm - 6	-2.5614	4.1381	-0.6190	0.5391	

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo

- Uma menor precisão da previsão vai ser compensada pela atribuição de um grau de confiança a esta previsão;
- Os intervalos de previsão (imóvel concreto ou valor médio) vão ser diferentes;
- Para um mesmo grau de confiança, a previsão de um valor médio requer um intervalo de menor amplitude (é mais fácil prever “em média” do que para um caso particular);
- A obtenção dos intervalos é um pouco mais complicada do que o cálculo das previsões pontuais;
- Para além do grau de confiança desejado e do tipo de previsão (imóvel concreto ou valor médio) vão ser muito importantes:
 - Características do imóvel;
 - Qualidade do modelo.

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão pontual

Os intervalos de previsão para $\hat{E}(y_0)$ calculam-se como os intervalos de confiança.

Para um grau de confiança de $(1 - \alpha) * 100\%$

$$\left(\hat{E}(y_0) - t_{\alpha/2} * s_0, \hat{E}(y_0) + t_{\alpha/2} * s_0 \right) \quad t_{\alpha/2} : P \left(T_{(n-k-1)} > t_{\alpha/2} \right) = \alpha/2$$

Como obter s_0 ?

- Utilizar como aproximação (por defeito) $s_0 = \frac{s}{\sqrt{n}}$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão pontual

Calcule o intervalo de confiança a 90% para a previsão do salário médio de uma mulher com 10 anos de escolaridade, 15 de experiência e 5 de antiguidade na empresa onde agora se encontra.

$$(1 - \alpha) = 0.9 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$$

$$t_{\alpha/2}: P\left(T_{(1000-4-1)} > t_{\alpha/2}\right) = 0.05 \Leftrightarrow t_{\alpha/2} = 1.645 \quad s_0 = \frac{s}{\sqrt{n}} = \frac{0.3636}{\sqrt{1000}} = 0.0115$$

$$\begin{aligned} \left(\hat{E}(y_0) - t_{\alpha/2} * s_0, \hat{E}(y_0) + t_{\alpha/2} * s_0\right) &\approx (7 - 1.645 * 0.0115, 7 + 1.645 * 0.0115) \\ &\approx (6.947, 7.0534) \end{aligned}$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão em média

Os intervalos de previsão para θ calculam-se como os intervalos de confiança.

Para um grau de confiança de $(1 - \alpha) * 100\%$

$$\text{Estatística teste: } \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} \sim t_{(n-k)}$$

$$\left(\hat{\theta} - t_{\alpha/2} * s_{\hat{\theta}}, \hat{\theta} + t_{\alpha/2} * s_{\hat{\theta}} \right)$$

$$\text{Com } t_{\alpha/2}: P\left(T_{(n-k-1)} > t_{\alpha/2}\right) = \alpha/2$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão em média

Intervalo de confiança para a previsão em média – Exerc.10.22

Para um grau de confiança de $(1 - \alpha) * 100\%$

$$\text{Estatística teste: } \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} \sim t_{(n-k)}$$

$$\text{Com } t_{0.05/2}: P \left(T_{\left(\frac{50-5-1}{44} \right)} > t_{\alpha/2} \right) = 2.0154$$

$$IC_{\theta}^{95\%} = (406.4044 - 2.0154 * 20.4454; 406.4044 + 2.0154 * 20.4454)$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

- Exercício 7 *Cap. 10* + 3 *Cap. 11*

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.994052024
R Square	0.988139427
Adjusted R Square	0.987348722
Standard Error	0.45093344
Observations	65

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	4	1016.456225	254.1141	1249.694
Residual	60	12.20045803	0.203341	
Total	64	1028.656683		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.046096374	0.253748635	0.181662	0.856461
dim	0.070004068	0.001065701	65.68827	1.35E-57
sat	1.419784503	1.160601656	1.223318	0.225993
sat2	4.348833678	1.166537451	3.727985	0.00043
server	0.123546013	0.159023561	0.776904	0.440268

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

- Exercício 7 *Cap. 10* + 3 *Cap. 11*

7 a) $H_0: \beta_1 \geq 0$ contra $H_1: \beta_1 < 0$ Estatística teste $\frac{b_1 - \beta_1}{s_{\beta_1}} \sim t_{\left(\frac{65-4-1}{60}\right)}$

$$t_{obs} = \frac{0,07 - 0}{0,0011} = 65,68$$

$$\text{Valor} - p = P(t_{(60)} < 65,68) \approx 1$$

ou para nível significância = $\alpha = 0.05$ $t_\alpha: P(t_{(60)} < t_\alpha) = 0.05 \Leftrightarrow t_\alpha = -1.671$

$$W = \{b_1: b_1 < 0 - 1.671 * 0.0011\} = \{b_1: b_1 < -0.0018\}$$

$0,07 \notin W \Rightarrow$ não se rejeita H_0

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.94243621
R Square	0.88818601
Adjusted R Square	0.88457911
Standard Error	1.36203245
Observations	65

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	2	913.6384744	456.8192	246.2462
Residual	62	115.0182085	1.855132	
Total	64	1028.656683		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	2.06265563	0.286657988	7.195528	9.86E-10
dim	0.06896647	0.003151866	21.88116	7.41E-31
server	-0.35196866	0.472471834	-0.74495	0.459116

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

$$7 \text{ b) } H_0: \beta_2 = \beta_3 = 0 \quad \text{contra } H_1: \exists \beta_j \neq 0 \\ j = 2,3$$

$$F = \frac{(VR_0 - VR_1) / \overbrace{(k-p)}^m}{VR_1 / (n - k - 1)} \sim F_{(m, n-k-1)} \quad R_0^2 = 1 - \frac{VR_0}{VT_0} = 0.8882 \Leftrightarrow VR_0 = 115.018$$

$$F_{obs} = \frac{(115.018 - 131.556) / \overbrace{(4-2)}^2}{131.556 / (65 - 4 - 1)} = 252.82 \quad \text{Valor - } p = P(F_{(2,60)} > 252.82) \approx 0$$

$$\text{ou } f_{0.05}: P(F_{(2,60)} > f_{0.05}) = 0.05 \Rightarrow f_{0.05} = 3.1504$$

$$W = \{f_{obs}: f_{obs} > 3.1504\} \Rightarrow 252.82 \in W$$

Rejeita-se H_0 , isto é, pelo menos uma das var.(s) que representam o grau de saturação da rede explica o tempo de transmissão de ficheiros entre 2 computadores

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

- Exercício 7 *Cap. 10* + 3 *Cap. 11*

a) $E[\mathbf{tempo} | \mathbf{dim}, \mathbf{sat}, \mathbf{sat}^2, \mathbf{d}] = \beta_0 + \beta_1 \mathbf{dim} + \delta \mathbf{server} + \beta_1 \mathbf{sat} + + \beta_1 \mathbf{sat}^2$

$$\mathbf{server} = \begin{cases} 1 & \text{ligação passa por um servidor da marca XYZ} \\ 0 & \text{caso contrário} \end{cases}$$

$\hat{\delta} = 0.1235$ Se a ligação passa por um servidor da marca XYZ, em média, o tempo aumenta aproximadamente 0.12 segundos.

$$H_0: \delta = 0 \quad \text{contra} \quad H_1: \delta \neq 0$$

$$\text{Estatística teste } \frac{\hat{\delta} - \delta}{s_{\delta}} \sim t_{\left(\frac{65-4-1}{60}\right)}$$

$$t_{obs} = \frac{0,1235 - 0}{0,159} = 0.7769$$

$$\text{Valor - } p = P(t_{(60)} < 0.7769) = 0.4403$$

$$\text{ou } t_{\alpha/2}: P(t_{(60)} > |t_{\alpha/2}|) = 0.05 \Leftrightarrow t_{\alpha/2} = 2$$

$$W = \{\hat{\delta} : |\hat{\delta}| > 0 + 2 * 0.159\} = \{\hat{\delta} : |\hat{\delta}| > 0.318\}$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

- Exercício 7 *Cap. 10* + 3 *Cap. 11*

$$b) E[\mathbf{tempo} | \mathbf{dim}, \mathbf{sat}, \mathbf{sat}^2, \mathbf{d}] = \beta_0 + \beta_1 \mathbf{dim} + \delta_{server} * \mathbf{dim} + \beta_1 \mathbf{sat} + + \beta_1 \mathbf{sat}^2$$

$$server = \begin{cases} 1 & \text{ligação passa por um servidor da marca XYZ} \\ 0 & \text{caso contrário} \end{cases}$$

$\hat{\delta} = 0.0019$ Se a ligação passa por um servidor da marca XYZ uma variação de 1 MB na dimensão faz o tempo variar, em média, aproximadamente 0.0019 segundos.

$$H_0: \delta = 0 \quad \text{contra} \quad H_1: \delta \neq 0$$

$$\text{Valor} - p = P(t_{(60)} > |0.8982|) = 0.3727$$

$$t_{obs} = \frac{0,019 - 0}{0,0021} = 0.8982$$

$$\text{Estatística teste } \frac{\hat{\delta} - \delta}{s_{\delta}} \sim t_{\left(\frac{65-4-1}{60}\right)}$$

$$\text{ou } t_{\alpha/2}: P(t_{(60)} > |t_{\alpha/2}|) = 0.05 \Leftrightarrow t_{\alpha/2} = 2$$

$$W = \{\hat{\delta} : |\hat{\delta}| > 0 + 2 * 0.0021\} = \{\hat{\delta} : |\hat{\delta}| > 0.0042\}$$

Capítulo 11 – Complementos sobre o Modelo de Regressão linear

Previsão por intervalo - Previsão em média

Exemplo:

$$tempo_i = \beta_0 + \beta_1 dim_i + \beta_2 sat_i + \beta_3 sat_i^2 + \varepsilon_i$$

$$IC_{\hat{E}(tempo_0)}^{0.95} \text{ quando } dim = 100MB, sat = 0.8 \quad s_0 = \frac{s}{\sqrt{n}} = \frac{0.4495}{\sqrt{65}} = 0.0557$$

$$\hat{E}(tempo_0) = 0.0603 + 0.07 * 100 + 1.523 * 0.8 + 4.22 + 0.8^2 = 10.9667$$

$$(1 - \alpha) = 0.95 \Rightarrow \alpha/2 = 0.025 \quad t_{\alpha/2}: P\left(T_{(65-3-1)} > t_{\alpha/2}\right) = 0.025 \Leftrightarrow t_{\alpha/2} = 1.9647$$

$$\left(\hat{E}(y_0) - t_{\alpha/2} * s_0, \hat{E}(y_0) + t_{\alpha/2} * s_0\right) \\ \approx (10.9667 - 1.9647 * 0.0557, 10.9667 + 1.9647 * 0.0557) = (10.857, 11.076)$$