

Prova modelo 1

1. Suponha que se pretendia estudar a relação entre a produtividade média dos trabalhadores nas fábricas (*prod*) e o número de horas de formação profissional que realizaram (*formação*) e que se estimou o seguinte modelo:

$$prod = \beta_0 + \beta_1 \text{formação} + u$$

a) Refira, em sua opinião, qual o efeito esperado da variável *formação* e que outros factores poderão eventualmente influenciar a produtividade dos trabalhadores e que poderiam ter sido considerados. Pensa que esses factores poderão estar correlacionados com as horas de formação profissional?

b) Supondo que alguma das variáveis omitidas estará correlacionada com a *formação* pronuncie-se, justificando, relativamente às propriedades do estimador dos mínimos quadrados de β_1 no modelo anterior.

2. Considere o seguinte modelo estimado que relaciona o peso do bebé à nascença, em kg, (*peso*) com o número médio de cigarros fumado por dia pela mãe (*cigs*):

$$p\hat{e}so = 3.38 - 0.018cigs; n = 605; R^2 = 0.026$$

a) Que peso se prevê para o bebé de uma mulher que fume um número de cigarros igual à média amostral, ou seja, 1.5 por dia? Comente a relação deste valor com o peso médio amostral dos bebés.

b) Escreva em termos teóricos o modelo de regressão que lhe permitiria estimar directamente a alteração percentual (aproximada) no peso do bebé motivada pelo consumo de mais um cigarro.

c) Em relação ao modelo estimado inicial, admita agora que a variável explicativa passa a ser medida em maços (1 maço = 20 cigarros) e que o peso passa a ser medido em gramas. Escreva a nova recta de regressão estimada e interprete os seus coeficientes.

3. Realizou-se um estudo sobre as tarifas de electricidade cobradas aos consumidores domésticos pelas várias empresas europeias do sector, procurando-se determinar quais os factores que são relevantes para a sua definição. Com base numa amostra de 95 observações para o ano de 2005 estimou-se o modelo

$$\hat{Y} = 70.495 + 0.005 \ln(X_1) + 0.004 \ln(X_2) + 0.280 \ln(X_3), R^2 = 0.5343,$$

(12.695) (0.001) (0.001) (0.129)

onde Y representa a tarifa praticada (em cêntimos de euros), X_1 a produção total de electricidade (em milhares de kWh), X_2 o preço unitário do factor trabalho (em euros) e X_3 o preço unitário do factor capital (em euros).

a) Calcule um intervalo de confiança a 99% para o efeito da variável $\ln(X_1)$ sobre a tarifa.

b) Através de um teste apropriado, verifique se o preço do factor trabalho influencia positivamente a tarifa cobrada (utilize um nível de significância de 5%).

4. O Ministério da Educação mandou elaborar um estudo sobre os principais factores que explicam a nota obtida na disciplina de Matemática no exame do 12º ano por alunos de escolas secundárias portuguesas. A equipa encarregada desta análise recolheu uma amostra aleatória formada por 897 alunos com base na qual estimou o modelo

$$nota = 0.05 + 0.25 \underset{(0.12)}{exper} - 0.15 \underset{(0.32)}{turma} + 1.83 \underset{(0.51)}{estudo} - 0.25 \underset{(0.11)}{faltas} + 0.12 \underset{(0.21)}{educp} + 0.35 \underset{(0.17)}{educm}$$

$$R^2 = 0.35, \quad SQT = 4975,$$

onde *exper* representa o número de anos de experiência de leccionação do professor, *turma* é o número de alunos da turma onde o aluno está inserido, *estudo* é o número de horas semanais dedicadas pelo aluno ao estudo da Matemática, *faltas* o número de faltas dadas pelo aluno durante o ano lectivo e *educp* e *educm* o número de anos de escolaridade do pai e da mãe, respectivamente. Em todos os testes que necessitar de realizar para responder às questões que se seguem use o nível de significância de 5%.

a) Teste se o modelo especificado é adequado para explicar as notas obtidas a Matemática.

b) Realizou-se ainda a seguinte regressão adicional:

$$nota = 1.11 + 0.29 \underset{(0.52)}{exper} - 0.14 \underset{(0.31)}{turma} + 1.99 \underset{(0.54)}{estudo} - 0.18 \underset{(0.09)}{faltas}, \quad SQR = 3260$$

Comparando esta nova regressão com a equação estimada inicialmente, que aspecto é possível testar? Realize esse teste e indique as suas conclusões.

5. Considere que se pretende analisar a relação entre o número de espectadores em estádios de futebol (medido em milhares de pessoas) e o número de vitórias das equipas. Usando dados relativos a jogos de 30 equipas num determinado país, obtiveram-se os seguintes resultados pelo método dos mínimos quadrados:

$$espectadores_i = 392.5098 + 12.35297 \text{ vitórias}_i + \hat{u}_i, \quad R^2 = 0.1910, \quad SQT = 196406.647$$

a) Calcule a soma de quadrados residual e a soma de quadrados explicada.

b) Obtenha uma estimativa da variância de $\hat{\beta}_1$.

FORMULÁRIO E TABELAS

Regressão linear simples	Regressão linear múltipla
$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$	$\hat{\beta} = (X'X)^{-1} X'Y$
$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum (X_i - \bar{X})^2}$	$\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$
$\underbrace{\sum (Y_i - \bar{Y})^2}_{SQT} = \underbrace{\hat{\beta}_1^2 \sum (X_i - \bar{X})^2}_{SQE} + \underbrace{\sum \hat{u}_i^2}_{SQR}$	$\underbrace{(Y'Y - n\bar{Y}^2)}_{SQT} = \underbrace{(\hat{\beta}'X'Y - n\bar{Y}^2)}_{SQE} + \underbrace{\hat{u}'\hat{u}}_{SQR}$

$$\hat{\beta} \pm t_{n-p}^{\alpha/2} \cdot \hat{\sigma}_{\hat{\beta}} \quad t = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \quad F = \frac{SQR_* - SQR}{SQR} \frac{n-p}{q} = \frac{R^2 - R_*^2}{1 - R^2} \frac{n-p}{q}$$

Valores críticos da distribuição t-Student

gl	0.10	0.05	0.025	0.01	0.005
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
90	1,291	1,662	1,987	2,368	2,632
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

Valores críticos da distribuição F-Snedcor

gl ^{num→} den↓	1%					5%				
	1	2	3	4	5	1	2	3	4	5
25	7,77	5,57	4,68	4,18	3,85	4,24	3,39	2,99	2,76	2,60
26	7,72	5,53	4,64	4,14	3,82	4,23	3,37	2,98	2,74	2,59
27	7,68	5,49	4,60	4,11	3,78	4,21	3,35	2,96	2,73	2,57
28	7,64	5,45	4,57	4,07	3,75	4,20	3,34	2,95	2,71	2,56
29	7,60	5,42	4,54	4,04	3,73	4,18	3,33	2,93	2,70	2,55
30	7,56	5,39	4,51	4,02	3,70	4,17	3,32	2,92	2,69	2,53
40	7,31	5,18	4,31	3,83	3,51	4,08	3,23	2,84	2,61	2,45
60	7,08	4,98	4,13	3,65	3,34	4,00	3,15	2,76	2,53	2,37
90	6,93	4,85	4,01	3,54	3,23	3,95	3,10	2,71	2,47	2,32
120	6,85	4,79	3,95	3,48	3,17	3,92	3,07	2,68	2,45	2,29
∞	6,63	4,61	3,78	3,32	3,02	3,84	3,00	2,60	2,37	2,21