

Introdução à Probabilidade e Estatística

Soluções da Ficha N°7: Regressão Linear Simples

Para as licenciaturas em: Eng. Civil, Eng. das Energias Renováveis, Eng. Geológica,
Eng. Informática e Eng. Mecatrónica

2º semestre 2014/15 — 2h Teóricas + 2h Práticas

Docentes: Patrícia Filipe e Ana Isabel Santos

1. Pretende-se modelar a resistência de um determinado tipo de plástico em função do tempo que decorre....

(a) $\hat{Y}_i = 153.904 + 2.417X_i$ (no teste o valor de $\tilde{\beta}_0$ estava em falta e foi aqui calculado utilizando a fórmula). $\tilde{\beta}_0$ diz-nos que a resistência do plástico imediatamente após a conclusão da moldagem é, em média, de 153.904.

$\tilde{\beta}_1$ significa que por cada aumento de 1 hora no tempo decorrido desde a conclusão da moldagem e a medição da resistência, verifica-se, em média, um aumento de 2.417 na resistência do plástico.

(b) $R = 0.9796$: Existe uma associação linear positiva muito forte entre a resistência do plástico e o tempo decorrido desde a conclusão da moldagem e a medição da resistência. $R^2 = 0.9596$: cerca 96% da variabilidade da resistência do plástico é explicada pela relação linear que possui com o tempo decorrido desde a conclusão da moldagem e a medição da resistência

(c) $H_0 : \beta_0 = 0 \text{ vs } H_1 : \beta_0 \neq 0$. o quadro do SPSS apresenta o intervalo de confiança a 95% para β_0 , uma vez que o 0 não está contido neste intervalo, também não estará contido no intervalo de confiança a 90% para β_0 , pelo que ao nível de significância de 10% pode concluir que a recta de regressão não passa pela origem

(d) $H_0 : \beta_1 = 0 \text{ vs } H_1 : \beta_1 \neq 0$, quadro do SPSS apresenta, para este teste, o $p\text{-value} \approx 0$, pelo que se rejeita a hipótese nula para qualquer um dos α usuais. Ou seja, é possível afirmar estatisticamente que o tempo desde a conclusão do processo de moldagem até à medição da resistência influencia linearmente de forma significativa a resistência do plástico.

(e) 238.499.

2. Uma empresa que avalia a qualidade de água das ETAR, analisa o ...

- (a) $\hat{Y}_i = 1.2813 + 0.5997X_i$. $\tilde{\beta}_0$ diz-nos que para a ausência de substâncias químicas utilizadas no tratamento da água, o nível de substâncias nocivas é 1.2813 mg/ 100ml. $\tilde{\beta}_1$ significa que por cada aumento de 1 mg/100 ml no nível de substâncias químicas utilizadas no tratamento da água verifica-se, em média, um aumento de aproximadamente 0.6 mg/ 100 ml no nível de substâncias nocivas detectadas na água.
- (b) $R = \sqrt{0.53} = 0.728$: Existe uma associação linear positiva moderada entre o nível de substâncias químicas utilizadas no tratamento da água e o nível de substâncias nocivas detectadas água. $R^2 = 0.53$: 53% da variabilidade do nível de substâncias nocivas detectadas água é explicada pela relação linear que possui com o nível de substâncias químicas utilizadas no tratamento da água.
- (c) $H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$. o quadro do SPSS apresenta o $p\text{-value} \approx 0$ para este teste, logo rejeita-se a hipótese nula para $\alpha = 5\%$ ($\alpha = 5\% > p\text{-value}$). Evidência estatística de que a recta de regressão não passa na origem.
- (d) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$. $t = 2.499 > z_{0.975} = 1.96$, pelo que se rejeita a hipótese nula para $\alpha = 5\%$. Ou seja, é possível afirmar estatisticamente que o nível de substâncias químicas utilizadas no tratamento da água influencia linearmente o nível de substâncias nocivas detectadas na água.
- (e) 0.345
- (f) 0.5997
3. O Presidente da Junta de Freguesia da pacata vila alentejana (e dono da única gasoleneira) ...
- (a) $\hat{Y}_i = 34.58 + 1.47X_i$. $\tilde{\beta}_0$ não faz sentido interpretar. $\tilde{\beta}_1$ significa que por cada aumento de 1 euro no preço da gasolina verifica-se, em média, um aumento de aproximadamente 1.47 l de gasolina vendida.
- (b) $R^2 = 0.96^2 = 0.9216$: 92.16%
- (c) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$. $t = 4.15 > z_{0.975} = 1.96$, pelo que se rejeita a hipótese nula para $\alpha = 5\%$. Ou seja, é possível afirmar estatisticamente o preço da gasolina e a quantidade vendida estão relacionados linearmente de modo significativo.
- (d) 37.5
4. Pretende-se obter um modelo que possa explicar a área foliar (em cm^2) através do comprimento da nervura principal (em cm) em folhas de videiras da casta Fernão Pires.
- (a) $\hat{Y}_i = -106.443 + 24.263X_i$. $\tilde{\beta}_0$ não tem interpretação, pois não faz sentido considerar que um comprimento da nervura principal igual a zero. $\tilde{\beta}_1$ significa que por cada aumento de 1 cm no comprimento da nervura principal a área foliar aumenta, em média, 24.263 cm^2 .
- (b) $R^2 = 0.94$. 94% da variabilidade da área foliar é explicada pelo comprimento da nervura principal.
- (c) 48.526

- (d) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$. $t = 12.550$, pelo que se rejeita a hipótese nula para $\alpha = 5\%$. Ou seja, é possível afirmar estatisticamente que o comprimento da nervura principal influencia linearmente a área foliar.
- (e) $] -177.9927; -34.8933[$
- (f) $\hat{Y}_{9,1} = 114.3519$, donde resulta um resíduo igual a 11.6497.
5. Para alguns países da Europa, foram registados alguns indicadores económicos, nomeadamente o PIBA...
- (a) $Y_i = 15,4623 - 0,1467X_i$.
- (b) $H_0 : \beta_1 = -0,2 \quad vs \quad H_1 : \beta_1 \neq -0,2$. Como $t_{obs} = 0,8417$ não se rejeita H_0 para $\alpha = 5\%$. Ou seja, não existe forte evidência estatística para afirmar (para $\alpha = 5\%$) que o declive da recta de regressão é diferente de -0,2.
- (c) Aproximadamente 29%.
6. No gráfico de dispersão abaixo encontram-se registados os tempos (em segundos) do vencedor na final da corrida de 400 metros masculinos...
- (a) A variável dependente corresponde aos tempos (em segundos) do vencedor na final da corrida dos 400 metros masculinos e a variável explicativa corresponde ao ano em que cada Olimpíada se realizou.
- (b) A relação entre as duas variáveis é negativa, pois à medida que os anos aumentam o tempo do vencedor na final da corrida dos 400 metros masculinos diminui.
- (c) $r^2 = 0,8277$. Aproximadamente 83% da variabilidade verificada nos tempos do vencedor na final da corrida dos 400 metros masculinos é explicada pelo tempo (ano da Olimpíada).
- (d) $\hat{Y}_i = 185,1177 - 0,07111X_i$. A interpretação de β_0 não faz sentido neste exercício, por correspondia ao tempo do vencedor para um ano zero. A interpretação de β_1 corresponde ao decréscimo médio, em segundos, para a Olimpíada seguinte.
- (e) $\hat{Y}_{2008} = 42,328$ segundos.
7. Considerando a idade (em anos) e a capacidade pulmonar (em L) de 9 crianças...
- (a) $r = 0,986$ e $r^2 = 0,972$.
- (b) $\hat{Y}_i = 0,093 + 0,163X_i$.
- (c) $]0,1394; 0,1866[$.
- (d) $t_{obs} = 1,147$. Não se rejeita a hipótese nula ao nível de significância de 1%. Ou seja, não existe forte evidência estatística para afirmar, com 99% de confiança, que a recta de regressão não passa na origem.
8. Pretende-se, se possível, modelar através de uma recta de regressão linear simples a quantidade de vidro...

- (a) $\hat{Y}_i = 10,629 + 56,131X_i$.
- (b) $r^2 = 0,986$.
- (c) $t_{obs} = 21,972$. Rejeita-se a hipótese nula ao nível de significância de 10%.
- (d) $\hat{Y}_{35} = 1975,214$ Kg.

9. Mediu-se o número de pulsações por minuto antes e depois de uma determinada prova de esforço...

- (a) $r = 0,934$ e $r^2 = 0,872$
- (b) $Y_i = 27,55 + 1,323X_i$
- (c) $]0,910; 1,736[$.
- (d) $H_0 : \beta_0 = 0 \quad vs \quad H_1 : \beta_0 \neq 0$. Como $t_{obs} = 2,676 < 3,355$ não se rejeita H_0 para $\alpha = 1\%$. Ou seja, existe forte evidência estatística para afirmar (para $\alpha = 1\%$) que a recta de regressão passa na origem.
- (e) $H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$.
Como (pela alínea c) $0 \notin]0,910; 1,736[$ rejeita-se H_0 . Ou seja, existe forte evidência estatística para afirmar (para $\alpha = 5\%$) que a variável X explica linearmente o comportamento da variável Y .