

Prof Dr. Paulo Canas Rodrigues

Respostas da terceira lista de exercício

Questão 01) Pretende-se, se possível, modelar através de uma reta de regressão linear simples a quantidade de vidro Y produzido num ponto de reciclagem (Kg), usando como variável independente x o número de dias sem despejar o mesmo. Para tal, registaram-se os seguintes dados.

x_i	2	3	4	5	10	15	20	25
Y_i	100	150	-	320	650	810	1040	1480

O valor de Y para $x = 4$ foi perdido, mas antes foram obtidos os seguintes resultados com base nos dados originais:

$$\sum_{i=1}^8 x_i = 84 \quad \sum_{i=1}^8 Y_i = 4800 \quad \sum_{i=1}^8 x_i^2 = 1404 \quad \sum_{i=1}^8 Y_i^2 = 4548000 \quad \sum_{i=1}^8 x_i Y_i = 79700$$

a) Escreva a reta de regressão estimada através do método dos mínimos quadrados.

Precisaremos encontrar os valores de \bar{x} , \bar{Y} , $\hat{\beta}_1$ e $\hat{\beta}_0$ para assim formar a reta de regressão $Y = \hat{\beta}_1 x + \hat{\beta}_0$:

$$\bar{x} = \frac{1}{8} \sum_{i=1}^8 x_i = \frac{84}{8} = 10,5$$

$$\bar{Y} = \frac{1}{8} \sum_{i=1}^8 Y_i = \frac{4800}{8} = 600$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^8 x_i Y_i - n \bar{x} \bar{Y}}{\sum_{i=1}^8 x_i^2 - n \bar{x}^2} = \frac{79700 - 8 \times 10,5 \times 600}{1404 - 8 \times 10,5^2} \approx 56,1303$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x} = 600 - 56,1303 \approx 10,6322$$

$$Y = 56,1303x + 10,6322$$

b) Acha que conseguiu um bom ajuste? Use o coeficiente de determinação.

Tem-se que o coeficiente de determinação é R^2 é dado como:

$$R^2 = \frac{(\widehat{\beta}_1)^2 \sum_{i=1}^8 x_i^2 - n\bar{x}^2}{\sum_{i=1}^8 Y_i^2 - n\bar{Y}^2} = \frac{56,1303^2 \times (1404 - 8 \times 10,5^2)}{4548000 - 8 \times 600^2} \approx 0,986$$

Dado o valor de R^2 obtido muito próximo de 1, pode-se afirmar que o ajuste foi bem sucedido.

c) Qual o valor da quantidade de vidro produzida no ponto de reciclagem que prevê ocorrer em 28 dias sem o despejar?

Não é possível extrapolar valores fora do intervalo usado para o ajuste. Assim sendo, para 28 dias, *não se pode prever o valor de Y produzido.*

d) Teste se o declive da reta de regressão obtida em (a) é zero, usando um nível de significância de 10%. Como interpreta a não rejeição dessa hipótese?

Temos as seguintes hipóteses:

- $H_0 : \beta_1 = 0$
- $H_a : \beta_1 \neq 0$

Cálculo da variância do erro:

$$\hat{\sigma}^2 = \frac{1}{n-2} \left\{ \left(\sum_{i=1}^8 Y_i^2 - n\bar{Y}^2 \right) - (\widehat{\beta}_1)^2 \left(\sum_{i=1}^8 x_i^2 - n\bar{x}^2 \right) \right\}$$
$$\hat{\sigma}^2 = \frac{1}{6} \{ (4548000 - 8 \times 600^2) - 56,1303^2 \times (1404 - 8 \times 10,5^2) \} \approx 3896,8797$$

Cálculo da estatística de teste T :

$$T = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^8 x_i^2 - n\bar{x}^2}}} \sim t_{n-2}$$
$$T = \frac{56,1303 - 0}{\sqrt{\frac{3896,8797}{1404 - 8 \times 10,5^2}}} \sim t_6 \approx 20,5356$$

Para a região crítica com $gl = 6$ e 10% de significância, temos o valor de $t_0 = 1,943$ (tabelado ou no python). Assim sendo, a região de rejeição será:

$$RR =] - \infty; -1,943[U \cup] 1,943; +\infty[$$

Então, como o valor de T obtido é 20,5356, está dentro da região de rejeição, **devemos rejeitar a hipótese nula H_0** . Isso quer dizer que o valor de β_1 é considerável na inclinação e não pode ser desconsiderado.

e) Qual o erro de previsão quando o ponto de reciclagem não é despejado durante 10 dias?

Para o caso, basta achar o valor previsto pelo ajuste e subtrair pelo valor real medido:

- Estimado $\rightarrow Y_{10}^e = 56,1303x_{10} + 10,6322 = 56,1303 \times 10 + 10,6322 = 571,9352$
- Tabelado $\rightarrow Y_{10}^t = 650$
- Diferença $\rightarrow \|\Delta Y_{10}\| = 571,9352 - 650 = 78,0648$

Questão 02) Considere o conjunto de dados “Wage” do pacote “ISLR2” do software R. Considere a variável “health_ins” como variável resposta e as variáveis “age”, “maritl”, “race”, “education”, “jobclass”, “health” e “logwage” como variáveis explicativas. Ajuste uma regressão logística, escreva o modelo final e interprete os coeficientes obtidos.

Questão 03) Numa empresa existem três máquinas para produzir um certo tipo de peças. Foram retiradas amostras aleatórias de dimensão cinco de cada uma das máquinas e foi medido o diâmetro (em mm) de cada uma das peças, resultando nos resultados abaixo: