

Enunciados e Resoluções de Problemas de Exames de Estatística 2 sobre Regressão Linear Simples e Regressão Não-Linear

2009/2010 – Recurso – P4

Um estudo sobre a evolução do número de corvos marinhos numa determinada ilha atlântica revelou um processo que, aparentemente, é de extinção desta espécie animal. Na tabela seguinte apresentam-se o número de corvos marinhos no final de cada um dos últimos 6 anos.

Ano	1	2	3	4	5	6
Número de Corvos	2300	1656	1545	1236	848	754

- Estime e teste um modelo que permita descrever uma evolução exponencial negativa do número de corvos marinhos em função do tempo.
- Sabendo que abaixo de 545 corvos a espécie se extinguirá inexoravelmente, estime a probabilidade de tal situação se atingir no final do ano 7 (admitindo que, entretanto, a evolução verificada nos anos anteriores não se altera).

2009/2010 – Especial – P4

Em resultado de uma auditoria efectuada num armazém de uma empresa, verificaram-se algumas discrepâncias entre o número de efectivo de unidades existentes e o respectivo número registado no inventário informático. Os resultados obtidos foram os seguintes:

Produto	A	B	C	D	E	F	G	H	I	J
Uni. Registadas	23	123	231	56	191	11	88	193	35	123
Uni. Existentes	26	102	218	59	185	15	88	179	38	121

$$\bar{x} = 107.4 \quad \bar{y} = 103.1 \quad s_{XX} = 54756.4 \quad s_{YY} = 46108.9 \quad s_{XY} = 50080.6$$

- Verifique se é ou não plausível a hipótese de que, em média, o número de unidades registadas no sistema informático (variável X) espelha correctamente o número de unidades efectivamente existentes no armazém (variável Y) (utilize o nível de significância de 5%).
- Considerando uma situação em que o valor registado no sistema informático é de 120 unidades, determine o intervalo de confiança a 95% para a previsão do número de unidades existentes no armazém.

2010/2011 – Normal – P4

Pretende estudar-se a relação entre a utilização de um dado fertilizante e a produção de um cereal. No quadro seguinte apresentam-se os dados referentes a 7 observações experimentais:

Fertilizante (kg/hectare)	(X)	100	200	300	400	500	600	700
Rendimento (ton/hectare)	(Y)	4.1	5.3	5.0	6.7	6.5	7.0	7.9

$$\bar{x} = 400 \quad \bar{y} = 6.07 \quad s_{XX} = 280000 \quad s_{YY} = 10.41 \quad s_{XY} = 1630$$

- Teste ao nível de significância de 5% se, de facto, o fertilizante aumenta a produção do cereal em causa.
- Admita que num ensaio posterior de confirmação se observou o valor de 6.8 ton/hectare para uma utilização de fertilizante de 300 kg/hectare. No entender de um analista que estudou este caso, o valor de rendimento é demasiado elevado, pelo que ele se propõe eliminar tal observação (que considera espúria) e repetir a experiência. Concorda com a posição do analista? Justifique.

Resolução

i) Sejam:

X : quantidade de fertilizante utilizada [$kg/hectare$]

Y : Rendimento da produção de milho [$ton./hectare$]

Na figura apresenta-se o gráfico (X,Y) construído com os dados fornecidos.

figura

O gráfico sugere que $Y = f(X)$ pode ser aproximada por uma relação linear, cujo modelo pode ser descrito por:

$$Y_n = \alpha + \beta \cdot (X_n - \bar{X}) + E_n$$

admitindo-se que os erros E_n são independentes e Normais, com média 0 e variância σ^2 .

Os parâmetros α e β podem ser estimados pelos seguintes estimadores:

$$A = \frac{1}{N} \cdot \sum_n Y_n = \bar{Y} \qquad B = \frac{S_{XY}}{S_{XX}}$$

Substituindo, vem

$$a = \hat{\alpha} = 6.07 \qquad b = \hat{\beta} = 1630/280000 = 0.00582$$

Vamos de seguida testar se, de facto, o fertilizante faz aumentar a produção de milho testando se o parâmetro β é positivo.

$$H_0 : \beta = 0$$

$$H_1 : \beta > 0$$

Quando H_0 é verdadeira, a estatística do teste de ET segue aproximadamente uma distribuição t_{N-2}

$$ET = \frac{\beta - \beta_0}{\frac{S}{\sqrt{S_{XX}}}} \rightsquigarrow t_{N-2}$$

Para efectuar este teste falta determinar o valor de S . Ora,

$$S^2 \approx DQMR = \frac{S_{YY} - b \cdot S_{XY}}{N - 2}$$

Vindo,

Substituindo,

$$ET = \frac{0.00582 - 0}{\frac{0.430}{\sqrt{280000}}} = \frac{0.00582}{0.000813} = 7.16$$

Como $ET = 7.16 > t_5(0.05) = 2.01$ a H_0 é rejeitada, concluindo-se, assim, ao nível de significância de 5% que a utilização de fertilizante contribui para o aumento da produção de milho. O valor de prova do teste é de $p = 0.04\%$.

A relação linear estimada entre X e Y será então

$$\hat{Y}_n = 34 + 2.288 \cdot (X_n - 400) + E_n$$

com $\sigma_E = 0.430$.

Uma vez que $\beta > 0$, isto é, que há uma relação linear significativa entre as variáveis X e Y , faz sentido calcular o respectivo coeficiente de determinação

$$R_{XY}^2 = \frac{B^2 \cdot S_{XX}}{S_{YY}}$$

Substituindo,

$$r_{XY}^2 = \frac{0.00582^2 \cdot 280000}{10.41} = \frac{9.49}{10.41} = 0.911 = 91.1\%$$

Estima-se assim que 91.1% da dispersão observada na produção de milho seja explicada pela relação linear entre esta variável e a utilização de um determinado fertilizante.

- ii) Para avaliar a validade da posição do analista vamos calcular a probabilidade de, para uma quantidade de fertilizante de 300 *kg/hectare*, ocorrer uma observação igual ou superior a 6.6 *ton./hectare*, $P(Y \geq 6.6 | X = 300)$.

Começemos por calcular uma previsão de Y quando X é igual a 300 ($Y | X = 300$) com base na expressão da relação linear entre X e Y estimada na alínea anterior,

$$\hat{Y} = \hat{\mu}_Y = 34 + 2.288 \cdot (300 - 400) = 5.49$$

O erro de previsão que se comete é dado pela diferença $\delta = (Y - \hat{Y})$. Então,

$$P(Y \geq 6.6 | X = 300) = P\left(\frac{Y - \hat{Y}}{\sqrt{\sigma^2_{(Y-\hat{Y})}}}\right) \geq \frac{6.6 - 5.49}{\sqrt{\sigma^2_{(Y-\hat{Y})}}}$$

Como se admitiu a Normalidade dos erros E_n , as variáveis Y_n e os erros de previsão $\delta = (Y - \hat{Y})$ são também Normais. Nestas condições, a variável padronizada $(Y - \hat{Y})/\sqrt{\sigma^2_{(Y-\hat{Y})}}$ segue uma distribuição t_{N-2} . A variância do erro de previsão é dada por $s^2_{(Y-\hat{Y})} = Var(Y - \hat{Y}) = [1 + 1/N + (X - \hat{X})^2/S_{XX}] \cdot \sigma^2$. Substituindo, temos

$$s^2_{(Y-\hat{Y})} = \left[1 + \frac{1}{7} + \frac{(300 - 400)^2}{280000}\right] \cdot 0.185 = 0.218$$

Finalmente,

$$P(Y \geq 6.6 | X = 300) = P\left(\frac{Y - \hat{Y}}{\sqrt{\sigma^2_{(Y-\hat{Y})}}}\right) \geq \frac{6.6 - 5.49}{\sqrt{0.218}} = P(t_5 > 2.81) = 0.0188 = 1.88\%$$

Em face desta probabilidade baixa, recomenda-se a eliminação da observação e a repetição da experiência.