Probabilidade e Estatística

Series de Problemas 6

21 de Maio de 2025

1. **Seja X a variável aleatória que representa o número semanal de avarias de um sistema eletrónico. Uma empresa electrotécnica propõe a hipótese H_0 de que**

$$P(X = x) = \frac{1}{2}\,(x+1)(x+2)\,(1-0.4)^x\,(0.4)^3, \quad x = 0, 1, 2, 3,$$

acumulando-se todas as probabilidades para $X > 3$. De 200 semanas obtiveram-se as frequências observadas abaixo:

x	0	1	2	3	> 3
Frequência observada	27	31	39	34	69
Frequência esperada	E_1	23.0	27.6	27.6	E_5

Após o cálculo das frequências esperadas E_1, E_5 (arredondadas às décimas), avalie se H_0 é consistente com os dados. Decida com base no valor-p.

A) **Rejeita-se para 1%, 5% e 10%**

B) **Rejeita-se para 10% e não se rejeita para 1% e 5%**

C) **Não se rejeita para 1%, 5% e 10%**

D) **Rejeita-se para 5% e 10% e não se rejeita para 1%**

Resolução:

Este problema envolve um teste de aderência (goodness-of-fit test) usando a distribuição qui-quadrado para verificar se os dados observados são consistentes com a distribuição proposta sob a hipótese nula H_0.

A função de probabilidade proposta é:

$$P(X = x) = \frac{1}{2}(x+1)(x+2)(0.6)^x(0.4)^3$$

Calculando as probabilidades para cada valor:

$$P(X=0) = \frac{1}{2} \cdot 1 \cdot 2 \cdot 1 \cdot 0.064 = 0.064$$
$$P(X=1) = \frac{1}{2} \cdot 2 \cdot 3 \cdot 0.6 \cdot 0.064 = 0.1152$$
$$P(X=2) = \frac{1}{2} \cdot 3 \cdot 4 \cdot 0.36 \cdot 0.064 = 0.1382$$
$$P(X=3) = \frac{1}{2} \cdot 4 \cdot 5 \cdot 0.216 \cdot 0.064 = 0.1382$$

Para $x > 3$, utilizamos o complementar:

$$P(X > 3) = 1 - (0.064 + 0.1152 + 0.1382 + 0.1382) = 0.5444$$

Com uma amostra de 200 semanas, as frequências esperadas são:

$$E_1 = 200 \times 0.064 = 12.8$$
$$E_2 = 200 \times 0.1152 = 23.0$$
$$E_3 = 200 \times 0.1382 = 27.6$$
$$E_4 = 200 \times 0.1382 = 27.6$$
$$E_5 = 200 \times 0.5444 = 108.9$$

A estatística de teste qui-quadrado é calculada por:

$$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i}$$

Substituindo os valores observados e esperados:

$$\chi^2 = \frac{(27-12.8)^2}{12.8} + \frac{(31-23.0)^2}{23.0} + \frac{(39-27.6)^2}{27.6}$$
$$+ \frac{(34-27.6)^2}{27.6} + \frac{(69-108.9)^2}{108.9}$$
$$= 15.75 + 2.78 + 4.71 + 1.48 + 14.62 = 39.34$$

Sob a hipótese nula H_0, a estatística χ^2 segue uma distribuição qui-quadrado com $k-1 = 5-1 = 4$ graus de liberdade.

O valor-p é:

$$\text{valor-p} = P(\chi^2_4 \geq 39.34) \ll 0.001$$

Este valor é extremamente pequeno, muito menor que qualquer nível de significância usual.

Como valor-p $\ll 0.001$, temos:

- Para $\alpha = 0.01$ (1%): valor-p < 0.01, logo rejeita-se H_0
- Para $\alpha = 0.05$ (5%): valor-p < 0.05, logo rejeita-se H_0
- Para $\alpha = 0.10$ (10%): valor-p < 0.10, logo rejeita-se H_0

Os dados observados apresentam discrepâncias significativas relativamente ao modelo proposto, particularmente nas categorias $x = 0$ (muito mais observações que esperado) e $x > 3$ (muito menos observações que esperado). A hipótese nula é rejeitada para todos os níveis de significância usuais.

Resposta: A opção correta é **(A) Rejeita-se para 1%, 5% e 10%**.

2. **Suspeita-se que o ruído máximo noturno, medido em decibéis (db) em zona residencial próxima de um aeroporto, é uma variável aleatória X que segue uma distribuição normal. Foram recolhidos os seguintes dados relativos a 100 medições noturnas, em dias escolhidos ao acaso:**

Ruído máximo noturno	≤ 50	$]50, 55]$	$]55, 60]$	> 60
Frequência observada	13	19	32	36

Avalie a hipótese H_0 de que $X \sim N(55, 7^2)$. Decida com base no valor-p aproximado e nas frequências esperadas (sob H_0) aproximadas às centésimas.

A) **Rejeita-se H_0 para 10% e não se rejeita H_0 para 1% e 5%**

B) **Rejeita-se H_0 para 5% e 10% e não se rejeita H_0 para 1%**

C) **Rejeita-se H_0 para 1%, 5% e 10%**

D) **Não se rejeita H_0 para 1%, 5% e 10%**

Resolução:

Este problema envolve um teste de aderência para verificar se os dados observados seguem uma distribuição normal com parâmetros específicos.

Sob a hipótese nula $H_0 : X \sim N(55, 7^2)$, calculamos as probabilidades teóricas para cada classe usando a função de distribuição acumulada da normal padrão.

Para $P(X \leq 50)$:

$$P(X \leq 50) = \Phi\left(\frac{50 - 55}{7}\right) = \Phi(-0.7143) \approx 0.2370$$

Para $P(50 < X \leq 55)$:

$$P(50 < X \leq 55) = \Phi(0) - \Phi(-0.7143) = 0.5000 - 0.2370 = 0.2630$$

Para $P(55 < X \leq 60)$:

$$P(55 < X \leq 60) = \Phi(0.7143) - \Phi(0) = 0.7630 - 0.5000 = 0.2630$$

Para $P(X > 60)$:

$$P(X > 60) = 1 - \Phi(0.7143) = 1 - 0.7630 = 0.2370$$

Com uma amostra de 100 observações, as frequências esperadas são:

$$E_1 = 100 \times 0.2370 = 23.70$$
$$E_2 = 100 \times 0.2630 = 26.30$$
$$E_3 = 100 \times 0.2630 = 26.30$$
$$E_4 = 100 \times 0.2370 = 23.70$$

A estatística qui-quadrado de aderência é:

$$\chi^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

Substituindo os valores observados e esperados:

$$\chi^2 = \frac{(13-23.70)^2}{23.70} + \frac{(19-26.30)^2}{26.30} + \frac{(32-26.30)^2}{26.30} + \frac{(36-23.70)^2}{23.70}$$
$$= 4.83 + 2.03 + 1.24 + 6.38 = 14.48$$

Sob a hipótese nula, a estatística χ^2 segue uma distribuição qui-quadrado com $k-1 = 4-1 = 3$ graus de liberdade (não há parâmetros estimados a partir dos dados).

O valor-p é:

$$\text{valor-p} = P(\chi^2_3 \geq 14.48) \approx 0.0023$$

Comparando com os níveis de significância:

- Para $\alpha = 0.01$ (1%): valor-p < 0.01, logo rejeita-se H_0
- Para $\alpha = 0.05$ (5%): valor-p < 0.05, logo rejeita-se H_0
- Para $\alpha = 0.10$ (10%): valor-p < 0.10, logo rejeita-se H_0

O valor-p extremamente pequeno indica evidência muito forte contra a hipótese de que os dados seguem uma distribuição normal com média 55 e desvio padrão 7.

Resposta: A opção correta é **(C) Rejeita-se H_0 para 1%, 5% e 10%**.

3. **Por forma a avaliar a relação entre a percentagem de hidrocarbonetos presentes no condensador principal de uma unidade de destilação (x) e a pureza do oxigênio produzido (Y), recolheu-se uma amostra casual de dimensão $n = 10$ que conduziu aos seguintes resultados:**

$$\sum_{i=1}^{10} x_i = 12.5, \quad \sum_{i=1}^{10} x_i^2 = 16.01, \quad \sum_{i=1}^{10} y_i = 930,$$

$$\sum_{i=1}^{10} y_i^2 = 86594, \quad \sum_{i=1}^{10} x_i y_i = 1168.3,$$

onde $x \in [0.9, 1.6]$.

Com base no método de regressão linear simples, determine um intervalo de confiança a 99% para o valor esperado de Y quando $x = 1.25$.

A) [**83.1367, 102.863**]
B) [**85.2048, 100.795**]
C) [**88.1632, 97.8368**]
D) [**91.4705, 94.5295**]

Resolução:

Este problema envolve a construção de um intervalo de confiança para o valor esperado de Y num ponto específico $x = 1.25$ usando regressão linear simples.

Comece-se por calcular as médias amostrais:

$$\bar{x} = \frac{12.5}{10} = 1.25, \quad \bar{y} = \frac{930}{10} = 93$$

Observe-se que $\bar{x} = 1.25$ coincide exatamente com o ponto onde pretendemos construir o intervalo. Isto simplifica os cálculos, pois a previsão pelo modelo de regressão linear simples será:

$$\hat{y}(1.25) = \bar{y} = 93$$

Para construir o intervalo de confiança, necessitamos do desvio padrão residual S. Calculemos as somas necessárias:

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 16.01 - 10 \times (1.25)^2 = 0.385$$

O coeficiente angular do modelo é:

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{S_{xx}} = \frac{1168.3 - 10 \times 1.25 \times 93}{0.385} = 15.065$$

Para calcular o desvio padrão residual, utilizamos:

$$S^2 = \frac{\text{SSE}}{n-2}, \quad \text{onde SSE} = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$$

Após os cálculos, obtém-se $S \approx 1.4415$.

Para um intervalo de confiança de 99% no valor esperado $E[Y|x = 1.25]$, utilizamos a distribuição t com $n-2 = 8$ graus de liberdade. O valor crítico é $t_{0.995,8} \approx 3.355$.

Como $x = 1.25 = \bar{x}$, o erro padrão da previsão simplifica-se para:

$$\text{SE} = S\sqrt{\frac{1}{n}} = \frac{1.4415}{\sqrt{10}} \approx 0.4558$$

A margem de erro é:

$$\text{ME} = t_{0.995,8} \times \text{SE} = 3.355 \times 0.4558 \approx 1.5295$$

O intervalo de confiança de 99% é:

$$\hat{y}(1.25) \pm \text{ME} = 93 \pm 1.5295 = [91.4705, 94.5295]$$

Resposta: A opção correta é **(D)** **[91.4705, 94.5295]**.

4. **Para testar a relação entre a altura das ondas (x, em metros) e o montante Y (em milhares de euros) dos estragos causados na orla costeira em dias de forte agitação marítima, foram obtidas observações relativas a 17 dias com forte agitação marítima que conduziram a:**

$$\sum_{i=1}^{17} x_i = 121, \quad \sum_{i=1}^{17} x_i^2 = 893, \quad \sum_{i=1}^{17} y_i = 359,$$

$$\sum_{i=1}^{17} y_i^2 = 7595, \quad \sum_{i=1}^{17} x_i y_i = 2555.$$

Ao testar a significância do modelo de regressão linear simples de Y sobre x (i.e., ao confrontar as hipóteses $H_0 : \beta_1 = 0$ e $H_1 : \beta_1 \neq 0$), recorrendo ao valor-p:

A) **Não se rejeita $H_0 : \beta_1 = 0$ a 1%, 5% e 10%**

B) **Rejeita-se $H_0 : \beta_1 = 0$ a 10% e não se rejeita a 1% e 5%**

C) **Rejeita-se $H_0 : \beta_1 = 0$ a 1%, 5% e 10%**

D) **Rejeita-se $H_0 : \beta_1 = 0$ a 5% e 10% e não se rejeita a 1%**

Resolução:

Este problema envolve um teste de significância para o coeficiente angular β_1 num modelo de regressão linear simples, testando se existe relação linear entre a altura das ondas e os estragos causados.

Calcule-se primeiro as médias de cada variável:

$$\bar{x} = \frac{121}{17} \approx 7.1176, \quad \bar{y} = \frac{359}{17} \approx 21.1176$$

Determine-se os somatórios centrados necessários:

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 893 - 17 \times (7.1176)^2 \approx 3.6948$$
$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 2555 - 17 \times 7.1176 \times 21.1176 \approx -0.0273$$

O estimador do coeficiente angular é:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx \frac{-0.0273}{3.6948} \approx -0.0074$$

Para calcular o erro padrão de $\hat{\beta}_1$, necessitamos da variância residual. O resíduo quadrático (SSE) é:

$$\text{SSE} = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i$$

onde $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$. Após os cálculos, obtém-se:

$$S^2 = \frac{\text{SSE}}{n-2} \approx 28.8935$$

O erro padrão de $\hat{\beta}_1$ é:

$$\text{SE}(\hat{\beta}_1) = \sqrt{\frac{S^2}{S_{xx}}} \approx \sqrt{\frac{28.8935}{3.6948}} \approx 2.7961$$

A estatística T segue uma distribuição t com $n-2=15$ graus de liberdade:

$$T = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)} \approx \frac{-0.0074}{2.7961} \approx -0.0026$$

O valor-p do teste bilateral é:

$$\text{valor-p} = 2 \times P(T_{15} \leq |-0.0026|) \approx 2 \times 0.499 \approx 0.998$$

Como o valor-p é extremamente elevado (aproximadamente 0.998), muito superior a todos os níveis de significância usuais (1%, 5% e 10%), não rejeitamos a hipótese nula $H_0 : \beta_1 = 0$ para nenhum destes níveis.

Isto indica que não há evidência estatística suficiente de uma relação linear significativa entre a altura das ondas e os estragos causados na orla costeira.

Resposta: A opção correta é **(A) Não se rejeita** $H_0 : \beta_1 = 0$ **a 1%, 5% e 10%**.

5. **A perda percentual de massa (Y) de uma certa substância metálica (quando exposta a oxigénio seco a 500°C) depende do período de exposição (x, em horas). Um conjunto de 6 medições conduziram a:**

$$\sum_{i=1}^{6} x_i = 16, \quad \sum_{i=1}^{6} x_i^2 = 48.5, \quad \sum_{i=1}^{6} y_i = 0.207,$$

$$\sum_{i=1}^{6} y_i^2 = 0.007777, \quad \sum_{i=1}^{6} x_i y_i = 0.612$$

Preencha a caixa abaixo com o valor do coeficiente de determinação com, pelo menos, 4 casas decimais.

Resolução:

Este problema envolve o cálculo do coeficiente de determinação R^2 numa regressão linear simples, que mede a proporção da variância de Y explicada pelo modelo.

Calculem-se primeiro as médias das variáveis:

$$\bar{x} = \frac{16}{6} = 2.6667, \quad \bar{y} = \frac{0.207}{6} = 0.0345$$

Em seguida, determinam-se os somatórios centrados necessários:

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 48.5 - 6 \times (2.6667)^2 = 5.8333$$
$$S_{yy} = \sum y_i^2 - n\bar{y}^2 = 0.007777 - 6 \times (0.0345)^2 = 0.0006355$$
$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 0.612 - 6 \times 2.6667 \times 0.0345 = 0.0599931$$

O coeficiente de determinação é dado pela fórmula:

$$R^2 = \frac{(S_{xy})^2}{S_{xx} \times S_{yy}}$$

Substituindo os valores calculados:

$$R^2 = \frac{(0.0599931)^2}{5.8333 \times 0.0006355} = \frac{0.003599}{0.003706} \approx 0.9711$$

Este resultado indica que aproximadamente 97.11

Resposta: O coeficiente de determinação é 0.9711.