Universidade do Minho

2ª FICHA DE AVALIAÇÃO

modelo 2022/23 – ESTATISTICA APLICADA – LEMec

Duração: 60 min

Leia com atenção cada pergunta e, sempre que possível, apresente todos os cálculos que tiver que efetuar

Nome: ____ Nº: ____________

PARTE A – EXERCÍCIO OBRIGATÓRIO [5 valores]

1. Foi realizado um teste para verificar a capacidade que um determinado polímero tem para remover resíduos tóxicos da água. O teste foi realizado para três temperaturas diferentes (baixa, média e alta temperatura representadas por 1, 2 e 3, respetivamente). Para cada temperatura foram recolhidos sete valores de percentagem de impurezas que foram removidas pelo polímero, num total de 21 observações independentes. Para cada uma das temperaturas foram obtidos os seguintes resultados:

$$\mathrm{T}_{1.} = 261 \qquad \mathrm{T}_{2.} = 254 \qquad \mathrm{T}_{3.} = 291 \qquad \sum_{i=1}^{3}\sum_{j=1}^{7} x_{ij}^2 = 31258$$

a. Qual a variável em análise?

b. Qual o tipo de planeamento experimental utilizar de forma a verificar se o polímero funciona igualmente bem em todas as três temperaturas?

c. Quais os pressupostos necessários ter em consideração na aplicação do planeamento experimental identificado anteriormente?

d. Formule as hipóteses (nula e alternativa) associadas ao teste estatístico.

e. Para o calculo da estatística do teste, F, complete as células em branco (identificadas pelas letras de "A" a "H") do quadro resumo da tabela ANOVA:

Fonte de variação	Soma dos quadrados	Graus de liberdade	Média dos quadrados	v.a. F
A=	B=	D=	G=	H=
Resíduos	C=	E=	11,80952	
Total	322,95238	F=		

f. Qual a decisão estatística em relação ao funcionamento do polímero nas três temperaturas? Considere o nível de significância de 5%.

g. Calcule o intervalo de confiança para a diferença das médias em relação à capacidade do polímero para remover resíduos tóxicos da água a baixa e alta temperatura (1 e 3), a uma probabilidade de 95%. Interprete o resultado obtido e relacione-o com o resultado da alínea anterior.

Para uma melhor explicação dos resultados obtidos, considerar a figura 1 com a distribuição dos dados.

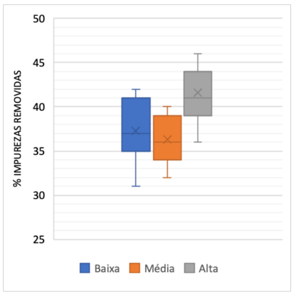

Figura 1. Distribuição das % impurezas removidas para as três temperaturas

PARTE B – <u>UM</u> EXERCÍCIO À ESCOLHA [2,5 valores]

2. Considere a v.a. X proveniente de uma população de distribuição normal $N(\mu, \sigma^2)$ com σ^2 conhecido e igual a 4.
A fim de testar a hipótese $H_0: \mu = 1$ contra a hipótese alternativa $H_1: \mu \neq 1$ ao nível de significância α, é extraída uma amostra $X_1, \ldots, X_n$ da população, e usa-se a seguinte regra:

rejeitar H_0 se $\bar{x} > 1 + c$ ou $\bar{x} < 1 - c$, onde $\bar{x}$ é a média dos valores da amostra observados.

a. Determine o valor de c em função da dimensão da amostra e do nível de significância. Identifique a região de rejeição para o caso em que $n = 64$ e $\alpha = 0{,}05$.

b. Determine a função potência do teste quando $\mu = 1$ e $\mu = 0{,}75$, para as condições consideradas na alínea anterior.
(Nota: se não conseguiu calcular o valor de c na alínea anterior, considere $c = 0.49$).

3. Pretende-se verificar a hipótese de o tempo necessário para a execução de várias tarefas numa linha de montagem (em min) seguir uma distribuição Normal. Para tal, foi recolhida uma amostra com 120 medições, distribuídas conforme ilustra a tabela.

Tempo necessário (min)	<75	75–100	100—125	>125	Total
frequência	5	46	65	4	120

A média e o desvio padrão amostral obtidos foram de 100min e 15min, respetivamente. Com o objetivo de verificar a veracidade da hipótese considerada, foram efetuados alguns cálculos, no entanto, não houve tempo para os terminar. Complete a tabela com os valores em falta ("A" e "B") de forma a permitir testar a hipótese inicialmente considerada. Utilize α=0.05 e apresente o teste completo.

Tempo necessário (min)	<75	75–100	100—125	>125	Total
frequência, f_i	5	46	65	4	120
p_i	A=	0,4522	B=	0,0478	1,0000
$n\,p_i$					120

$E.T.:\quad Q=\sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i}$ onde $e_i=np_i$ $\quad R.R.:\quad Q>\chi^2_{\alpha;\,gl}$ onde $gl=k-1-p$, p corresponde ao número de parâmetros estimados

PARTE C – UM EXERCÍCIO À ESCOLHA [2,5 valores]

4. Com o objetivo de determinar como a habilidade para executar uma determinada tarefa complexa é influenciada pela quantidade de treino, foram usados 15 indivíduos aos quais foi dado um treino que variava de 3 a 12 horas. Depois do treino foram registados os tempos que cada um deles gastou a executar a tarefa. Representando por x a duração do treino (em horas) e por y o tempo gasto na execução da tarefa (em minutos), os resultados obtidos de uma forma resumida, foram os seguintes:

$$\sum_{i=1}^{15}x_1^2=811.2;\ \sum_{i=1}^{15}y_1^2=31350.6;\ \sum_{i=1}^{15}x_iy_i=4867.6;\ \bar{x}=7.2;\ \bar{y}=45.6;\ \sum_{i=1}^{15}(y_i-\hat{y}_i)^2=62.824;\ \sum_{i=1}^{15}(y_i-\bar{y})^2=160.2$$

a. Com os dados disponíveis, calcule o coeficiente de correlação e interprete o valor encontrado no contexto do problema.
b. Os dados fornecem evidência suficiente para concluir que existe uma correlação negativa entre a duração do treino e o tempo de execução da tarefa? Use α=0.01, e conclua.

Considerando: coeficiente de correlação de Pearson: $R=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})^2}}$, coeficiente de determinação: $R^2=1-\frac{\sum(Y_i-\hat{Y}_i)^2}{\sum(Y_i-\bar{Y})^2}=\frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sum(Y_i-\bar{Y})^2}$

Teste de hipótese: $T=\frac{R\sqrt{n-2}}{\sqrt{1-R^2}}\ \sim\ t_{n-2}$

5. O artigo "*Chronological Trend in Blood Lead Levels*" (N. Engl. J. Med., p. 1373-1377) apresenta os dados a seguir sobre y = nível médio de chumbo no sangue de crianças brancas entre 6 meses e 5 anos de idade e x = quantidade de chumbo usada na produção de gasolina (em 1000 t) em sete períodos de seis meses:

$\boldsymbol{x}$	48	59	79	91	97	102	107
$\boldsymbol{y}$	9.3	11.0	12.8	14.6	13.6	16.0	18.2

E de uma forma resumida:

$$\sum_{i=1}^{7}(x_i-\bar{x})^2=3013.429;\ \sum_{i=1}^{7}(x_i-\bar{x})y_i=380.014;\ \sum_{i=1}^{7}(y_i-\bar{y})^2=53.797;\ \sum_{i=1}^{7}(x_i-\bar{x})(y_i-\bar{y})=380.014.$$

a. Com os dados disponíveis, estime uma equação de regressão que possa ser usada para prever o nível médio de chumbo no sangue em função da quantidade de chumbo usada na produção de gasolina e interprete os parâmetros estimados no contexto do problema.
b. Estime o nível médio de chumbo no sangue para uma quantidade 100 de chumbo usada na produção de gasolina. O que pode dizer relativamente ao nível médio de chumbo no sangue para uma quantidade de 150 de chumbo usada na produção de gasolina?

Para o modelo de regressão estimado do tipo $\hat{Y}_i=\hat{\beta}_0+\hat{\beta}_1X_i$, $\quad\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}$, $\quad\hat{\beta}_1=\frac{\sum(X_1-\bar{X})Y_i}{\sum(X_i-\bar{X})^2}$