DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º Semestre - 2022/2023
Exame Época de Recurso

Data: 22 de julho de 2023 Duração: 2 horas e 30 minutos

Resolução

O exame foi resolvido recorrendo ao software R: ver script_ExameERecurso_ME_22_23.

1. (a) O *consumo máximo* foi de 121.1 kWh, registado na lâmpada 74, de LED. Tem um nível de iluminação de 995.8 lux e a temperatura ambiente do local onde se situa a lâmpada é de 26º celsius. Habitualmente está acesa 8 horas por dia e ilumina um espaço com 4 trabalhadores.

   O *consumo de energia mediano* é de 78.3 kWh e coincide com a lâmpada 46, do tipo Fluorescente, com 582.2 lux, a uma temperatura ambiente de 15ºC. Habitualmente está ligada 5 horas e ilumina um espaço com 7 trabalhadores.

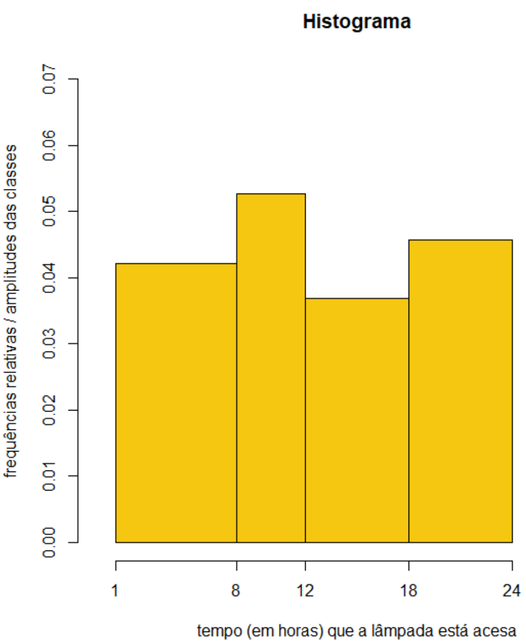
   (b) Tabela de frequências:

i	Classes x_i	Freq. Absoluta n_i	Freq. Relativa f_i	FR/Amplitude (∗) $\frac{f_i}{\text{amplitude das classes}}$	Freq. Absol. Acum. N_i	Freq. Relat. Acum. F_i
1	[1,8[	28	0.2947	0.0421	28	0.2947
2	[8,12[	20	0.2105	0.0526	48	0.5052
3	[12,18[	21	0.2211	0.0368	69	0.7263
3	[18,24]	26	0.2737	0.0228	95	1
		$n = 95$	1			

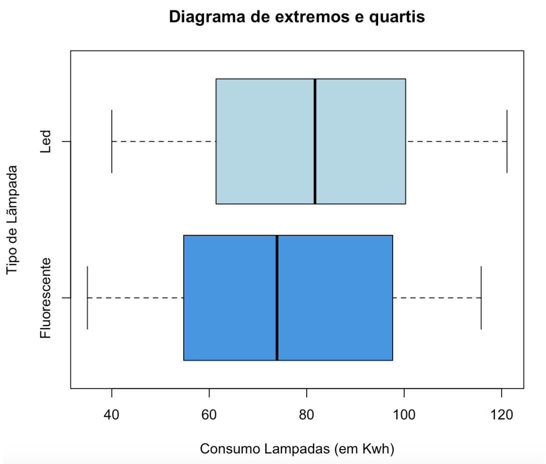
(∗) Eixo dos yy do histograma

Representação gráfica:

Como o tempo (em horas) que habitualmente as lâmpadas estão acesas, por dia, está representado em intervalos (como uma variável aleatória contínua), a representação gráfica adequada será o histograma. Note-se ainda que, as classes têm amplitudes diferentes, pelo que as frequências relativas são as mais adequadas para representar os dados, no entanto devem ser proporcionais à amplitude das classes de modo a área total das barras do histograma corresponder a 1.



(c) Tendo em conta o tipo de iluminação, obtém-se a seguinte representação gráfica para o consumo:



	Consumo das lâmpadas	
	LED	Fluorescente
máximo	121.1 kWh	115.8 kWh
mínimo	40 kWh	35 kWh
amplitude total	81.1 kWh	80.8 kWh
1º quartil	58.5 kWh	54.75 kWh
mediana	81.7 kWh	73.90 kWh
3º quartil	102.1 kWh	97.65 kWh
amplitude inter-quartil	43.6 kWh	42.9 kWh

Em nenhum dos tipos de lâmpadas, o consumo apresenta valores considerados "outliers". Nas lâmpadas do tipo LED, o consumo tem os valores do máximo, do mínimo e dos quartis maiores do que os registados nas lampadas Fluorescente. Também a dispersão dos dados em relação às amplitudes (total e inter-quartil) é ligeiramente superior nas lâmpadas LED. De salientar ainda que, nas lâmpadas LED o consumo apresenta uma forma simétrica, enquanto nas lâmpadas fluorescente regista-se uma ligeira assimetria positiva.

A amplitude do consumo de energia que contém 50% das observações centrais, é a Amplitude Inter-quartil. Assim, 50% das observações centrais das lâmpadas tipo *LED* apresenta uma amplitude de 43.6 Kwh, enquanto que nas lâmpadas do tipo *Fluorescente* essa amplitude é de 42.9 Kwh.

(d) Como $b_1 = 0.0484 \simeq 0$, pode-se dizer que a distribuição do consumo de energia é simétrica. Desta forma, uma sugestão para modelar estes dados é a distribuição Normal pois é uma distribuição simétrica.

Seja $X \rightarrow$ o consumo, em kWh

Hipóteses: $H_0 : X \sim Normal$ contra $H_1 : X \nsim Normal$

Escolha do Teste: Como se pretende testar se X se comporta de acordo com uma Distribuição Normal (sem mais informação) e $n = 95 > 50$, vamos recorrer ao teste de ajustamento Lilliefors.

Tomada de Decisão: se valor$-p \leq \alpha$, então Rejeita-se H_0, ou seja, a sugestão não é válida.

Como valor$-p = 0.1349$, então a partir de $\alpha \geq 0.1349$ Rejeita-se H_0, ou seja, rejeita-se a hipótese do consumo (em kWh) seguir uma distribuição Normal (a sugestão não é válida).

(e) Populações:

$X_F \rightarrow$ a iluminação, em lux, das lâmpadas de tipo Fluorescente e

$X_L \rightarrow$ a iluminação, em lux, das lâmpadas de tipo LED

Amostras:

$n_F = 56$ e $n_L = 39$

amostras independentes

Hipóteses:

$$\begin{cases} H_0 : \mu_F = \mu_L \longrightarrow & \text{em média, o nível de iluminação das lâmpadas do tipo Fluorescente} \\ & \text{é idêntico ao nível de iluminação das lâmpadas do tipo LED} \\ \text{contra} & \\ H_1 : \mu_F \neq \mu_L \longrightarrow & \text{em média, o nível de iluminação das lâmpadas do tipo Fluorescente} \\ & \text{é diferente do nível de iluminação das lâmpadas do tipo LED} \end{cases}$$

$$\Leftrightarrow \qquad H_0 : \mu_F - \mu_L = 0 \quad \text{contra} \quad H_1 : \mu_F - \mu_L \neq 0$$

nível de significância $= \alpha = 0.03$

Tipo de teste: o teste é bilateral

Estatística de Teste: Como temos amostras independentes, $n_F = 56 \geq 30$ e $n_L = 39 \geq 30$ e considerando Populações Quaisquer com σ_F e σ_L desconhecidos, a estatística de teste é:

$$Z = \frac{\left(\overline{X}_F - \overline{X}_L\right) - (\mu_F - \mu_L)}{\sqrt{\frac{S_F^2}{n_F} + \frac{S_L^2}{n_L}}} \dot{\sim} N(0,1)$$

Tomada de Decisão pelo valor-p:

Como valor-p $= 0.2331 > 0.03 = \alpha$, então Não se Rejeita H_0.

<u>OU</u>

Tomada de Decisão pela Região crítica:

$$RC = \left]-\infty, z_{\frac{\alpha}{2}}\right] \cup \left[z_{1-\frac{\alpha}{2}}, +\infty\right[= \left]-\infty, z_{\frac{0.03}{2}}\right] \cup \left[z_{1-\frac{0.03}{2}}, +\infty\right[=]-\infty, -2.1701] \cup [2.1701, +\infty[$$

Como $Z_{obs} = --1.1925 \notin RC$, então Não se Rejeita H_0.

Conclusão: Assim, com 3% de significância e com base nas amostras, conclui-se que não existe evidência estatística que, em média, o nível de iluminação das lâmpadas do tipo Fluorescente seja diferente do nível de iluminação das lâmpadas do tipo LED.

(f) Considerando:

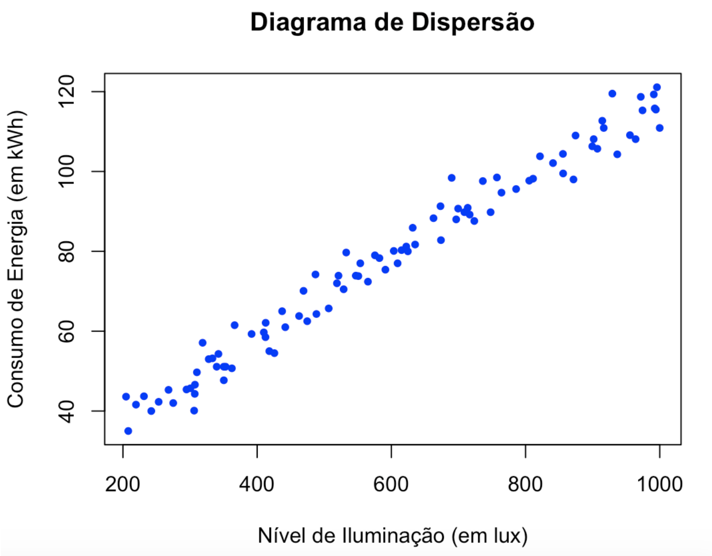
X - Illumination = nível de iluminação da lâmpada (em lux)

Y - Consumption = Consumo de energia, em kWh, da lâmpada

Amostra: $n = 95$

Verificar se o modelo de regressão linear poderá ser adequado:

- diagrama de dispersão:



Com base no diagrama de dispersão parece existir uma forte correlação linear positiva entre as variáveis pois é fácil imaginar uma reta com declive positivo a passar pela nuvem de pontos.

- coeficiente de correlação linear de Pearson:

$$r_{XY} = 0.9875$$

Com base no coeficiente de correlação linear de Pearson, as variáveis apresentam uma correlação linear positiva quase perfeita pois $r_{XY} \simeq 1$

Como graficamente e numericamente chegamos à mesma conclusão, podemos concluir que as variáveis apresentam uma correlação linear positiva muito forte logo o modelo de regressão linear simples é adequado para modelar estes dados.

X - Illumination = nível de iluminação da lâmpada (em lux) a Variável Independente

Y - Consumption = Consumo de energia, em kWh, da lâmpada a Variável Dependente

Em que o modelo de regressão linear é: $\hat{y} = a + bx$ com $b = 0.09998$ e $a = 18.287$. Então a reta de regressão linear simples é:

$$\hat{y} = 18.287 + 0.09998x$$

Previsão:

$$\hat{y}\,(100) = 18.287 + 0.09998 \times 100 = 28.28512 \text{ kWh}$$

o consumo previsto para um nível de iluminação de 100 lux é 28.28512 kWh. Embora o modelo de regressão linear tenha sido considerado adequado para modelar os dados, como $x = 100$ não cai dentro do intervalo de valores observados ($100 \notin [204.6, 999.9]$), não podemos considerar que a previsão efetuada seja de confiança. Não sabemos se a reta de regressão encontrada é adequada tem torno de $x = 100$.

2. (a) Considere a variável aleatória contínua X = consumo de energia do aparelho de ar condicionado (em kWh).

$$\begin{aligned} E\,[X] &= \int_{-\infty}^{+\infty} x f\,(x)\,dx = \int_{-\infty}^{0} x \times 0dx + \int_{0}^{1} x \times xdx + \int_{1}^{2} x \times (2-x)dx + \int_{2}^{+\infty} x \times 0dx = \\ &= \int_{-\infty}^{0} 0dx + \int_{0}^{1} x^2dx + \int_{1}^{2} \left(2x - x^2\right)dx + \int_{2}^{+\infty} 0dx = 0 + \left[\frac{x^3}{3}\right]_0^1 + \left[x^2 - \frac{x^3}{3}\right]_1^2 + 0 = \\ &= \frac{1}{3} - 0 + \left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right) = 1 \end{aligned}$$

$$V[1-2X] = (-2)^2 \times V[X] = 4 \times \left(E\left[X^2\right] - E^2\,[X]\right) = 4 \times \left(\frac{7}{6} - 1^2\right) = \frac{2}{3} = 0.6667$$

(b) Considere a variável aleatória contínua Y = temperatura nas salas desse edifício (em $^\circ C$),

$$Y \sim N(25, 3)$$

pois $\mu = E[Y] = 25^\circ C$ e $\sigma = \sqrt{V[Y]} = 3^\circ C$.

i. Considere a variável aleatória discreta W = número de salas com temperatura de pelo menos $26^\circ C$, num piso do edifício com 8 salas,

$$W \sim B(8, 0.3694)$$

pois

$$\begin{aligned} &n = 8 \text{ salas} \\ &p = P(\text{sucesso}) = P\,(Y \geq 26) = 1 - P\,(Y < 26) \underset{\text{v.a. contínua}}{=} 1 - F_Y(26) = 0.3694 \end{aligned}$$

portanto a probabilidade pretendida é

$$P\,(W = 4) = f_W(4) = 0.2061$$

ii. População: $Y \sim N(25, 3)$

Amostra: $n = 15$

É necessário determinar a distribuição amostral de $\bar{X}$: como a População é Normal e $\sigma = 3$ é conhecido tem-se

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \Leftrightarrow Z = \frac{\bar{X} - 25}{\frac{3}{\sqrt{15}}} \sim N(0,1)$$

A probabilidade pretendida é

$$P\left(24 < \bar{X} < 27\right) = P\left(\frac{24-25}{\frac{3}{\sqrt{15}}} < Z < \frac{27-25}{\frac{3}{\sqrt{15}}}\right) = P\left(-1.291 < Z < 2.582\right) \underset{\text{v.a. contínua}}{=}$$

$$= \Phi(2.582) - \Phi(-1.291) = 0.8967$$

(c) Considere a variável aleatória discreta V = número de avarias consecutivas num aparelho de ar condicionado, por ano,

$$V \sim P(3)$$

pois $E[V] = \lambda = 3$ avarias por ano.

Considere a variável aleatória contínua T = tempo (em anos) entre avarias consecutivas do aparelho de ar condicionado,

$$T \sim Exp\left(\frac{1}{3}\right)$$

pois, pela relação entre a distribuição de Poisson e a distribuição Exponencial, tem-se

$$E[T] = \theta = \frac{t}{\lambda} = \frac{1}{3} \text{ anos entre avarias}$$

Como 12 meses = 1 ano e 10 meses = $\frac{5}{6}$ ano, a probabilidade pretendida é

$$P\left(T \geq \frac{5}{6} + 1 | T \geq 1\right) \underset{(*)}{=} P\left(T \geq \frac{5}{6}\right) = 1 - P\left(T < \frac{5}{6}\right) \underset{\text{v.a. contínua}}{=} 1 - F\left(\frac{5}{6}\right) = 0.0821$$

$(*)$ propriedade "falta de memória" da distribuição Exponencial

3. Amostra:

dimensão da amostra: $n = 40$

26 preferem a "cor branca" e $40 - 26 = 14$ preferem a "cor quente"

(a) População: p = proporção de colaboradores que tem preferência pela "cor branca"

Amostra: proporção amostral = $p^* = \frac{26}{40} = 0.65$

Pretende-se construir um intervalo de confiança para p e pretende-se saber se pode ser $p = 0.70$.

Escolha do Intervalo de confiança:

Como temos uma População Binomial e $n = 40 \geq 30$ o intervalo de confiança a $(1 - \alpha) \times 100\%$ para p é dado por:

$$\left]p^* - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*q^*}{n}}; p^* + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*q^*}{n}}\right[$$

logo, o intervalo de confiança a 90% para p é:

$$]0.526, 0.774[$$

Como $0.70 \in]0.526, 0.774[$, então, com 90% de confiança há a possibilidade de 70% dos colaboradores preferir a "cor branca".

(b) Pretende-se determinar o grau de confiança tal que a margem de erro do intervalo calculado na alínea anterior seja no máximo de 0.1

A margem de erro do intervalo de confiança é:

$$\text{margem de erro} = \frac{\text{amplitude do IC}}{2} = \frac{\left(p^* + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*q^*}{n}}\right) - \left(p^* - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*q^*}{n}}\right)}{2} = z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*q^*}{n}}$$

como pretendemos que a margem de erro seja no máximo de 0.1, tem-se:

$$\text{margem de erro} \leq 0.1 \Leftrightarrow z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*q^*}{n}} \leq 0.1$$

como $n = 40$, $p^* = \frac{26}{40} = 0.65$ e $q^* = 1 - p^* = 1 - 0.65 = 0.35$, vem:

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{0.65 \times 0.35}{40}} \leq 0.1 \Leftrightarrow z_{1-\frac{\alpha}{2}} \leq 1.326 \Leftrightarrow 1 - \frac{\alpha}{2} \leq \Phi(1.326) \Leftrightarrow 1 - \frac{\alpha}{2} \leq 0.9076 \Leftrightarrow$$

$$\Leftrightarrow \alpha \geq 2 \times (1 - 0.9076) \Leftrightarrow \alpha \geq 0.1848 \Leftrightarrow 1 - \alpha \leq 1 - 0.1848 \Leftrightarrow 1 - \alpha \leq 0.8152$$

Portanto o grau de confiança a considerar tem de ser no máximo de 81.52%.

(c) População: X = duração das lâmpadas, em anos
Amostra: $n = 12$

i. Uma estimativa pontual para a média da duração das lâmpadas é

$$\bar{x} = 4.95 \text{ anos}$$

Uma estimativa pontual para o desvio padrão da duração das lâmpadas é

$$s = 0.9577 \text{ anos}$$

ii. Interpretação das Hipóteses:

$$\begin{cases} H_0 : \sigma \geq 1.5 \rightarrow & \text{o desvio padrão da duração das lâmpadas é no mínimo de 1.5 anos} \\ \text{contra} & \\ H_1 : \sigma < 1.5 \rightarrow & \text{o desvio padrão da duração das lâmpadas é inferior a 1.5 anos} \end{cases}$$

Teste das Hipóteses:

$$H_0 : \sigma \geq 1.5 \quad vs \quad H_1 : \sigma < 1.5 \quad \Leftrightarrow \quad H_0 : \sigma^2 \geq 1.5^2 \quad vs \quad H_1 : \sigma^2 < 1.5^2$$

Nível de significância: $\alpha = 0.01$

Para poder efetuar o teste de hipóteses paramétrico para a variância é necessário que se possa considerar que a População é Normal. Então vamos começar por recorrer a um teste de ajustamento para verificar se esta amostra poderá vir de uma População Normal:

$$H_0 : X \sim Normal \quad vs \quad H_1 : X \nsim Normal$$

Como $n = 12 < 50$ e apenas queremos testar se é Normal sem especificar parâmetros, então vamos recorrer ao teste de ajustamento Shapiro-Wilk. Como $valor - p = 0.2302 > \alpha = 0.01$, então não se rejeita H_0. Com base na amostra e para um nível de significância de 1%, há evidência estatísticas que é possível considerar que a amostra vem de uma População Normal.

$$H_0 : \sigma^2 \geq 1.5^2 \quad vs \quad H_1 : \sigma^2 < 1.5^2$$

Estatística de Teste: População Normal (visto através do teste de ajustamento Shapiro-Wilk)

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \Leftrightarrow X^2 \sim \chi^2_{(11)}$$

Tipo de teste: o teste é unilateral esquerdo

Tomada de Decisão pela Região crítica:

$$RC = \left[0, x^2_{(n-1);\alpha}\right] = \left[0, x^2_{(11);0.01}\right] = [0, 3.0535]$$

Como $X^2_{obs} = 4.4844 \notin RC$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 1%, existe evidência estatística de que o desvio padrão da duração das lâmpadas é no mínimo de 1.5 anos.