DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º SEMESTRE - 2022/2023
Exame de Época Normal

Data: 11 de julho de 2023 Duração: 2 horas e 30 minutos

Resolução

O exame foi resolvido recorrendo ao software R: ver script_ExameENormal_ME_22_23.

1. (a) Tabela de frequências:

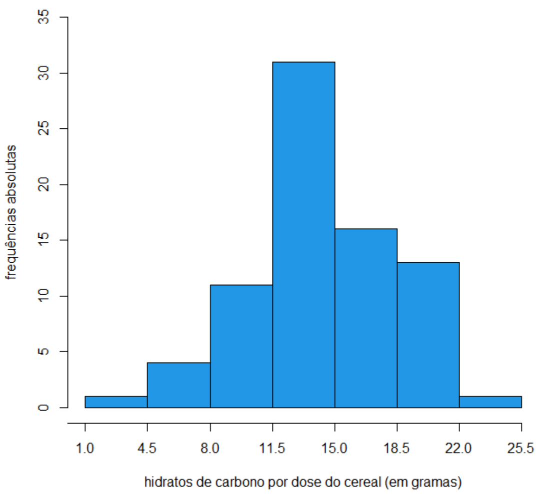
i	Prateleira x_i	Freq. Absoluta n_i	Freq. Relativa f_i	Freq. Absoluta Acum. N_i	Freq. Relativa Acum. F_i
1	mais baixa	20	0.25974	20	0.25974
2	do meio	21	0.27273	41	0.53247
3	mais alta	36	0.46753	77	1
		$n = 77$	1		

Como os dados são qualitativos, a única medida adequada é a Moda:

$$\text{moda} = \text{prateleira mais alta}$$

(pois $n_3 = 36$ é a maior frequência absoluta).

(b) Histograma



O histograma aparenta que os dados se distribuem de forma simétrica, com uma ligeira assimetria negativa (ou enviesada para a esquerda).

(c) Populações:

X_I = quantidade de açúcar nos cereais de pequeno almoço destinados aos consumidores infantis

X_A = quantidade de açúcar nos cereais de pequeno almoço destinados aos consumidores adultos

Amostras:

amostras de dimensão $n_I = 30$ e $n_A = 47$

amostras aleatórias independentes

Hipóteses:

$$\begin{cases} H_0 : \mu_I \leq \mu_A & \text{os cereais de pequeno almoço destinados ao consumidor infantil} \\ & \text{não contêm mais açúcar do que os destinados a adultos} \\ \text{contra} & \\ & \\ H_1 : \mu_I > \mu_A & \text{os cereais de pequeno almoço destinados ao consumidor infantil} \\ & \text{contêm mais açúcar do que os destinados a adultos} \end{cases}$$

$$\Leftrightarrow \qquad H_0 : \mu_I - \mu_A \leq 0 \quad \text{contra} \quad H_1 : \mu_I - \mu_A > 0$$

Nível de significância $= \alpha = 0.04$

Tipo de teste: o teste é Unilateral Direito

Estatística de Teste: Como temos amostras independentes, $n_I = 30 \geq 30$ e $n_A = 47 \geq 30$ e considerando Populações Quaisquer com σ_I e σ_A desconhecidos, a estatística de teste é:

$$Z = \frac{\left(\overline{X}_I - \overline{X}_A\right) - (\mu_I - \mu_A)}{\sqrt{\frac{S_I^2}{n_I} + \frac{S_A^2}{n_A}}} \dot{\sim} N\left(0, 1\right)$$

Tomada de Decisão pela Região crítica:

$$RC = [z_{1-\alpha}, +\infty[= [z_{1-0.04}, +\infty[= [1.7507, +\infty[$$

Como $Z_{obs} = 0.5335 \notin RC$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 4%, não há evidência estatística que a suspeita possa ser considerada verdadeira, ou seja, não há evidência estatística que, em média, os cereais de pequeno almoço destinados ao consumidor infantil contenha mais açúcar do que os destinados a adultos.

(d) População:

$p =$ proporção de cereais na prateleira mais alta destinados aos consumidores infantis

Os responsáveis pretendem $p = 0.10$.

Amostra aleatória:

amostra de dimensão $n = 36$

Escolha do Intervalo de confiança:

Como temos população binomial e $n = 36 \geq 30$ 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 96% para p é:

$$]0.0799, 0.3645[$$

Como $0.10 \in]0.0799, 0.3645[$, então, com 96% de confiança há a possibilidade do hipermercado estar a cumprir a orientação dos responsáveis.

(e) Variáveis:

shelf - variável qualitativa ordinal

client - variável qualitativa nominal

Escolha do Teste:

Como as variáveis são qualitativas e pretende-se verificar se as variáveis estão associadas, vamos recorrer ao teste de independência do Qui-Quadrado.

Hipóteses a testar:

H_0 : Não existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina, ou seja, as variáveis shelf e client são independentes

contra

H_1 : Existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina, ou seja, as variáveis shelf e client não são independentes

Tabela de contingência entre shelf e client:

		Consumidor		
		Adulto	Infantil	Total
prateleira	mais baixa	15	5	20
	do meio	6	15	21
	mais alta	31	5	36
	Total	52	25	77

a tabela de contingência tem $r = 3$ linhas e $c = 2$ colunas

Estatística de Teste:

$$Q \sim \chi^2_{(r-1)\times(c-1)} \Leftrightarrow Q \sim \chi^2_{(2)}$$

Tomada de decisão: se valor-p$\leq \alpha$, então Rejeita-se H_0, ou seja, existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina.

Como valor-p$= 0.0011$, então a partir de $\alpha \geq 0.0011$ Rejeita-se H_0, ou seja, a partir de $\alpha \geq 0.0011$ há evidência estatística que existe associação entre a prateleira onde se encontra a embalagem de cereais e o consumidor a quem se destina.

(f) Considerando:

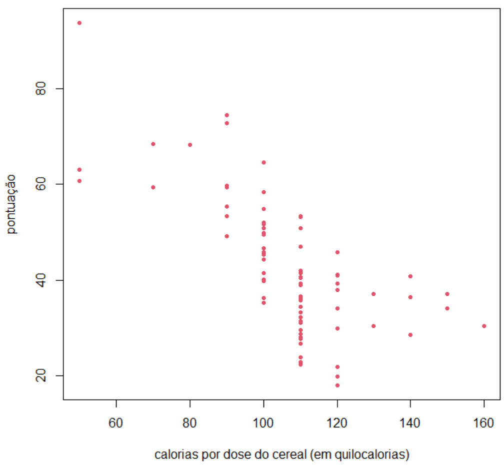
X - calories = calorias por dose do cereal (em quilocalorias)

Y - rating = pontuação de um cereal de pequeno almoço

Amostra: $n = 77$

Verificar se o modelo de regressão linear poderá ser adequado:

- diagrama de dispersão:



Com base no diagrama de dispersão parece existir correlação linear negativa entre as variáveis pois é possível imaginar uma reta com declive negativo a passar pela nuvem de pontos.

- coeficiente de correlação linear de Pearson:

$$r_{XY} = -0.6894$$

Com base no coeficiente de correlação linear de Pearson, as variáveis apresentam uma correlação linear negativa moderada pois $r_{XY} < 0$ e $-0.8 < r_{XY} < -0.5$.

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

X - calorias por dose do cereal (em quilocalorias) a Variável Independente

Y - pontuação de um cereal de pequeno almoço a Variável Dependente

Em que o modelo de regressão linear é: $\hat{y} = a + bx$

$$b = -0.497 \quad \text{e} \quad a = 95.789$$

Então a reta de regressão linear simples é:

$$\hat{y} = 95.789 - 0.497x$$

Previsão:

$$\hat{y}\,(125) = 95.789 - 0.497 \times 125 = 33.6607$$

a pontuação prevista de um cereal de pequeno almoço que apresente 125 quilocalorias por dose é 33.6607. Como o modelo de regressão linear foi considerado adequado para modelar os dados e como $x = 125$ cai dentro do intervalo de valores observados ($125 \in [50, 160]$), podemos considerar que a previsão efetuada é de confiança.

2. Seja X a variável aleatória discreta, que representa a procura diária de um dado modelo de *smartphone* na loja I-TEL.

   (a)
$$P\,(X < 3\,|\,X \geq 1) = \frac{P\,(X < 3 \wedge X \geq 1)}{P\,(X \geq 1)} = \frac{P\,(1 \leq X < 3)}{1 - P\,(X < 1)} = \frac{f\,(1) + f\,(2)}{1 - f\,(0)} = \frac{0.125 + 0.5}{1 - 0.25} = \frac{5}{6}$$

   (b)
$$E[Y] = E[-3X + 2] = -3E[X] + 2 \underset{\text{calc. aux.}}{=} -3 \times 1.5 + 2 = -2.5$$

   cálculo auxiliar:
   $E[X] = 0 \times 0.25 + 1 \times 0.25 + 2 \times 0.5 + 3 \times 0.125 = 1.5$

   (c) Seja $T-$ tempo entre entradas consecutivas de clientes, em minutos,
   então, $T \sim Exp\,(15)$ porque $E[T] = \theta = 15$

   Considere-se a variável aleatória discreta $W-$ número de clientes que entram na loja, em 3 horas. Pela relação entre a distribuição Exponencial e a distribuição de Poisson tem-se

$$W \sim P\left(\lambda = \frac{t}{\theta} = \frac{3 \times 60}{15} = 12\right)$$

$$P\,(W \geq 10) = 1 - P(W < 10) = 1 - P(W \leq 9) \underset{\text{v.a. discreta}}{=} 0.7576$$

   (d) Seja $L-$ tempo de produção de um lote, em dias, onde

$$L \sim N\,(50, 5)$$

   i. Pretende-se determinar k (prazo de entrega), tal que:

$$P\,(L \leq k) = 0.95 \quad \Leftrightarrow F(k) = 0.95 \Leftrightarrow k = F^{-1}(0.95) \Leftrightarrow k = 58.224$$

   Pelo que para garantir que em 95% das vezes o prazo é cumprido, este terá de ser cerca de 58 dias.

   ii. Seja V- número de lotes em 10, que demoram mais de 60 dias em produção.
   Por outro lado,

$$p = P(\text{lote demorar mais de 60 dias}) = P(L > 60) = 1 - P(L \leq 60) = 0.0227 \text{ e } n = 10.$$

   Assim $\quad V \sim B\,(10, 0.0227)$ e

$$P\,(V = 3) = 0.0012$$

3. População

X = velocidade do vento (em nós), $X \sim N(\mu, \sigma)$

Amostra: $n = 11$

(a) Como a População pode ser considerada Normal, então o Intervalo de confiança a $(1-\alpha) \times 100\%$ para σ^2 é

$$\left]\frac{(n-1)s^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right[$$

e sabe-se que o intervalo de confiança obtido para o desvio padrão é $]7.6878; 16.5710[$, ou seja, para a variância fica $]7.6878^2; 16.5710^2[$ então considerando, por exemplo, o limite superior tem-se

$$\frac{(n-1)s^2}{\chi^2_{n-1;\frac{\alpha}{2}}} = 16.5710^2 \Leftrightarrow \frac{(11-1)\times 108.2}{\chi^2_{11-1;\frac{\alpha}{2}}} = 16.5710^2 \Leftrightarrow \chi^2_{10;\frac{\alpha}{2}} = \frac{10 \times 108.2}{16.5710^2} \Leftrightarrow$$

$$\Leftrightarrow \frac{\alpha}{2} = F\left(3.940305\right) \underset{X^2 \sim \chi^2_{(10)}}{\Leftrightarrow} \frac{\alpha}{2} = 0.05 \Leftarrow \alpha = 0.10$$

O grau de confiança utilizado foi de 90%.

(b) Interpretação das Hipóteses:

$$\begin{cases} H_0 : \mu = 40 \rightarrow & \text{a velocidade média do vento é 40 nós} \\ \text{contra} & \\ H_1 : \mu = 38 \rightarrow & \text{a velocidade média do vento é igual a 38 nós, ou seja, inferior a 40 nós} \end{cases}$$

Teste das Hipóteses:

$$H_0 : \mu = 40 \quad \text{contra} \quad H_1 : \mu < 40$$

Estatística de Teste: População Normal e σ desconhecido

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_(n-1)$$

Nível de significância: $\alpha = 0.01$

Tipo de teste: o teste é unilateral esquerdo

Decisão: como valor$-p = 0.731 > 0.01$ então Não se Rejeita H_0

OU

Tomada de Decisão pela Região crítica:

$$RC = \left]-\infty, t_{(n-1);\alpha}\right] = \left]-\infty, t_{(11-1);0.01}\right] =]-\infty, -2.763769]$$

Como $T_{obs} = 0.637693 \notin RC$, então Não se Rejeita H_0.

Conclusão: Com base na amostra e para um nível de significância de 1%, existe evidência estatística de que a média da velocidade do vento não é inferior a 40 nós, ou seja, não é 38 nós.