1º Teste

1. A distonia cervical, também chamada de torcicolo espasmódico, é uma condição dolorosa na qual os músculos do pescoço contraem-se involuntariamente, fazendo com que a cabeça torça ou vire para o lado, também pode fazer com que a cabeça se incline incontrolavelmente para frente ou para trás. Para comparar tratamentos, foram selecionados aleatoriamente pacientes e foram formados 3 grupos aos quais foi injetado no músculo afetado 5 000 unidades de toxina botulínica B, 10 000 unidades de toxina botulínica B ou um placebo. Para analisar os resultados obtidos aceda ao ficheiro "distonia.txt", que se encontra no Moodle, e tem os seguintes campos:
   - id = identificação do paciente
   - treat = tipo de tratamento atribuído ao paciente (1= placebo, 2 = 5 000 unidades de botox B, 3 = 10 000 unidades de botox B)
   - age = idade do paciente, em anos
   - sex = género do paciente (1 = Feminino, 2 = Masculino)
   - twstrs = pontuação atribuída aos sintomas para medir a gravidade, a dor e a incapacidade da distonia cervical (pontuações mais altas significam mais sintomas)

[1.0] (a) Identifique a Amostra indicando a sua dimensão, a unidade estatística e as variáveis estatísticas, classifcando-as.

[1.5] (b) Considera-se que a distonia cervical ocorre com mais frequência nas mulheres do que nos homens. Fazendo uma análise descritiva, com base em tabelas de frequências completas e em medidas adequadas, diga, justificando, se concorda com esta afirmação.

[1.5] (c) A distonia cervical é um distúrbio raro que pode ocorrer em qualquer idade, sendo no entanto mais frequente em pessoas com idade no intervalo $[45, 60[$. Recorrendo a um histograma verifique se a afirmação é verdadeira.

[1.5] (d) Com base no diagrama de extremos e quartis, compare por tipo de tratamento a pontuação atribuída aos sintomas. Comente os resultados obtidos.

[1.5] (e) Devido à forma como se distribuem as idades dos pacientes considera-se que esses dados poderiam ser modelados por uma distribuição semelhante ao gráfico da função densidade de probabilidade da distribuição Normal. Recorrendo a medidas de localização adequadas diga, justificando, se concorda com esta afirmação.

1. (a) Amostra: pacientes com distonia cervical que estão a fazer tratamento

   Dimensão da Amostra: $n = 631$ pacientes

   Unidade estatística: pacientes com distonia cervical

   Variável estatística: treat. Classificação: Qualitativa nominal (mas se considerarmos que Placebo corresponde a 0 unidades de Botox B, pode ser classificada como qualitativa ordinal).

   Variável estatística: age. Classificação: Quantitativa contínua (mas foi recolhida como quantitativa discreta)

   Variável estatística: sex. Classificação: Qualitativa nominal

   Variável estatística: twstrs. Classificação: Quantitativa contínua (mas foi recolhida como quantitativa discreta, também pode ser entendida como uma escala logo pode ser classificada como qualitativa ordinal).

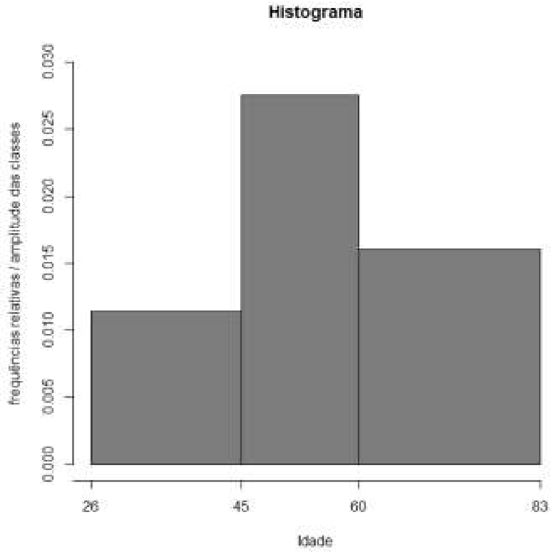
   (b) Tabela de frequências da variável sex:

i	Género x_i	Frequência Absoluta n_i	Frequência Relativa f_i
1	Feminino	395	0.626
2	Masculino	236	0.374
		$n = 631$	1

   moda = Feminino

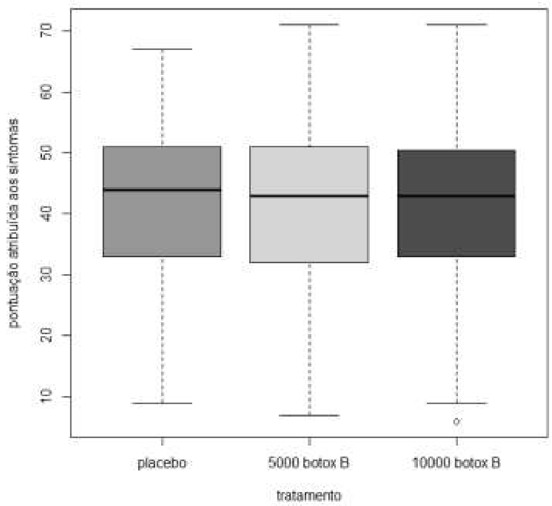
   Com base na tabela de frequências pode-se comprovar que a moda é Feminino, pois é o que apresenta maior frequênmcia (absoluta e relativa). Logo a afirmação parece ser verdadeira, nesta amostra a distonia cervical ocorre com mais frequência nas mulheres do que nos homens.

(c) Como pretende-se ver o que acontece na classe $[45, 60[$, vamos construír um histograma com 3 classes: [mínimo dos dados, 45[, [45, 60[, [60, máximo dos dados]



Com base no histograma, a classe que apresenta maior frequência relativa proporcional à sua amplitude é o intervalo $[45, 60[$. A afirmação parece ser verdadeira, nesta amostra e com base no histograma, a distonia cervical é mais frequente em pessoas com idade no intervalo $[45, 60[$ quando comparado com os outros intervalos.

(d) Diagrama de extremos e quartis da pontuação atribuída aos sintomas por tipo de tratamento:



Não parece haver diferenças significativas em relação à pontuação atribuída aos sintomas nos diferentes tratamentos, os quartis são quase idêncticos. Os tratamentos com toxina botulínica B apresentam os valores mais baixos (6 e 7 pontos) e mais altos dessa pontuação (71 pontos). No caso do tratamento com 10 000 unidades da toxina botulínica B, o valor 6 aparece como o único "outlier" e é um "outlier" considerado moderado.

(e) O gráfico da função densidade de probabilidade da distribuição Normal é um gráfico simétrico. Então vamos analisar a simetria com base nas medidas de localização central:

- moda = 57 anos
- mediana = 56 anos
- média = 55.616 anos

Como os valores são todos próximos, podemos dizer que os dados não se afastam muito da simetria no entanto apresentam uma ligeira assimetria negativa pois:

$$\text{média} < \text{mediana} < \text{moda}$$

Como a assimetria negativa não é muito acentuada ($b_1 = -0.0224 \approx 0$), então concordo que os dados da idade talvez possam ser modelados por uma distribuição Normal.

2. A crescente prevalência da dor na região do pescoço entre as pessoas tem motivado um projeto de pesquisa com o objetivo de desenvolver uma nova prótese de disco cervical. No âmbito desse projeto, sabe-se que o número de próteses defeituosas produzidas por uma máquina em período experimental é uma variável aleatória discreta, X, com a seguinte função de distribuição:

$$F(x) = \begin{cases} 0 & , x < 1 \\ 0.10 & , 1 \leq x < 2 \\ 0.40 & , 2 \leq x < 3 \\ 0.70 & , 3 \leq x < 4 \\ 0.90 & , 4 \leq x < 5 \\ 1 & , x \geq 5 \end{cases}$$

[1.0] (a) Calcule a função de probabilidade da variável aleatória X.

[1.5] (b) Sabendo que não foram produzidas mais de quatro próteses defeituosas, qual a probabilidade de terem sido produzidas pelo menos duas próteses defeituosas?

[1.5] (c) Calcule $V\left[\frac{7-2X}{3}\right]$.

[1.5] (d) O tempo, em minutos, que uma destas próteses demora a ser fabricada, é uma variável aleatória Y com a seguinte função densidade de probabilidade:

$$f(y) = \begin{cases} \frac{1}{40} - \frac{y}{3200} & , 0 \leq y \leq 80 \\ 0 & , \text{caso contrário} \end{cases}$$

Qual a probabilidade de uma prótese demorar mais de uma hora a ser fabricada?

2. (a) A variável aleatória discreta X = número de próteses defeituosas produzidas por uma máquina em período experimental, tem domínio $D_X = \{1, 2, 3, 4, 5\}$ e a sua função de probabilidade é:

x	1	2	3	4	5
$f(x)$	0.1	0.3	0.3	0.2	0.1

pois

- $f(1) = F(1) = 0.10$
- $f(2) = F(2) - F(1) = 0.40 - 0.10 = 0.30$
- $f(3) = F(3) - F(2) = 0.70 - 0.40 = 0.30$
- $f(4) = F(4) - F(3) = 0.90 - 0.70 = 0.20$
- $f(5) = F(5) - F(4) = 1 - 0.90 = 0.10$

(b) Pretende-se

$$P(X \geq 2 | X \leq 4) = \frac{P(X \geq 2 \wedge X \leq 4)}{P(X \leq 4)} = \frac{P(2 \leq X \leq 4)}{F(4)} = \frac{P(1 < X \leq 4)}{0.90} =$$
$$= \frac{F(4) - F(1)}{0.90} = \frac{0.90 - 0.10}{0.90} = \frac{8}{9} = 0.8889$$

(c) Pretende-se

$$V\left[\frac{7-2X}{3}\right] = V\left[\frac{7}{3} + \left(-\frac{2}{3}\right)X\right] = \left(-\frac{2}{3}\right)^2 \times V[X] = \frac{4}{9} \times \left(E\left[X^2\right] - E^2[X]\right) =$$
$$= \frac{4}{9} \times \left(9.7 - 2.9^2\right) = 0.5733 \text{ defeitos}^2$$

pois

$$E[X] = \sum_x x f(x) = 1 \times f(1) + 2 \times f(2) + 3 \times f(3) + 4 \times f(4) + 5 \times f(5) =$$
$$= 1 \times 0.1 + 2 \times 0.3 + 3 \times 0.3 + 4 \times 0.2 + 5 \times 0.1 = 2.9 \text{ defeitos}$$
$$E\left[X^2\right] = \sum_x x^2 f(x) = 1^2 \times f(1) + 2^2 \times f(2) + 3^2 \times f(3) + 4^2 \times f(4) + 5^2 \times f(5) =$$
$$= 1 \times 0.1 + 4 \times 0.3 + 9 \times 0.3 + 16 \times 0.2 + 25 \times 0.1 = 9.7 \text{ defeitos}^2$$

(d) Seja Y = tempo, em minutos, que uma destas próteses demora a ser fabricada.
Como 1 hora = 60 minutos, pretende-se

$$P(Y>60)=\int_{60}^{+\infty} f(y)dy=\int_{60}^{80}\left(\frac{1}{40}-\frac{y}{3200}\right)dy+\int_{80}^{+\infty}0dy=\left[\frac{1}{40}\times y-\frac{1}{3200}\times\frac{y^2}{2}\right]_{60}^{80}+0=$$
$$=\left(\frac{1}{40}\times 80-\frac{1}{3200}\times\frac{80^2}{2}\right)-\left(\frac{1}{40}\times 60-\frac{1}{3200}\times\frac{60^2}{2}\right)=0.0625$$

3. No projeto de investigação pretende-se avaliar a resistência de um novo material utilizado nas próteses de disco cervical. Sabe-se que o número de falhas numa prótese de disco cervical com este novo material é uma variável aleatória com distribuição de Poisson de média 2 falhas por ano.

[1.5] (a) Foi colocada uma prótese de disco cervical num paciente. Qual a probabilidade de ocorrerem três falhas em 5 anos?

[1.5] (b) Sabendo que a segunda falha numa prótese ocorreu há pelo menos 4 meses, calcule a probabilidade do tempo entre a segunda e terceira falhas ser superior a 6 meses?

(c) Os resultados experimentais indicam que a resistência das próteses produzidas pelo material A segue uma distribuição normal com uma média de 300 MPa (megapascals) e um desvio padrão de 20 MPa.

[2.0] i. Sabe-se que a protese pode sofrer fratura prematura se a resistência for inferior a 280 MPa. Qual é a probabilidade de, num grupo de 7 próteses produzidas pelo material A, encontrar pelo menos uma prótese que sofra fratura prematura?

[1.0] ii. O engenheiro que trabalha neste projeto deseja garantir que 5% das próteses produzidas com o material A tenham uma resistência inferior a um determinado valor. Qual é esse valor?

[1.5] iii. Sabe-se que a resistência das próteses produzidas com um material B também segue um comportamento normal com média 290 MPa. Sabendo que 80% das próteses produzidas com o material B apresentam uma resistência superior a 250 MPa, qual o desvio padrão da resistência das próteses produzidas com o material B?

3. Seja X = número de falhas, por ano, numa prótese de disco cervical, $X\sim P(2)$ pois $E[X]=2=\lambda$.

(a) X' = número de falhas, em 5 anos, numa prótese de disco cervical. $X'\sim P(10)$ pois

$$\begin{array}{lcl} 1 \text{ ano} & \mapsto & \lambda=2 \\ 5 \text{ anos} & \mapsto & \lambda=2\times 5=10 \end{array}$$

Pretende-se

$$P(X'=3)=f(3)=0.0076$$

(b) Seja T = tempo, em meses, entre falhas consecutivas. Pela relação entre a distribuiçao de Poisson e a distribuição Exponencial tem-se $T\sim Exp(6)$ pois $\theta=\frac{12}{2}=6$. Logo

$$P(T>6|T\geq 4)\underset{(*)}{=}P(T>6-4)=P(T>2)=1-P(T\leq 2)=1-F(2)=0.7165$$

(∗) propriedade "Falta de memória" da distribuição Exponencial

(c) Seja Y = resistência, em MPa (megapascals), das próteses produzidas pelo material A, tem-se $Y\sim N(300,20)$ pois $\mu=E[Y]=300$ e $\sigma=\sqrt{V[X]}=20$.

i. Seja W = número de próteses com fratura prematura, num grupo de 7 próteses, tem-se $W\sim B(7,0.1587)$ pois

$$n=7$$
$$p=P(\text{fratura prematura})=P(Y<280)\underset{\text{v.a. contínua}}{=}F(280)=0.1587$$

Pretende-se

$$P(W\geq 1)=1-P(W<1)\underset{\text{v.a. discreta}}{=}1-P(W\leq 0)=1-F(0)=0.7016$$

ii. Pretende-se determinar m tal que

$$P(Y<m)=0.05\underset{\text{v.a. contínua}}{\Leftrightarrow}F(m)=0.05\Leftrightarrow m=267.1029 \text{ MPa}$$

iii. Seja V = resistência das próteses produzidas com um material B, tem-se $V \sim N(290, \sigma)$ pois $\mu = E[V] = 290$. Tem-se

$$V \sim N(290, \sigma) \Leftrightarrow Z = \frac{V - 290}{\sigma} \sim N(0, 1)$$

Pretende-se determinar σ tal que

$$P(V > 250) = 0.80 \Leftrightarrow 1 - P(V \leq 250) = 0.80 \Leftrightarrow P(V \leq 250) = 0.20 \Leftrightarrow$$
$$\Leftrightarrow P\left(Z \leq \frac{250 - 290}{\sigma}\right) = 0.20 \Leftrightarrow \Phi\left(\frac{-40}{\sigma}\right) = 0.20 \Leftrightarrow \frac{-40}{\sigma} = z_{0.20} \Leftrightarrow$$
$$\Leftrightarrow \frac{-40}{\sigma} = -0.8416 \Leftrightarrow \sigma = 47.52732 \text{ MPa}$$