Departamento de Matemática da Universidade de Coimbra

**Estatística LEI – LECD**

Proposta de resolução da primeira frequência de Estatística (LEI+LECD) do ano letivo 24/25

**I**

1. (a) Considerem-se os seguintes acontecimentos:

$D =$ "o chip é defeituoso"; $A =$ "o chip foi aprovado na inspeção"

Temos $P(A/D) = 0,03$; $P(A|\overline{D}) = 0{,}98$; $P(\overline{A}) = 0,07$

Pretende-se mostrar que: $P(D) = \frac{1}{19}$.

Tem-se $P(A) = 1 - P(\overline{A}) = 1 - 0,07 = 0,93$.

Com efeito

$$\begin{aligned} P(A) &= P(D)P(A/D) + P(\overline{D})P(A/\overline{D}) \\ \Leftrightarrow 0,93 &= P(D)0,03 + (1 - P(D))0,98 \\ \Leftrightarrow 0,95P(D) &= 0,05 \Rightarrow P(D) = \frac{0,05}{0,95} = \frac{1}{19} \end{aligned}$$

(b) Pretende-se determinar:

$$P(D/A) = \frac{P(D \cap A)}{P(A)} = \frac{P(D)P(A/D)}{P(A)} = \frac{0,03/19}{0,93} = \frac{1}{589}$$

(c) Seja $X =$ "número de chips defeituosos entre os 6 escolhidos (em mais de 3000)".

$X \sim H(n, M, B)$, com $n = 6$, $\frac{B}{M} = P(D) = \frac{1}{19}$ uma vez que os chips são retirados sem reposição. Como $n/M < 6/3000 < 0,1$, $X \overset{\bullet}{\sim} B(6, 1/19)$.

Assim, considerando a variável aleatória $X' \sim B(6, \frac{1}{19})$, obtemos

$$\begin{aligned} P(X \leq 1) &\approx P(X' \leq 1), \\ &= P(X' = 0) + P(X' = 1) = \binom{6}{0} p^0 (1-p)^6 + \binom{6}{1} p^1 (1-p)^5 \\ &= \left(\frac{18}{19}\right)^6 + 6 \cdot \frac{1}{19} \cdot \left(\frac{18}{19}\right)^5 \approx 0,9639 \end{aligned}$$

2. Definimos a variável $Y =$ "número de lotes para exportação ...".

Sabemos que $Y \sim \mathcal{P}(\lambda), \lambda > 0$, e $E(Y^2) = 4\,E(Y)$. Queremos calcular $P(Y < 2)$.

Pela fórmula de Koenig, temos

$$\begin{aligned} & V(Y) = E(Y^2) - E(Y)^2 \\ \Leftrightarrow\ & V(Y) = 4\,E(Y) - E(Y)^2, && \text{porque } E(Y^2) = 4\,E(Y) \\ \Leftrightarrow\ & \lambda = 4\,\lambda - \lambda^2, && \text{porque, como } Y \sim \mathcal{P}(\lambda),\ \text{temos } E(Y) = V(Y) = \lambda \\ \Leftrightarrow\ & \lambda^2 - 3\,\lambda = 0 \\ \Leftrightarrow\ & \lambda\,(\lambda - 3) = 0 \\ \Leftrightarrow\ & \lambda = 0 \vee \lambda = 3 \end{aligned}$$

Como se tem $\lambda > 0$, concluímos que $\lambda = 3$. Logo, $Y \sim \mathcal{P}(3)$ e

$$P(Y < 2) = P(Y \leq 1) = 0.1991 \quad \text{(a partir da tabela).}$$

3. Sejam $V \sim N(4,3)$ e $W \sim N(1,2)$

(a) Temos

$$\begin{aligned} P(V > 3.4) &= P\left(\frac{V-4}{3} > \frac{3.4-4}{3}\right) \\ &= P\left(U > -0.2\right), && \text{onde } U = \frac{V-4}{3} \sim N(0,1) \\ &= P\left(U < 0.2\right), && \text{pela simetria da lei } N(0,1) \\ &= F_U(0.2) \\ &= 0.5793. && \text{pela tabela da lei } N(0,1) \end{aligned}$$

(b) Uma vez que $P(V < 2W) = P(V - 2W < 0)$ vamos definir $D = V - 2W$.
Pela ELN, como $V$ e $W$ são leis normais independentes, temos $D \sim N(m_D, \sigma_D)$, com

$$\begin{aligned} m_D &= E(D) = E(V-2W) = E(V) - 2E(W) = 4 - 2\times 1 = 2, \\ {\sigma_D}^2 &= V(D) = V(V-2W) = V(V) + (-2)^2\, V(W) = 9 + 4\times 4 = 25 \\ &\Rightarrow \sigma_D = 5 \end{aligned}$$

(Note que $V(V)$ representa a variância da variável aleatória $V$)
Portanto $D \sim N(2,5)$ e temos

$$\begin{aligned} P(D < 0) &= P\left(\frac{D-2}{5} > \frac{0-2}{5}\right) \\ &= P\left(U < -0.4\right), && \text{onde } U = \frac{D-2}{5} \sim N(0,1) \\ &= F_U(-0.4) \\ &= 1 - F_U(0.4), && \text{pela simetria da lei } N(0,1) \\ &= 1 - 0.6554, && \text{pela tabela da lei } N(0,1) \\ &= 0.3446 \end{aligned}$$

## II

1. $\forall x > \theta$, temos

$$P(X > x) = 1 - P(X \le x) = 1 - \int_\theta^x \frac{3\theta^3}{t^4} dt = 1 - \left[-\frac{\theta^3}{t^3}\right]_\theta^x = 1 + \left[\frac{\theta^3}{t^3}\right]_\theta^x = 1 + \left(\frac{\theta^3}{x^3} - 1\right) = \left(\frac{\theta}{x}\right)^3.$$

2. Sabe-se que $E(X) = \frac{3}{2}\theta$. Portanto, $\theta = \frac{2}{3}E(X)$. Escrevemos, assim, $\theta$ em função do momento $E(X)$. Em consequência, o respetivo estimador é $T_n = \frac{2}{3}\overline{X}$.

3. (a) Uma estimativa de $E(X)$ é a média da amostra, $\overline{x}$, ou seja 35.4 milhares de horas. Para uma estimativa de $\theta$ usamos o estimador construído na alínea 2). Assim, $\widehat{\theta} = \frac{2}{3}\times\overline{x} = \frac{2}{3}\times 35.4 = 23.6$

(b) Usando o resultado da alínea 1), temos que a estimativa solicitada é $\widehat{p} = \left(\frac{\widehat{\theta}}{30}\right)^3 = \left(\frac{23.6}{30}\right)^3 \approx 0.48$

## III

1. Temos

$$A \subseteq B \Longrightarrow \overline{B} \subseteq \overline{A} \Longrightarrow \overline{A} \cap \overline{B} = \overline{B}$$

pelo que

$$P(\overline{A}/\,\overline{B}) = \frac{P(\overline{A}\cap\overline{B})}{P(\overline{B})} = \frac{P(\overline{B})}{P(\overline{B})} = 1.$$

2. Se na seguinte propriedade da variância

$$V(aX + b) = a^2V(X), \forall a, b \in \mathbb{R},$$

considerarmos $a = -1$ e $b = 2$, obtemos $V(2 - X) = (-1)^2V(X) = V(X)$.

3. O suporte de uma variável aleatória discreta $X$ é o conjunto dos números que $X$ assume com probabilidade estritamente positiva (logo a probabilidade de $X$ assumir valores fora do suporte é zero!) . Se $X$ segue uma lei binomial, então o suporte de $X$ é constituído por números inteiros positivos (ou $S_X \subseteq \mathbb{N}$) — (podemos consultar o formulário). Assim conclui-se que $P(X = 1.5) = 0$.

## IV

1. Um outlier à esquerda de uma amostra é uma observação $x$ que verifica

$$x < q_1 - 1.5 \times (q_3 - q_1)$$

e um outlier à direita é uma observação $y$ que verifica

$$y > q_3 + 1.5 \times (q_3 - q_1).$$

Ora, para esta amostra temos

$$q_1 - 1.5 \times (q_3 - q_1) = 51.323 - 1.5 \times (61.812 - 51.323) = 35.5895$$

e

$$q_3 + 1.5 \times (q_3 - q_1) = 61.812 + 1.5 \times (61.812 - 51.323) = 77.5455.$$

Como o mínimo da amostra é $40.145 > 35.5895$ não existe nenhuma observação $x$ que seja outlier à esquerda desta amostra. Como o máximo da amostra é $74.827 < 77.5455$, não existe nenhuma observação $y$ que seja outlier à direita desta amostra. A amostra não tem outliers.

2. Como o histograma sugere a densidade de uma lei normal (podemos desenhar o esboço do gráfico) e ao diagrama de dispersão que constitui o QQ-plot normal se pode ajustar uma reta, validamos a hipótese de $X$ seguir uma lei normal. ( Como se sabe, mais tarde aprendemos a validar esta hipótese mais formalmente, usando um teste de ajustamento –Shapiro-Wilk ou Anderson-Darling)

3. O Boxplot $A$ indica a presença de outliers à esquerda o que não se verifica nesta amostra. O Boxplot $B$ evidencia uma amplitude entre a terceiro quartil e a mediana $(q_3 - q_2)$ muito superior à amplitude entre a mediana e o primeiro quartil $(q_2 - q_1)$, o que não se verfica nesta amostra, onde estas amplitudes são valores muito próximos. (Também podemos argumentar que o Boxplot $B$ não apresenta a simetria que se destaca nos valores amostrais apresentados na tabela ou que a assimetria patente no Boxplot $B$ não é compatível com a simetria evidenciada pelo histograma). No Boxplot $C$ podemos notar que o terceiro quartil é um valor próximo de 65 e portanto muito superior ao terceiro quartil da amostra que é 61.812.