IPS Instituto Politécnico de Setúbal
Escola Superior de Tecnologia de Setúbal

DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º SEMESTRE - 2020/2021
1.º TESTE

Teste modelo — Duração: 2h00

Resolução

1. Dados relativos ao número de intervenções cirúrgicas realizadas diariamente num bloco operatório de um hospital durante o período de 350 dias.

   (a) Tabela de frequências:

i	x_i	n_i	N_i	f_i	F_i
1	0	42	42	$\frac{42}{350}=0.12$	0.12
2	1	$91-42=49$	91	$\frac{49}{350}=0.14$	$0.12+0.14=0.26$
3	2	$0.24\times 350=84$	$91+84=175$	0.24	$0.26+0.24=0.50$
4	3	$350-(175+21)=154$	$175+154=329$	$\frac{154}{350}=0.44$	$0.50+0.44=0.94$
5	4	21	$329+21=350$	$\frac{21}{350}=0.06$	$0.94+0.06=1$
		$n=350$		1	

   (b) Variável em estudo: número de intervenções cirúrgicas realizadas diariamente num bloco operatório de um hospital.

   Classificação da variável: Variável quantitativa discreta.

   Distribuição Teórica: Distribuição Poisson.

   (c) Número de dias em que o bloco foi utilizado para intervenções cirúrgicas $= n - n_1 =$ $= 350 - 42 = 308$ dias

   (d) Percentagem de dias em que o máximo de intervenções realizadas no bloco foi uma $=$ $= F_2 \times 100\% = 0.26 \times 100\% = 26\%$ dos dias

   (e) Medidas de localização central: moda, média e mediana

$$mo = x_4 = 3 \text{ intervenções cirúrgicas} \qquad (n_4 = 154 \text{ é o maior})$$

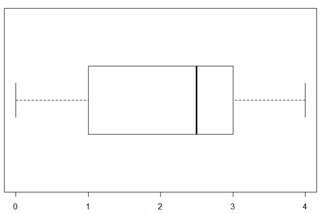
$$\bar{x} = \frac{0\times42+1\times49+2\times84+3\times154+4\times21}{350} = 2.18 \text{ intervenções cirúrgicas}$$

$$\tilde{x} = \mathrm{Q}_2 = Q_{0.50} = \frac{x_3+x_4}{2} = \frac{2+3}{2} = 2.5 \text{ intervenções cirúrgicas} \qquad (F_3 = 0.5)$$

   Como $\bar{x} < \tilde{x} < mo$ a distribuição dos dados diz-se assimétrica negativa.

   (f) Diagrama de extremos e quartis:

$min = x_1 = 0$ intervenções cirúrgicas
$\mathrm{Q}_1 = Q_{0.25} = x_2 = 1$ intervenção cirúrgica $\quad (F_2 = 0.26 \geq 0,25)$
$\mathrm{Q}_2 = 2.5$ intervenções cirúrgicas
$\mathrm{Q}_3 = Q_{0.75} = x_4 = 3$ intervenções cirúrgicas $\quad (F_4 = 0.94 \geq 0,75)$
$max = x_5 = 4$ intervenções cirúrgicas



(g) Como a variável "tempo em minutos que o bloco operatório esteve ocupado por dia" é uma variável quantitativa contínua, então é necessário construir classes, logo o número de linhas da tabela de frequências coincide com o número de classes e como $n = 350$, recorrendo à regra de Sturges tem-se:

$$k = \left\lfloor 1 + \frac{\ln(350)}{\ln(2)} \right\rfloor = 9 \, \text{classes}$$

2. Variáveis: Y o custo de produção (em unidades monetárias) como variável dependente e X a quantidade produzida de próteses dentárias por dia (em dezenas) como variável independente. Amostra: $n = 12$ ensaios.

(a) Verificar se o modelo de regressão linear é adequado:

- Diagrama de dispersão: deveria ser feito mas como os dados não foram disponibilizados não é possível
- Coeficiente de correlação linear:

$$r_{xy} = \frac{13 - 12 \times \frac{4}{12} \times \frac{25}{12}}{\sqrt{\left(3 - 12 \times \left(\frac{4}{12}\right)^2\right) \times \left(68.215 - 12 \times \left(\frac{25}{12}\right)^2\right)}} = 0.9$$

Com base no coeficiente de correlação linear podemos dizer que a correlação linear existente é positiva ($r_{xy} > 0$) e forte ($0.8 < r_{XY} < 1$), então parece ser adequado o modelo de regressão linear, no entanto convinha ter os dados para confirmar com o diagrama de dispersão.

Reta de regressão:

$$b = \frac{13 - 12 \times \frac{4}{12} \times \frac{25}{12}}{3 - 12 \times \left(\frac{4}{12}\right)^2} = 2.8 \qquad \text{e} \qquad a = \frac{25}{12} - 2.8 \times \frac{4}{12} = 1.15$$

$$\hat{y} = 2.8x + 1.15$$

(b) Previsão para a máquina A:

$$x = 2 \, \text{dezenas} \rightarrow \hat{y}(2) = 2.8 \times 2 + 1.15 = 6.75 \text{ unidades monetárias}$$

Previsão para a máquina B:

$$x = 2 \, \text{dezenas} \rightarrow \hat{y}(2) = 2.32 \times 2 + 1.63 = 6.27 \text{ unidades monetárias}$$

Máquina recomendada: como ambos os modelos de regressão apresentam uma correlação linear forte logo, com base no coeficiente de correlação linear não há motivos para duvidar das retas de regressão ajustadas nem das previsões efetuadas ($2 \in [0, 5]$), portanto a máquina recomendada é a máquina B pois o custo de produção previsto é inferior.

Com base nos modelos de regressão, a máquina A seria preferível quando:

$$\hat{y}_A < \hat{y}_B \Leftrightarrow 2.8x + 1.15 < 2.32x + 1.63 \Leftrightarrow x < 1$$

ou seja, quando a produção é inferior a 10 unidades.

(c) Como as variáveis estão em unidades de medida diferentes (dezenas e unidades monetárias), para comparar a dispersão vamos recorrer ao coeficiente de variação.

Quantidade produzida:

$\bar{x} = \frac{4}{12} = 0.333$ dezenas

$s_x = \sqrt{\frac{3-12\times\left(\frac{4}{12}\right)^2}{12-1}} = 0.389$ dezenas

$CV_x = \frac{0.389}{0.333} \times 100\% = 116.81\%$

Custo da produção:

$\bar{y} = \frac{25}{12} = 2.083$ u.m.

$s_y = \sqrt{\frac{68.215-12\times\left(\frac{25}{12}\right)^2}{12-1}} = 1.211$ u.m.

$CV_y = \frac{1.211}{2.083} \times 100\% = 58.14\%$

Como $CV_x > CV_y$ logo os dados relativos à quantidade produzida apresentam uma maior dispersão do que os dados relativos ao custo de produção.

3. Variável aleatória discreta: $X =$ número de defeitos por automóvel

(a) 30% destes automóveis apresentam no mínimo 3 defeitos:

$$P(X \geq 3) = 0.30 \Leftrightarrow 1 - P(X < 3) = 0.30 \Leftrightarrow P(X \leq 2) = 0.7 \Leftrightarrow F(2) = 0.7 \Leftrightarrow k = 0.7$$

$$P(X < 3 \,|\, X \geq 1) = \frac{P(X < 3 \,\mathrm{e}\, X \geq 1)}{P(X \geq 1)} = \frac{P(1 \leq X < 3)}{1 - P(X < 1)} = \frac{P(0 < X \leq 2)}{1 - P(X \leq 0)} = \frac{F(2) - F(0)}{1 - F(0)} = \frac{0.7 - 0.25}{1 - 0.25} = 0.6$$

(b) A variável aleatória X segue uma distribuição Uniforme Discreta, $X \sim U_{(4)}$, pois $D_x = \{0, 1, 2, 3\}$ logo

$$F(2) = f(0) + f(1) + f(2) \Leftrightarrow k = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} = 0.75$$

Como o domínio da variável aleatória X são inteiros consecutivos, então

$$V[X] = \frac{(3 - 0 + 1)^2 - 1}{12} = 1.25$$

logo

$$V[1 - 3X] = (-3)^2\, V[X] = 9 \times 1.25 = 11.25$$

(c) Valor esperado da quantidade de CO_2 emitida por um automóvel deste modelo:

$$\begin{aligned} E[Y] &= \int_{-\infty}^{+\infty} y f(y)\, dy = \int_{-\infty}^{0} y \times 0 dy + \int_0^1 y \times y dy + \int_1^2 y \times (2 - y)\, dy + \int_2^{+\infty} y \times 0 dy = \\ &= 0 + \int_0^1 y^2 dy + \int_1^2 (2y - y^2)\, dy + 0 = \left[\frac{y^3}{3}\right]_0^1 + \left[y^2 - \frac{y^3}{3}\right]_1^2 = \\ &= \frac{1^3}{3} - 0 + 2^2 - \frac{2^3}{3} - \left(1^2 - \frac{1^3}{3}\right) = \frac{1}{3} + 4 - \frac{8}{3} - 1 + \frac{1}{3} = 1\,\mathrm{hg/km} \end{aligned}$$

4. -

(a) Seja X a variável aleatória discreta que representa o número de meteoritos que entram na atmosfera terrestre em certa região do globo, por hora, tal que $X \sim P(80)$ pois $E[X] = 80$.

Assim Y a variável aleatória discreta que representa o número de meteoritos que entram na atmosfera terrestre em certa região do globo em períodos de 15 minutos, é tal que $Y \sim P(20)$ pois $\lambda = \frac{15 \times 80}{60} = 20$.

$$P(Y \geq 25) = 1 - P(Y < 25) = 1 - P(Y \leq 24) = 1 - F(24) = 1 - 0.8432 = 0.1568$$

(b) Seja W a variável aleatória discreta que representa o número de meteoritos, em 15, que são visíveis a olho nu, sob a forma de estrela cadente, tal que $W \sim B(15, 0.10)$ pois $n = 15$ e $p = P(\textit{sucesso}) = P(\text{de um meteorito ser visível a olho nu}) = 0.10$. Pretende-se

$$P(W > 1) = 1 - P(W \leq 1) = 1 - F(1) = 1 - 0.5490 = 0.451.$$

(c) Seja V a variável aleatória contínua que representa o tempo, em minutos, entre entradas consecutivas de meteoritos na atmosfera, recorrendo à relação entre a distribuição Poisson e a distribuição Exponencial tem-se $V \sim Exp(0.75)$, pois $\theta = \frac{t}{\lambda} = \frac{60}{80} = 0.75$. A função de distribuição da variável aleatória V é dada por

$$F(v) = \begin{cases} 0 & , \quad v < 0 \\ \\ 1 - e^{-\frac{v}{0.75}} & , \quad v \geq 0 \end{cases}.$$

Pretende-se

$$P(V \geq 3) = 1 - P(V < 3) = 1 - F(3) = 1 - (1 - e^{-\frac{3}{0.75}}) = 0.0183$$