IPS Instituto Politécnico de Setúbal
Escola Superior de Tecnologia de Setúbal

DEPARTAMENTO DE MATEMÁTICA
MÉTODOS ESTATÍSTICOS
2.º Semestre - 2020/2021
Exame Época Normal

Data: 2 de julho de 2021 Duração: 2 horas e 30 minutos

Resolução

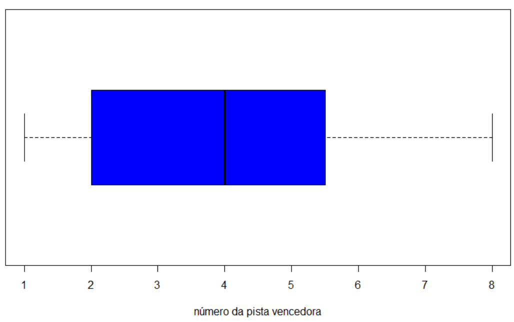
1. Variável: número da pista vencedora num circuito circular com 8 pistas.

   (a) Tabela de Frequências

i	Número da Pista x_i	Freq. Absoluta n_i	Freq. Relativa f_i	Freq. Abs. Acumulada N_i	Freq. Rel. Acumulada F_i
1	1	29	$\frac{29}{144} = 0.2014$	29	$\frac{29}{144} = 0.2014$
2	2	19	$\frac{19}{144} = 0.1319$	$29 + 19 = 48$	$\frac{48}{144} = 0.3333$
3	3	18	$\frac{18}{144} = 0.1250$	$48 + 18 = 66$	$\frac{66}{144} = 0.4583$
4	4	25	$\frac{25}{144} = 0.1736$	$66 + 25 = 91$	$\frac{91}{144} = 0.6319$
5	5	17	$\frac{17}{144} = 0.1181$	$91 + 17 = 108$	$\frac{108}{144} = 0.7500$
6	6	10	$\frac{10}{144} = 0.0694$	$108 + 10 = 118$	$\frac{118}{144} = 0.8194$
7	7	15	$\frac{15}{144} = 0.1042$	$118 + 15 = 133$	$\frac{133}{144} = 0.9236$
8	8	11	$\frac{11}{144} = 0.0764$	$133 + 11 = 144$	$\frac{144}{144} = 1$
		$n = 144$	1		

   (b) Para construirmos o diagrama de extremos e quartis, teremos de calcular o mínimo, o máximo, o 1º quartil, mediana e 3º quartil:

   - **Mínimo:** $min(x_i) = x_1 = 1$
   - **Máximo:** $max(x_i) = x_8 = 8$
   - **1º Quartil:** $Q_{0.25} = x_2 = 2$ pois $F_2 = 0.3333 > 0.25$
   - **Mediana:** $Q_{0.50} = x_4 = 4$ pois $F_4 = 0.6319 > 0.50$
   - **3º Quartil:** $Q_{0.75} = \frac{x_5+x_6}{2} = \frac{5+6}{2} = 5.5$ pois $F_5 = 0.75$



(c) Consideremos $X-$ número da pista vencedora num circuito circular de 8 pistas, com domínio $D_X = \{1, 2, 3, 4, 5, 6, 7, 8\}$. Pretende-se averiguar se determinadas pistas têm ou não mais hipóteses de vencer a corrida, ou seja, pretende-se averiguar se o número da pista vencedora não segue ou segue uma distribuição uniforme discreta. Assim:

Hipóteses a testar

$$\begin{cases} H_0 : X \sim U_{(8)} \rightarrow & \text{Os apostadores não têm razão} \\ vs & \\ H_1 : X \nsim U_{(8)} \rightarrow & \text{Os apostadores têm razão} \end{cases}$$

Dados

- Total de dados: $n = 144$
- Distribuição Uniforme discreta: $f(x) = P(X = x) = \frac{1}{8}$ se $x \in \{1, 2, 3, 4, 5, 6, 7, 8\}$
- Número de parâmetros a estimar: $r = 0$
- Nível de significância: $\alpha = 0.05$

Estatística de teste

$$Q = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(k-1-r)}$$

A estatística de teste, sob a hipótese H_0, tem distribuição Qui-Quadrado com

$$k - 1 - r = 8 - 1 - 0 = 7 \text{ graus de liberdade}$$

ou seja

$$Q \sim \chi^2_{(7)}$$

Construção da tabela de frequências observadas e esperadas

Domínio x_i	Frequências Observadas $O_i = n_i$	Probabilidade $p_i = f(x_i)$	Frequências Esperadas $E_i = n \times p_i$	Valor Observado da Estatística de Teste $\frac{(O_i-E_i)^2}{E_i}$
1	29	$f(1) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(29-18)^2}{18} = 6.7222$
2	19	$f(2) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(19-18)^2}{18} = 0.0556$
3	18	$f(3) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(18-18)^2}{18} = 0$
4	25	$f(4) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(25-18)^2}{18} = 2.7222$
5	17	$f(5) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(17-18)^2}{18} = 0.0556$
6	10	$f(6) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(10-18)^2}{18} = 3.5556$
7	15	$f(7) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(15-18)^2}{18} = 0.5000$
8	11	$f(8) = \frac{1}{8}$	$144 \times \frac{1}{8} = 18$	$\frac{(11-18)^2}{18} = 2.7222$
	$n = 144$	1	$n = 144$	$Q_{obs} = 16.3334$

Regra de Decisão através do valor-p

$$valor - p = P(Q \geq Q_{obs}) = P(Q \geq 16.3334) = 1 - P(Q < 16.3334) \underset{\text{v.a. contínua}}{=} 1 - F(16.3334) \approx$$
$$\approx 1 - F(16.0) = 1 - 0.975 = 0.025$$

Conclusão: Como $valor - p = 0.025 < 0.05 = \alpha$ então rejeita-se a hipótese H_0. Com base na amostra e ao nível de significância de 5%, X não segue distribuição $U_{(8)}$, pelo que os apostadores têm razão.

2. Consideremos

$$X - \text{ quilometragem de um carro usado (em milhares de quilómetros), com } x \in [9, 65]$$

e

$$Y - \text{ preço de venda (em centenas de euros), com } y \in [80, 300]$$

$n = 10$ (dimensão da amostra)
Como se pretende prever o preço de venda de um carro usado com base na quilometragem, tem-se:

$$X - \text{ variável independente}$$

e

$$Y - \text{ variável dependente}$$

Para verificar se a quilometragem explica o preço de venda é necessário calcular o coeficiente de correlação linear de Pearson:

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) \times \left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}} = \frac{\sum_{i=1}^{10} x_i y_i - 10 \times \overline{x}\,\overline{y}}{\sqrt{\left(\sum_{i=1}^{10} x_i^2 - 10\overline{x}^2\right) \times \left(\sum_{i=1}^{10} y_i^2 - 10\overline{y}^2\right)}} =$$

$$= \frac{39100 - 10 \times \frac{339}{10} \times \frac{1530}{10}}{\sqrt{\left(14931 - 10 \times \left(\frac{339}{10}\right)^2\right) \times \left(290700 - 10 \times \left(\frac{1530}{10}\right)^2\right)}} = -0.915$$

Uma vez que $r_{xy} = -0.915$, concluímos que existe uma correlação linear negativa forte, porque $-1 < r_{xy} < -0.8$, portanto a quilometragem explica o preço de venda. Ou seja, o modelo de regressão linear parece ser adequado aos dados.

Note-se que a confirmação da existência de uma correlação linear negativa forte entre X e Y deveria ser sempre acompanhada pelo diagrama de dispersão, mas neste caso não dispomos dos dados.

Pretende-se ainda determinar o preço de venda previsível para um carro usado com 45000 km. Para isso, precisamos de determinar a reta de regressão linear simples: $\widehat{y} = a + bx$. Assim

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{10} x_i y_i - 10\overline{x}\,\overline{y}}{\sum_{i=1}^{10} x_i^2 - 10\overline{x}^2} = \frac{39100 - 10 \times \frac{339}{10} \times \frac{1530}{10}}{14931 - 10 \times \left(\frac{339}{10}\right)^2} = -3.7125$$

e

$$a = \overline{y} - b\overline{x} = \frac{1530}{10} - (-3.7125) \times \frac{339}{10} = 278.8538$$

Logo a reta pretendida será

$$\widehat{y} = a + bx \Leftrightarrow \widehat{y} = 278.8538 - 3.7125x$$

E portanto determinar o preço de venda de um carro usado com $x = 45000$ km $= 45$ milhares de km será

$$\widehat{y}(45) = 278.8538 - 3.7125 \times 45 = 111.7913 \text{ centenas de euros}$$

Prevê-se que um carro com 45000 km seja vendido por aproximadamente 11179 euros.

No que diz respeito à previsão do preço de venda para um carro com 45000 km, esta previsão terá boa qualidade uma vez que 45 milhares de quilómetros está entre os valores observados pois estes situam-se entre 9 e 65 milhares de quilómetros e o modelo de regressão linear simples foi considerado adequado.

3. Seja X a variável aleatória que representa o consumo semanal de um certo bem alimentar, em dezenas de unidades, em famílias com filhos de idade inferior a 5 anos.

   (a) Tem-se 25 unidade = 2.5 dezenas de unidades e 10 unidades = 1 dezena de unidades, portanto pretende-se

$$P(X < 2.5|X \geq 1) = \frac{P(X < 2.5 \wedge X \geq 1)}{P(X \geq 1)} = \frac{P(1 \leq X < 2.5)}{1 - P(X < 1)} \underset{\text{v.a. contínua}}{=}$$

$$= \frac{F(2.5) - F(1)}{1 - F(1)} = \frac{\frac{2.5^3}{27} - \frac{1^3}{27}}{1 - \frac{1^3}{27}} = 0.5625$$

   (b) Para determinar o valor esperado de X precisamos de encontrar a função densidade probabilidade de X. Assim:

$$f(x) = F'(x) = \begin{cases} 0' & , x \leq 0 \\ \left(\frac{x^3}{27}\right)' & , 0 < x < 3 \\ 1' & , x \geq 3 \end{cases} \Leftrightarrow f(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{3x^2}{27} & , 0 < x < 3 \\ 0 & , x \geq 3 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow f(x) = \begin{cases} \frac{x^2}{9} & , 0 < x < 3 \\ 0 & , \text{caso contrário} \end{cases}$$

   Assim, temos:

$$E[X] = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_{-\infty}^{0} x \times 0 dx + \int_{0}^{3} x \times \left(\frac{x^2}{9}\right) dx + \int_{3}^{+\infty} x \times 0 dx =$$

$$= 0 + \int_{0}^{3} \left(\frac{1}{9}x^3\right) dx + 0 = \frac{1}{9} \times \left[\frac{x^4}{4}\right]_0^3 = \frac{1}{9} \times \left(\frac{3^4}{4} - 0\right) = 2.25 \text{ dezenas de unidades} = 22.5 \text{ unidades}$$

   (c) Como $Y = -1 - 2X$ e $E[X^2] = 5.4$ dezenas de unidades^2, tem-se

$$V[Y] = V[-1 - 2X] = (-2)^2 V[X] = 4 \times \left(E\left[X^2\right] - E^2[X]\right) = 4 \times \left(5.4 - 2.25^2\right) = 1.35$$

4. Seja X a variável aleatória discreta que representa o número de buracos por quilómetro, $X \sim P(6)$ pois $E[X] = 6 = \lambda$.

   (a) Seja Y a variável aleatória discreta que representa o número de buracos em 250 metros, $Y \sim P(1.5)$ pois

$$\begin{array}{lcl} 1 \text{ km} = 1000 \text{ metros} & \mapsto & \lambda = 6 \\ 250 \text{ metros} & \mapsto & \lambda_Y = \frac{250 \times 6}{1000} = 1.5 \end{array}$$

   Pretende-se

$$P(Y = 3) = f(3) = 0.1255$$

   (b) Seja D a variável aleatória contínua que representa a distância, em metros, percorrida por uma dada viatura até encontrar o primeiro buraco, recorrendo à relação entre a distribuição de Poisson e a distribuição Exponencial tem-se $D \sim Exp\left(\frac{500}{3}\right)$, pois $\theta = \frac{1000}{6} = \frac{500}{3}$. A função de distribuição da variável aleatória D é dada por

$$F(d) = \begin{cases} 0 & , \quad d < 0 \\ 1 - e^{-\frac{d}{\frac{500}{3}}} & , \quad d \geq 0 \end{cases}.$$

Pretende-se

$$P(D > 750|D \geq 200) \underset{(*)}{=} P(D > 750 - 200) = P(D > 550) =$$

$$= 1 - P(D \leq 550) = 1 - F(550) = 1 - \left(1 - e^{-\frac{550}{\frac{500}{3}}}\right) = 0.0369$$

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

(c) Seja W a variável aleatória contínua que representa o tempo (em minutos) que uma viatura leva a percorrer um troço de 10 quilómetros que está em obras, $W \sim N(9.5, \sqrt{2.1})$ pois $E[W] = \mu = 9.5$ minutos e $V[W] = \sigma^2 = 2.1 \Leftrightarrow \sigma = \sqrt{2.1}$ minutos. Portanto tem-se

$$W \sim N(9.5, \sqrt{2.1}) \Leftrightarrow Z = \frac{W - 9.5}{\sqrt{2.1}} \sim N(0, 1)$$

i. Seja S a variável aleatória discreta que representa o número de vezes que demora no mínimo 11 minutos a percorrer os 10 km, num conjunto de 7 vezes que esse percurso é feito, $S \sim B(7, 0.15)$ pois

$n = 7$ vezes que o percurso de 10 km é feito

$$p = P(W \geq 11) = 1 - P(W < 11) = 1 - P\left(Z < \frac{11 - 9.5}{\sqrt{2.1}}\right) = 1 - P(Z < 1.04) \underset{\text{v.a. contínua}}{=}$$

$$= 1 - \Phi(1.04) = 1 - 0.8508 = 0.15$$

Pretende-se

$$P(S \geq 4) = 1 - P(S < 4) \underset{\text{v.a. discreta}}{=} 1 - P(S \leq 3) = 1 - F(3) = 1 - 0.9879 = 0.0121$$

ii. Amostra: dimensão $n = 15$

Seja $\overline{T}$ - Tempo que em média 15 veículos levam a percorrer o percurso de 10 quilómetros. Pretende-se $P\left(\overline{T} \leq 9\right)$.

Como a População é Normal e $\sigma = \sqrt{2.1}$ conhecido, tem-se

$$Z = \frac{\overline{T} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

então

$$P\left(\overline{T} \leq 9\right) = P\left(Z \leq \frac{9 - 9.5}{\frac{\sqrt{2.1}}{\sqrt{15}}}\right) = P(Z \leq -1.34) = \Phi(-1.34) = 1 - \Phi(1.34) = 1 - 0.9099 = 0.0901$$

5. Amostra de dimensão $n = 40$ balcões de "teste de sabor"e em 26 desses locais a bebida com sabor a limão foi a preferida.

(a) População: $Y \sim$ Binomial, onde p representa proporção de público que tem preferência pelo sabor a limão.

Como tem-se: População Binomial e $n = 40 \geq 30$, o Intervalo de confiança a $(1 - \alpha) \times 100\%$ para p é:

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

Como grau de confiança $= 1 - \alpha = 0.90$, tem-se nível de significância $= \alpha = 0.10$ e

$$z_{1-\frac{\alpha}{2}} = z_{1-\frac{0.10}{2}} = z_{0.95} = 1.645$$

A proporção amostral é

$$p^* = \frac{26}{40} = 0.65 \qquad \text{e} \qquad q^* = 1 - p^* = 1 - 0.65 = 0.35$$

Então o Intervalo de confiança a 90% para p, a proporção de público que tem preferência pelo sabor a limão, é:

$$\left]0.65 - 1.645 \times \sqrt{\frac{0.65 \times 0.35}{40}}, 0.65 + 1.645 \times \sqrt{\frac{0.65 \times 0.35}{40}}\right[=]0.5259, 0.7741[$$

Com 90% de confiança, a percentagem de público que tem preferência pelo sabor a limão está entre 52.59% e 77.41%. Como o intervalo apenas contempla valores superiores a 50%, o que indica que com um grau de confiança a 90% o público prefere a bebida com sabor a limão.

(b) Pretende-se determinar o grau de confiança, $1 - \alpha$, tal que o erro de estimação = margem de erro é no máximo de 0.1:

$$\text{margem de erro} \leq 0.1 \Leftrightarrow \frac{\left(p^* + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}}\right) - \left(p^* - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}}\right)}{2} \leq 0.1 \Leftrightarrow$$

$$\Leftrightarrow z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^* \times q^*}{n}} \leq 0.1 \Leftrightarrow z_{1-\frac{\alpha}{2}}\sqrt{\frac{0.65 \times 0.35}{40}} \leq 0.1 \Leftrightarrow z_{1-\frac{\alpha}{2}} \leq 1.33 \Leftrightarrow$$

$$\Leftrightarrow 1 - \frac{\alpha}{2} \leq \Phi(1.33) \Leftrightarrow 1 - \frac{\alpha}{2} \leq 0.9082 \Leftrightarrow \alpha \geq 0.1836 \Leftrightarrow 1 - \alpha \leq 0.8164$$

ou seja, um grau de confiança menor ou igual a 81.64%.

(c) População: X- duração, em anos, da máquina industrial de engarrafamento, com $X \sim N(\mu, \sigma)$. Amostra: $n = 5$.

i. Estimativa pontual para σ^2 = variabilidade do tempo de vida da máquina industrial:

$$s^2 = \frac{(4.6 - 5.28)^2 + (4.3 - 5.28)^2 + (6.6 - 5.28)^2 + (4.7 - 5.28)^2 + (6.2 - 5.28)^2}{5 - 1} = 1.087 \text{ anos}^2$$

pois

$$\bar{x} = \frac{4.6 + 4.3 + 6.6 + 4.7 + 6.2}{5} = 5.28 \text{ anos}$$

ii. Hipóteses:

$$\begin{cases} H_0 : \sigma \geq 1.5 \rightarrow & \text{o desvio padrão do tempo de vida da máquina é no mínimo de 1.5 anos} \\ vs & \\ H_1 : \sigma < 1.5 \rightarrow & \text{o desvio padrão do tempo de vida da máquina é inferior a 1.5 anos} \end{cases}$$

tem-se

$$\begin{cases} H_0 : \sigma \geq 1.5 \\ vs \\ H_1 : \sigma < 1.5 \end{cases} \Leftrightarrow \begin{cases} H_0 : \sigma^2 \geq 1.5^2 \\ vs \\ H_1 : \sigma^2 < 1.5^2 \end{cases}$$

Estatística de Teste: População Normal e μ desconhecido

$$X^2 = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)} \Leftrightarrow X^2 \sim \chi^2_{(5-1)} \Leftrightarrow X^2 \sim \chi^2_{(4)}$$

Valor observado da Estatística de Teste, considerando a amostra recolhida e supondo a hipótese H_0 verdadeira:

$$X^2_{\text{obs}} = (5 - 1) \times \frac{1.087}{1.5^2} = 1.93$$

Região Crítica: o teste é Unilateral Esquerdo e nível de significância = $\alpha = 0.01$

$$RC = \left[0, x^2_{(n-1);\alpha}\right] = \left[0, x^2_{(4);0.01}\right] = [0, 0.297]$$

Como o $X^2_{\text{obs}} = 1.93 \notin RC$, então não se rejeita a H_0. Com base na amostra e para um nível de significância de 1%, há evidência estatística que o desvio padrão do tempo de vida da referida máquina industrial pode ser considerado no mínimo de 1.5 anos.