2ª Freq. + Exame de Época Normal

2014/2015

$\boxed{1.}$ X – nº anual de instituições científicas q̄ participam no OCJF

Y – nº " " alunos q̄ participam no OCJF

$n_1 = n_2 = 18$

a.i) $\bar{y} = \dfrac{\sum_{i=1}^{18} y_i}{n_2} = \dfrac{13214}{18} = 734,11$ – logo em média o nº de alunos q̄ participaram anualmente no OCJF é de cerca 734 alunos

$\tilde{y} = P_{50} = 747,50$ – portanto, em 50% dos anos, o nº anual de alunos q̄ participaram no OCYF é inferior ou igual a 747,5 alunos e nos restantes anos é superior ou igual a esse valor.

ii) $s_x = ?$

Como $s_x^2 = \dfrac{1}{n_1 - 1}\left(\sum_{i=1}^{18} x_i^2 - n_1 \bar{x}^2\right) = \dfrac{1}{17}\left(75409 - 18 \times \bar{x}^2\right)$

e $\bar{x} = \dfrac{\sum_{i=1}^{18} x_i}{n_1} = \dfrac{1091}{18} = 60,6111$, então tem-se que

$s_x^2 = \dfrac{1}{17}\left(75409 - 18 \times 60,6111^2\right) = 546,0178 \Rightarrow s_x = 23,367$

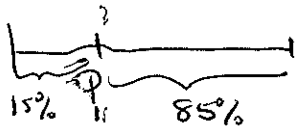
logo, o desvio típico em torno da média do nº anual de instituições científicas q̄ participam no OCYF é cerca de 23,367 instituições.

$max_x = ?$

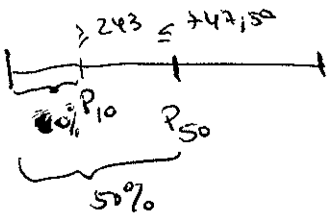
Como $a = máx - min \Leftrightarrow 83 = máx - 11 \Leftrightarrow máx = 94$

portanto, o nº máximo de instituições científicas q̄ participaram anualmente no OCYF foi 94 instituições

b. i) Em 85% dos anos, o nº anual de alunos ... foi $\geq$ a $318{,}90 \approx$ $\underline{319}$ alunos.



ii) Em cerca de $\underline{40}$ % dos anos ...



c) X – nº anual Inst. científicas

Assimetria : como $G_B = \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} = \dfrac{83{,}25 + 44 - 2\times 62}{39{,}25} =$

$= \dfrac{3{,}25}{39{,}25} = 0{,}083 > 0$ logo a distribuição do nº anual de Inst. científicas é ligeira/ assimétrica positiva.

Achatamento :

Como $k_p = \dfrac{\text{Kurtosis}}{\text{std. Error kurtosis}} = \dfrac{-0{,}454}{1{,}038} = -0{,}4374$, então $-1{,}96 < k_p = -0{,}4374 < 1{,}96$. Portanto, a distribuição do nº anual de Inst. científicas é mesocúrtica.

d) Como $CV_Y = 42{,}68\%$ e $CV_X = \dfrac{23{,}367}{60{,}6111} \times 100\% = 38{,}55\%$

então temos que $CV_X < CV_Y$, portanto podemos concluir que os dados relativos ao nº anual de instituições q participam no OCJF é mais homogéneo q o conj. de dados relativos ao nº anual de alunos. logo, a afirmação é verdadeira.

2. X - v.a.

x	$1-k$	$k-1$	k	$2k$
$f(x)$	p	$3p$	p	p

a) $E[X] = 8 \qquad p = ? \qquad k = ?$

Como $\sum_{i=1}^{4} f(x_i) = 1$ então tem-se $p + 3p + p + p = 1 \Leftrightarrow 6p = 1$

e $E(X) = 8 \Leftrightarrow \sum_{i=1}^{4} x_i f(x_i) = 8 \Leftrightarrow (1-k)p + (k-1)3p + kp + 2kp = 8$

$\Leftrightarrow (1 - k + 3k - 3 + k + 2k)\, p = 8 \Leftrightarrow (-2 + 5k)\, p = 8$

logo vem $\begin{cases} 6p = 1 \\ (-2+k5)p = 8 \end{cases} \Leftrightarrow \begin{cases} p = 1/6 \\ -2+5k = 48 \end{cases} \Leftrightarrow \begin{cases} p = 1/6 \\ k = \frac{50}{5} = 10 \end{cases}$

Portanto, tem-se

x	-9	9	10	20
$f(x)$	$1/6$	$1/2$	$1/6$	$1/6$

b) $\text{var}(X) = ?$

Por def. tem-se $\text{Var}(X) = E(X^2) - E^2(X) = ?$

Como $E(X^2) = \sum_{i=1}^{4} x_i^2 f(x_i) = (-9)^2 \times \frac{1}{6} + 9^2 \times \frac{1}{2} + 10^2 \times \frac{1}{6} + 20^2 \times \frac{1}{6} =$

$= \frac{81}{2} + \frac{1}{6}(81 + 100 + 400) = \underset{(\times 3)}{\frac{81}{2}} + \frac{581}{6} = \frac{824}{6} = \frac{412}{3}$, então

tem-se $\text{var}(X) = \frac{412}{3} - 8^2 = \frac{412}{3} - \underset{(\times 3)}{64} = \frac{412 - 192}{3} = \frac{220}{3} \simeq 73{,}3$

c) $P(X \leq x) \geq 0{,}75$

Como $P(X \leq -9) = P(X = -9) = 1/6$

$P(X \leq 9) = f(-9) + f(9) = \frac{1}{6} + \frac{1}{2} = \frac{4}{6} = \frac{2}{3} \simeq 0{,}7$

$P(X \leq 10) = P(X \leq 9) + P(X = 10) = \frac{4}{6} + \frac{1}{6} = \frac{5}{6} \simeq 0{,}8$

Portanto, tem-se que $x = 10$.

d) $Y = X^3$

Como $x = -9 \Rightarrow y = (-9)^3 = -9^3 \Rightarrow f_Y(-9^3) = 1/6$

Como $x = 9 \Rightarrow y = 9^3$ logo $f_Y(9^3) = 1/2$

$x = 10 \Rightarrow y = 10^3$ logo $f_Y(10^3) = 1/6$

$x = 20 \Rightarrow y = 20^3$ logo $f_Y(20^3) = 1/6$

Dado $f_{X,Y}(x,y) = P(X=x, Y=y)$ e como $Y = X^3$ então X e Y não são independentes, portanto tem-se

$f_{X,Y}(-9,9^3) = P(X=-9, Y=-9^3) = 1/6$

$f_{X,Y}(-9,9^3) = P(X=-9, Y=9^3) = 0 = f_{X,Y}(-9,10^3) = f_{X,Y}(-9,20^3)$

$f_{X,Y}(9;-9^3) = P(X=9, Y=-9^3) = 0 = f_{X,Y}(9,10^3) = f_{X,Y}(9,20^3)$

$f_{X,Y}(9,9^3) = P(X=9, Y=9^3) = 1/2$

$f_{X,Y}(10,-9^3) = P(X=10, Y=-9^3) = 0 = f_{X,Y}(10,9^3) = f_{X,Y}(10,20^3)$

$f_{X,Y}(10,10^3) = P(X=10, Y=10^3) = 1/6$

$f_{X,Y}(20,-9^3) = 0 = f_{X,Y}(20,9^3) = f_{X,Y}(20,10^3)$ e $f_{X,Y}(20,20^3) = 1/6$

Portanto, a função de probabilidade conjunta é definida por

X \ Y	-9^3	9^3	10^3	20^3
-9	1/6	0	0	0
9	0	1/2	0	0
10	0	0	1/6	0
20	0	0	0	1/6

3. X - tempo de vida de um computador

$X \sim N(\mu, \sigma^2 = 0{,}7225)$ $\quad \sigma^2 = 0{,}7225 \Rightarrow \sigma = 0{,}85$

a) $P(X > \frac{\mu}{2}) = 0{,}983 \Leftrightarrow 1 - P(X \leq \mu/2) = 0{,}983 \Leftrightarrow$

$\Leftrightarrow P(X \leq \mu/2) = 1 - 0{,}983 \Leftrightarrow P\left(Z \leq \frac{\mu/2 - \mu}{0{,}85}\right) = 0{,}017 \Leftrightarrow \phi\left(\frac{\mu/2}{0{,}85}\right) = 0{,}983 =$

$= \phi(2{,}12) \Leftrightarrow \frac{\mu/2}{0{,}85} = 2{,}12 \Leftrightarrow \frac{\mu}{2} = 1{,}802 \Leftrightarrow \mu = 3{,}604 \simeq 3{,}6$

logo, o tempo médio de vida dos computadores é 3,6 anos.

b. i) $P(1{,}5 \leq X \leq 4) = P\left(\frac{1{,}5 - 3{,}6}{0{,}85} \leq Z \leq \frac{4 - 3{,}6}{0{,}85}\right) =$

$= P(-2{,}47 \leq Z \leq 0{,}47) = \phi(0{,}47) - \phi(-2{,}47) =$

$= \phi(0{,}47) - 1 + \phi(2{,}47) = 0{,}6808 - 1 + 0{,}9932 = 0{,}674$

Não considerar!!

$P(X \leq 3{,}5) = P\left(Z \leq \frac{3{,}5 - 3{,}6}{0{,}85}\right) = P(Z \leq -0{,}18) = \phi(-0{,}18) =$

$= 1 - \phi(0{,}18) = 1 - 0{,}5714 = 0{,}4286$

ii) $P(X < 1) = P\left(Z < \frac{1 - 3{,}6}{0{,}85}\right) = P(Z < -3{,}06) = \phi(-3{,}06) =$

$= 1 - \phi(3{,}06) = 1 - 0{,}9989 = 0{,}0011$

c) $n = 25$

Y - v.a. q representa o nº de computadores substituídos

$Y \sim B(n = 25, p = 0{,}0011)$, pois dado 1 computador só há duas possibilidades, ou é substituído ou não é substituído. logo, $f(y) = P(Y = y) = {}^{25}C_y (0{,}0011)^y (1 - 0{,}0011)^{25-y}$, $y = 0, \ldots, 25$

$P(Y < 3) = P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) =$

$= {}^{25}C_0 (0{,}0011)^0 \times (1 - 0{,}0011)^{25} + {}^{25}C_1 (0{,}0011)^1 (1 - 0{,}0011)^{24} + {}^{25}C_2 (0{,}0011)^2 \times$

$(1-0{,}0011)^{23} = \cdot\ 1 \cdot 0{,}9989^{25} + 25 \times 0{,}0011 \times 0{,}9989^{24} + 300 \times 0{,}0011^2 \times$

$\times\ 0{,}9989^{23} \cong 0{,}97286 + 0{,}02678 + 0{,}00035 \cong 0{,}99999$

$\boxed{4.}$

$n_1 = 12$ X – % de indivíduos, com idades entre 16 e 74 anos, com acesso à internet em casa, em Itália

$n_2 = 12$ Y – % " ", com idades entre 16 e 74 anos, " " à internet em casa, em Portugal

a) Teste da normalidade dos dados (*)

C.A.:
$\alpha = 10\% = 0{,}10$
$\Downarrow$
$1-\frac{\alpha}{2} = 1-0{,}05 = 0{,}95$

Teste: $H_0: \mu = 50\%$ vs $H_1: \mu \neq 50\%$

Estatística de Teste: $T = \dfrac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} \sim t_{12-1} = t_{11}$

(*) (*) $H_0: Y \sim N(\mu_2, \sigma_2)$ vs $H_1: Y \not\sim N(\mu_2, \sigma_2)$

Regra de decisão: Rejeitar H_0 se p-value $\leq \alpha$.

Como $n = 12$, então aplica-se o Teste de Shapiro-Wilk, portanto

$p\text{-value}_{SW} = 0{,}312 > \begin{cases} \alpha = 0{,}01 \\ \alpha = 0{,}05 \\ \alpha = 0{,}10 \end{cases}$. Donde, aos níveis de significância usuais, não se rejeita H_0,

ou seja, assume que Y segue uma distribuição Normal.

Regra de decisão: Rejeitar H_0 se $|T_{obs}| > t_{11;0,95}$

Como $t_{11;0,95} = 1{,}796$ então R.A. $= [-1{,}796 ; 1{,}796]$ e

R.R. : $]-\infty ; -1{,}796[\cup]1{,}796, +\infty[$.

Dado que $T_{obs} = \dfrac{32{,}83 - 50}{4{,}651} = -3{,}692 \in$ R.R.

logo rejeita-se a hipótese H_0, ao nível de significância de 10%.

Ou) Como $I.C._{95\%}(\mu_2) =]22{,}60 ; 43{,}07[\supset I.C._{90\%}(\mu_2) \not\ni 50$

então, rejeita-se a hipótese H_0, ao nível de significância de 10%.
Portanto, existe evidência estatística para afirmar que a percentagem média de indivíduos com internet em casa em Portugal é diferente de 50%.

20 min

b) $\alpha = 5\% = 0{,}05 \Rightarrow 1-\alpha = 0{,}95$

Teste da normalidade para X:

$H_0: X \sim \mathcal{N}(\mu_1; \sigma_1)$ vs $H_a: X \not\sim \mathcal{N}(\mu_1; \sigma_1)$

Regra de decisão: Rejeitar H_0 se p-value $\leq \alpha$.

Como $n_1 = 12 < 30$, então aplica-se o Teste de Shapiro-Wilk.

Portanto, como $\text{p-value}_{S\text{-}W} = 0{,}206 > \begin{cases} \alpha = 0{,}01 \\ \alpha = 0{,}05 \\ \alpha = 0{,}10 \end{cases}$, então aos níveis de significância usuais, não se rejeita H_0, ou seja, assume que $X \sim \mathcal{N}(\mu_1; \sigma_1)$

Teste: $H_0: \sigma_1^2 \leq 144$ vs $H_1: \sigma_1^2 > 144$

Est. Teste: $\chi^2 = \dfrac{(n-1)\cdot s^2}{\sigma_0^2} = \dfrac{11 \times s^2}{144} \sim \chi^2_{12-1} = \chi^2_{11}$

Regra de decisão: Rejeitar H_0 se $\chi^2_{obs} > \chi^2_{11;0{,}95}$

Como $\chi^2_{11;0{,}95} = 19{,}68$, então R.A. $= [0; 19{,}68]$ e R.R. $=]19{,}68, +\infty[$. Dado que $\chi^2_{obs} = \dfrac{11 \times 12{,}281^2}{144} = 11{,}52$ $\in$ R.A., então não se rejeita a hipótese H_0, ou seja, ao nível de significância de 5%, não existe evidência estatística para afirmar que a variância da percentagem de indivíduos de Itália com acesso à internet excede 144.

c) $\text{p-value} = P(\chi^2 > \chi^2_{obs}) = P(\chi^2 > 11{,}52) = 1 - P(\chi^2 \leq 11{,}52)$
$= 1 - 0{,}6 = 0{,}4$

d) $\alpha = 10\% \Rightarrow 1-\alpha = 0{,}90$

Teste: $H_0: \mu_1 - \mu_2 = 0$ vs $H_1: \mu_1 - \mu_2 \neq 0$

Como X e Y seguem distribuições normais tem-se que se verificar se podemos assumir a igualdade das variâncias.

Teste de Levene:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad vs \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

Regra de decisão: Rejeitar Ho se $p\text{-value}_{T.\ Levene} \leq \alpha$.

Como $p\text{-value}_{T.L.} = 0,220 > \begin{cases} \alpha = 0,01 \\ \alpha = 0,05 \\ \alpha = 0,10 \end{cases}$, então não se rejeita Ho, ou seja, aos níveis de significância usuais, assume-se a igualdade das variâncias. Assim, tem-se que

Est. Teste:

$$T = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{\frac{11.S_1^2 + 11.S_2^2}{22}}\sqrt{\frac{1}{12} + \frac{1}{12}}} \sim t_{22}$$

Regra de decisão: Rejeitar Ho se $p\text{-value} \leq \alpha$.

Como $p\text{-value} = 0,778 > 0,10 = \alpha$, então não se rejeita Ho, ou seja, ao nível de significância de 10%, não existe evidência estatística para afirmar que existem diferenças significativas entre Itália e Portugal quanto à percentagem média de indivíduos com acesso à internet.

OU Regra de decisão: Rejeitar Ho se $|T_{obs}| > t_{22;\,0,90}$

Como $I.C._{95\%}(\mu_1 - \mu_2) =]-1,106 - \varepsilon\ ;\ 13,795[\ = (*)$

$13,795 = \bar{X} - \bar{Y} + \varepsilon \Leftrightarrow \varepsilon = 13,795 - (\bar{X} - \bar{Y}) = 13,795 - (3,545 - 4,651)$

$= 13,795 - (-1,106) = 14,901$

$(*) =]-16,007\ ;\ 13,795[\ \supset I.C._{90\%}(\mu_1 - \mu_2) \ni 0$, logo não se rejeita Ho.

5.

X - nº estágios por área científica

$k = 5$ - classes

a) $A = O_1 = ? = 33$

$O_1 - E_1 = -4,4 \Leftrightarrow O_1 = E_1 - 4,4 = 37,4 - 4,4 = 33.$

$B = E_2 = ? = O_2 - (O_2 - E_2) = 17 - (-4,8) = 21,8$

$C = E_3 = ? = n - (E_1 + E_2 + E_4 + E_5) = 104 - (37,4 + 21,8 +$

$+ 22,9 + 16,6) = 5,3 \Rightarrow D = O_3 - E_3 = 8 - 5,3 = 2,7$

$E = O_4 = n - O_1 - O_2 - O_3 - O_5 = 104 - 33 - 17 - 8 - 16 = 30$

$F = O_4 - E_4 = 30 - 22,9 = 7,1$

$G = df = k - 1 = 5 - 1 = 4$

b) Teste de ajustamento: $\alpha = 5\%$

$H_0: p_1 = 0,36 \wedge p_2 = 0,21 \wedge p_3 = 0,05 \wedge p_4 = 0,22 \wedge p_5 = 0,16$

vs

$H_1: p_1 \neq 0,36$ ou $p_2 \neq 0,21$ ou $p_3 \neq 0,05$ ou $p_4 \neq 0,22$ ou $p_5 \neq 0,16$

Estatística teste: $$\chi^2 = \sum_{i=1}^{5} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_4$$

Regra de decisão: Rejeitar H_0 se $\chi^2_{obs} > \chi^2_{4;0,95}$

Como $\chi^2_{obs} = 5,347$ e $\chi^2_{4;0,95} = 11,14$, então

$\chi^2_{obs} < \chi^2_{4;0,95}$, logo não se rejeita H_0, ou seja, ao nível de significância de 5%, não existe evidência estatística para afirmar que a distribuição do número de estágios por área científica sofreu alteração significativa.

ou

Regra de decisão: Rejeitar H_0 se p-value $\leq \alpha$.

Como p-value $= 0,253 > 0,05 = \alpha$, então não se rejeita H_0, ao nível de significância de 5%.

6. X - faixa etária $\quad X_1$ - até aos 25 anos

X_2 - mais de 25 anos $\quad L=2$

Y - classificação dos pacientes $\quad Y_1$ - Recuperado

Y_2 - $\bar{N}$ recup. mas em trata/

$c=3 \quad Y_3$ - $\bar{N}$ " e desistiu do trata/

a) o teste de hipóteses q permite concluir se o estado dos pacientes ao fim de 3 anos de trata/ é independente da idade é o teste de independência

H_0 : X e Y são independentes

vs

H_1 : X e Y não são independentes

b) Est. Teste :

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{1\times 2} = \chi^2_2$$

Regra de decisão : Rejeitar H_0 se $\left(\chi^2_{obs} > \chi^2_2\right)$ p-value $\leq \alpha$.

Como

		Y_1	Y_2	Y_3	Total
X	X_2	$O_{11}=56$ $E_{11}=58{,}3$	$O_{12}=72$ $E_{12}=60{,}2$	$O_{13}=52$ $E_{13}=61{,}5$	$180=O_{1\cdot}$
	X_1	$O_{21}=34$ $E_{21}=31{,}7$	$O_{22}=21$ $E_{22}=32{,}8$	$O_{23}=43$ $E_{23}=33{,}5$	$98=O_{2\cdot}$
	Total	$O_{\cdot 1}=90$	$O_{\cdot 2}=93$	$O_{\cdot 3}=95$	$n=278$

$$E_{12} = \frac{O_{1\cdot} \times O_{\cdot 2}}{n} = \frac{180 \times 93}{278} = 60{,}2 \qquad E_{22} = \frac{98 \times 93}{278} = 32{,}8$$

$$E_{13} = \frac{O_{1\cdot} \times O_{\cdot 3}}{n} = \frac{180 \times 95}{278} = 61{,}5 \qquad E_{23} = \frac{98 \times 95}{278} = 33{,}5$$

então Tem-se que

$$\chi^2_{obs} = \frac{(56-58{,}3)^2}{58{,}3} + \frac{(72-60{,}2)^2}{60{,}2} + \frac{(52-61{,}5)^2}{61{,}5} + \frac{(34-31{,}7)^2}{31{,}7} +$$

$$+ \frac{(21-32{,}8)^2}{32{,}8} + \frac{(43-33{,}5)^2}{33{,}5} = 0{,}0907 + 2{,}3130 + 1{,}4675 + 0{,}1669 + 4{,}2451 + 2{,}6940 = 10{,}977$$

Portanto, como $P(\chi^2 > \chi^2_{obs})$ = p-value = $P(\chi^2 > 10,977)$ =
$= 1 - P(\chi^2 \leq 10,977) = 1 - 0,995 = 0,005 < \alpha = 0,05$
então, rejeita-se a hipótese H_0, ou seja, ao nível de significância de 5%, existe evidência estatística para afirmar que o estado dos pacientes ao fim de 3 anos de trata/ não é independente da idade.