

Intervalos de confiança e testes de hipóteses (resolução de mais alguns exercícios de ficha 6)

7

Exercício 2

- a) X - v.a. que representa o valor do pH das águas analisadas
 X é normal (μ desconhecido; σ^2 desconhecido)
 Amostra: $n = 16$

- i) Construção do intervalo de confiança para a média populacional μ , com um grau de confiança de 99%.

$$1 - \alpha = 99\% \Rightarrow \alpha = 1\%$$

• Escolha da variável funcional

parâmetro a estimar: μ

Tipo população: Normal

dimensão amostra: $n = 16 \leq 30$

σ^2 desconhecido

$$T = \frac{\bar{X} - \mu}{\frac{s_e}{\sqrt{n}}} \sim t_{n-1}$$

• Determinação da média amostral e desvio padrão corrigido amostral

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{107,6}{16} = 6,725$$

$$s^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = \frac{730,64}{16} - 6,725^2 = 0,439375$$

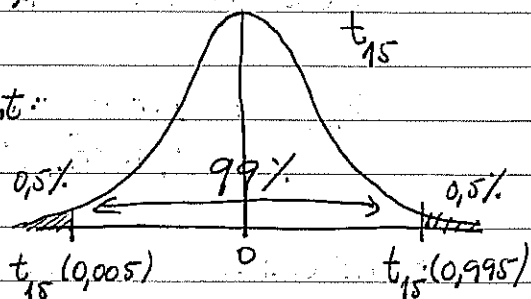
$$s_e^2 = \frac{n}{n-1} s^2 = \frac{16}{15} \times 0,439375 \approx 0,4682$$

• Obtenção dos quantis da dist. t de Student

Atuando de Tab. 5, $t_{15}(0,995) = 2,947$

Utilizando a simetria da dist. t de Student,

$$t_{15}(0,005) = -t_{15}(0,995) = -2,947$$



• Construção do I.C. para μ a 99% com base na amostra:

$$P\left(-2,947 \leq \frac{6,725 - \mu}{\sqrt{0,4687}/\sqrt{16}} \leq 2,947\right) = 0,99 \Leftrightarrow$$

$$\Leftrightarrow P\left(-2,947 \times \frac{\sqrt{0,4687}}{4} - 6,725 \leq -\mu \leq 2,947 \times \frac{\sqrt{0,4687}}{4} - 6,725\right) = 0,99 \Leftrightarrow$$

$$\Leftrightarrow P\left(-7,2294 \leq -\mu \leq -6,2206\right) = 0,99 \Leftrightarrow$$

$$\Leftrightarrow P\left(6,2206 \leq \mu \leq 7,2294\right) = 0,99$$

Assim $\mu \in [6,2206 ; 7,2294]$ a 99%

iii) • nível significância = erro de 1ª espécie = 5%

• Hipóteses do Teste para o parâmetro populacional μ

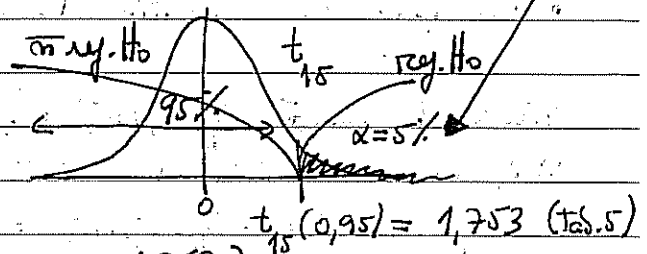
$H_0: \mu = 7$ contra $H_1: \mu > 7$ (teste unilateral direito)

• Escolha da variável fucional (ou estatística de teste) adequada:

$$T = \frac{\bar{X} - \mu}{s_e/\sqrt{n}} \sim t_{n-1}$$

• valor observado da variável fucional (supondo H_0 verdadeira)

$$t_{\text{obs}} = \frac{6,725 - 7}{\sqrt{0,4687}/\sqrt{16}} \approx -1,607$$



• determinar do quantil apropriado

• Região de rejeição $R = \left\{ u^{(16)} \in R^{16} : \frac{\bar{X} - 7}{\sqrt{0,4687}/\sqrt{16}} > 1,753 \right\}$

• Conclusão estatística (para $\alpha = 5\%$)

Como $t_{\text{obs}} \notin R$, não existe evidência estatística para $\text{rej. } H_0$.

b) Admitamos que $X \sim \text{Normal}(\mu=7; \sigma^2=9,04)$

$$\begin{aligned} \text{i)} \quad P(6,5 < X < 7,1) &= P\left(\frac{6,5-7}{\sqrt{9,04}} < Z < \frac{7,1-7}{\sqrt{9,04}}\right) = P(-2,5 < Z < 0,5) = \\ &= F(0,5) - F(-2,5) = F(0,5) - 1 + F(2,5) = \\ &= 0,6915 - 1 + 0,9938 = 0,6853 \end{aligned}$$

tab.4

ii) Atendendo à pergunta é necessário uma nova variável aleatória,

Y - n.º. que representa o n.º análises, em 25, em que $\text{pH} < 7$

$Y \sim \text{Binomial}(n=25; p=0,5)$

Justificação: $p = P(X < 7) = 0,5$
Med = $\mu = 7$

$$P(Y=18) = \binom{25}{18} \cdot 0,5^{18} \cdot 0,5^7 = 0,014325976 \approx 0,01433$$

ou $P(Y=18) = \text{binompdf}(25, 0,5, 18) \approx 0,01433$

Exercício 3

Amostra: $n = 31$ dias (mês Julho)

$$\sum_{i=1}^{31} T_i = 907,5^\circ\text{C} \quad \text{e} \quad \sum_{i=1}^{31} T_i^2 = 27146,95 (^\circ\text{C})^2$$

proveniente de uma pop. Normal

a) $H_0: \mu = 31$ contra $H_1: \mu = 30$ ($\mu < 31$) teste unilateral
esquerdo

• Escolha da variável funcional

parâmetro a estimar - μ

Tipo população - Normal

dimensão amostra: $n = 31 > 30$

σ^2 desconhecido

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

- Determinação da média e variância amostral

$$\bar{x} = \frac{1}{31} \sum_{i=1}^{31} x_i = \frac{907,5}{31} \approx 29,274^\circ\text{C}$$

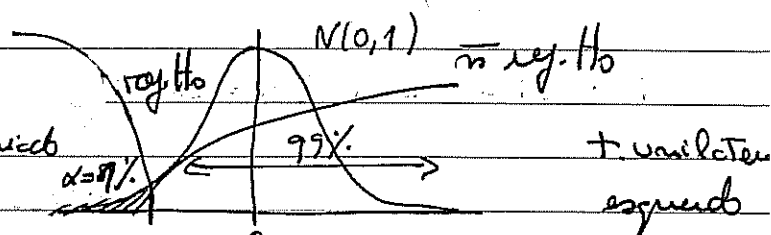
$$s^2 = \frac{\sum_{i=1}^{31} x_i^2}{n} - \bar{x}^2 = \frac{27146,95}{31} - 29,274^2 \approx 18,74$$

- valor observado da variável fatorial (ou estatística de teste)
(supondo H_0 verdadeiro)

$$z_{\text{observ}} = \frac{29,274 - 31}{\sqrt{18,74}/\sqrt{31}} \approx -2,22$$

Supondo $\alpha = 1\%$

- determinação do quantil apropriado



$$-2,33 \approx -z_{0,99} = z_{0,01}$$

tab. 4

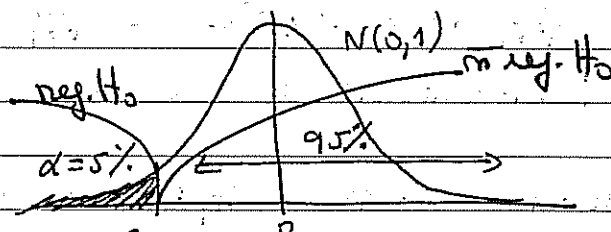
- Região de rejeição $\mathcal{R} = \left\{ \mathbf{x}^{(31)} \in \mathbb{R}^{31} : \frac{\bar{x} - 31}{s/\sqrt{31}} < -2,33 \right\}$ para $\alpha = 1\%$

- Conclusão estatística ($\alpha = 1\%$)

Como $z_{\text{observ}} \notin \mathcal{R}$, logo não se rejeita a hipótese H_0 .

Supondo $\alpha = 5\%$

- determinação do quantil



$$-1,645 \approx -z_{0,95} = z_{0,05}$$

- Região de rejeição $\mathcal{R} = \left\{ \mathbf{x}^{(31)} \in \mathbb{R}^{31} : \frac{\bar{x} - 31}{s/\sqrt{31}} < -1,645 \right\}$ para $\alpha = 5\%$

- Conclusão estatística ($\alpha = 5\%$)

Como $z_{\text{observ}} \in \mathcal{R}$, logo rejeita-se H_0 pelo que existe evidência estatística significativa para rejeitar H_0 .

b) p-value ou valor p é o menor nível de significância α que leva à rejeição da hipótese H_0

$$p\text{-value} = P(\text{rej. } H_0 \mid H_0 \text{ verd}) = \text{normsldf}(-\infty, -2.22) \approx 0,01321$$

máq.

Interpretação: para níveis de significância (α) superiores ou iguais a 1,321% devemos rejeitar a hipótese nula (H_0) (ou seja rej. H_0 para $\alpha \geq 1,321\%$)

c) Construção do intervalo de confiança para a variância da população, com um grau de confiança de 98%.

$$1-\alpha = 0,98 \Rightarrow \alpha = 2\%$$

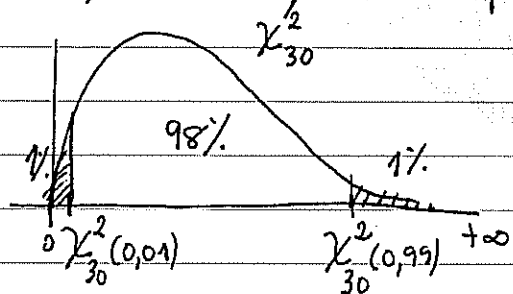
- Escolha da variável funcional
parâmetro a estimar: σ^2
Tipo população: Normal
dimensão amostral: $m = 31$
 σ^2 desconhecido

$$U = \frac{(m-1) S_e^2}{\sigma^2} \sim \chi_{m-1}^2$$

- cálculo da variância amostral corrigida

$$s_e^2 = \frac{m}{m-1} s^2 = \frac{31}{30} \times 18,74 \approx 19,36$$

- obtenção dos quantis apropriados da dist. do Qui-Quadrado com $m-1$ graus de liberdade



Consultando a Tabela 6 do Qui-Quadrado

$$\chi_{30}^2(0,01) = 14,95$$

$$\chi_{30}^2(0,99) = 50,89$$

- Construção do I.C. para σ^2 a 98% com base na amostra:

$$P\left(14,95 \leq \frac{(31-1) \times 19,36}{\sigma^2} \leq 50,89\right) = 0,98 \quad \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{14,95}{30 \times 19,36} \leq \frac{1}{\sigma^2} \leq 50,89\right) = 0,98 \quad \Leftrightarrow$$

$$\Leftrightarrow P\left(\frac{30 \times 19,36}{50,89} \leq \sigma^2 \leq \frac{30 \times 19,36}{14,95}\right) = 0,98 \quad \Leftrightarrow$$

$$\Leftrightarrow P(11,413 \leq \sigma^2 \leq 38,85) = 0,98$$

Assim $\sigma^2 \in [11,413 ; 38,85]$ a 98%.

d) Almea suplementar

$$H_0: \sigma^2 = 10 \text{ contra } H_1: \sigma^2 \neq 10, \quad \alpha = 2\%$$

(teste bilateral)

Uma vez que se trata de um teste bilateral e $\alpha = 2\%$, posso utilizar o I.C. construído na almea anterior.

Como $\sigma^2 = 10 (H_0) \notin [11,413 ; 38,85]$ a 98%, rejeita-se a hipótese nula, pelo que existe evidência estatística de que a variância desta população não é igual a $10(^{\circ}\text{C})^2$.

Exercício 4

X - v.a. que representa o consumo mensal de água, de uma família de quatro elementos (em m^3)

Amostra: $n = 12$

proveniente de uma pop. normal (μ desconhecido, σ^2 desconhecido)

- a) Construção do intervalo de confiança, com grau de confiança de 95%, para o desvio padrão do consumo mensal de água
 $1 - \alpha = 0,95 \Rightarrow \alpha = 5\%$ parâmetro desvio padrão da população

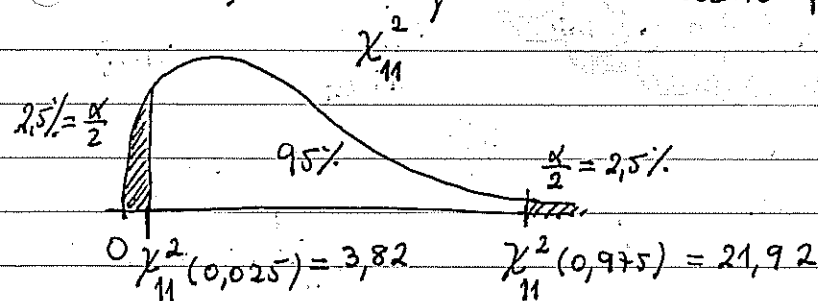
- Escolha da variável fatorial

$$\left. \begin{array}{l} \text{parâmetro } \sigma^2 \\ \text{pop. Normal} \\ n = 12 \text{ (qualquer)} \\ \mu \text{ e desconhecido} \end{array} \right\} U = \frac{(n-1) S_c^2}{\sigma^2} \sim \chi_{n-1}^2$$

- Cálculo da variância amostral corrigida

$$s_c^2 = \frac{\sum_{i=1}^{12} v_i^2}{12} - \frac{\left(\sum_{i=1}^{12} v_i \right)^2}{12 \times 12} = \frac{2065}{11} - \frac{155^2}{12 \times 11} \approx 5,7197$$

- Obtenção dos quantis de ordem apropriados da dist. do Qui-Quadrado com $n-1$ graus de liberdade (tabel. 6)



- Construção do I.C. para σ a 95% com base na amostra:

$$P\left(3,82 \leq \frac{(12-1) \times 5,7197}{\sigma^2} \leq 21,92 \right) = 0,95 \Leftrightarrow$$

$$P\left(\frac{3,82}{11 \times 5,7197} \leq \frac{1}{\sigma^2} \leq \frac{21,92}{11 \times 5,7197} \right) = 0,95 \Leftrightarrow P\left(2,8703 \leq \sigma^2 \leq 16,47 \right) = 0,95$$

Assim $\sigma^2 \in [2,8703 ; 16,47]$ a 95%.

pelo que $\sigma \in [\sqrt{2,8703} ; \sqrt{16,47}]$ a 95%.

$\sigma \in [1,694 ; 4,058]$ a 95%.

b) $H_0: \sigma = 1$ contra $H_1: \sigma \neq 1$ (teste bilateral) $\alpha = 5\%$

Através da alínea anterior:

Como $\sigma = 1 \notin$ I.C. para σ a 95%, rejeita-se H_0 , pelo que existe evidência estatística de que a variância populacional é diferente de 1.

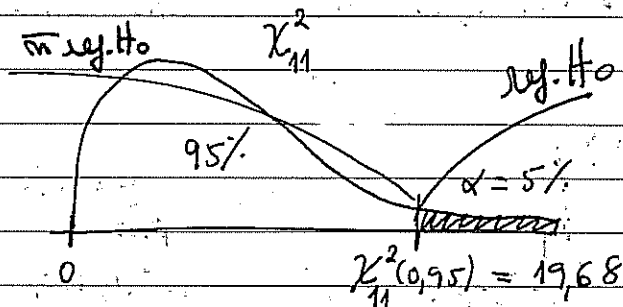
c) $H_0: \sigma = 2$ contra $H_1: \sigma > 2$

ou
 $H_0: \sigma^2 = 4$ contra $H_1: \sigma^2 > 4$ Teste unilateral direito $\alpha = 5\%$

- Escolha da variável fulcral (a mesma da alínea a)
- Cálculo da variância amostral corrigida (feita alínea a)
- valor observado da variável fulcral (supondo H_0 verdadeira)

$$U_{\text{observ}} = \frac{11 \times 5,7197}{4} \approx 15,729$$

- quantil de ordem apropriada da dist. do Qui-Quadrado com 11 g.l. (consulte tabela 6)



- Região de rejeição $R_0 = \{12^{(12)} \in 12^{12} : \frac{11 \times S_e^2}{4} > 19,68\}$ para $\alpha = 5\%$

- conclusão estatística ($\alpha = 5\%$)

Como $U_{\text{observ}} \notin R_0$, não se rejeita H_0 , pelo que não existe evidência estatística de que a variância da população é superior a 4.

d) $H_0: \sigma^2 = 3$ contra $\sigma^2 < 3$ teste unilateral esquerda $\alpha = 2,5\%$

Supõe-se que $\mu = 13,5 \text{ m}^3$ (segundo enunciado)

• Escolha da variável fatorial

$$\left. \begin{array}{l} \text{parâmetro } \sigma^2 \\ \text{população Normal} \\ n = 12 \text{ (qualquer)} \\ \mu \text{ é conhecida } (\mu = 13,5 \text{ m}^3) \end{array} \right\} U = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_m^2$$

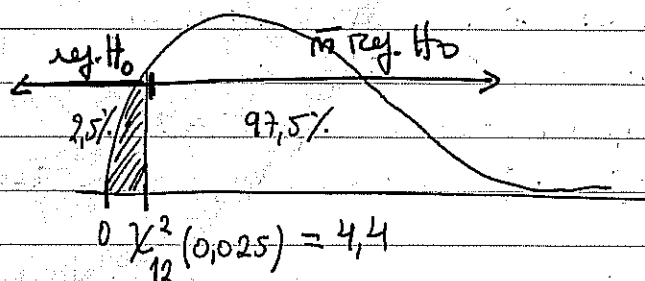
• Relativamente à amostra fornecida, cálculo de:

$$\begin{aligned} \sum_{i=1}^{12} (X_i - \mu)^2 &= \sum_{i=1}^{12} (X_i - 13,5)^2 = \sum_{i=1}^{12} (X_i^2 - 2 \times 13,5 \times X_i + 13,5^2) = \\ &= \sum_{i=1}^{12} X_i^2 - 27 \cdot \sum_{i=1}^{12} X_i + \sum_{i=1}^{12} 182,25 = \\ &= 2065 - 27 \times 155 + 12 \times 182,25 = 67 \end{aligned}$$

• valor observado da variável fatorial (supondo H_0 verdadeira)

$$U_{\text{obs}} = \frac{67}{3} \approx 22,33$$

• quantil de ordem apropriada da dist. do Qui-Quadrado com 12 g.l. (consulta tabela 6)



• região de rejeição $R = \{u^{(12)} \in \mathbb{R}^k : \sum_{i=1}^{12} (X_i - 13,5)^2 / 3 < 4,4\} \quad \alpha = 2,5\%$

• conclusão estatística ($\alpha = 2,5\%$)

Como $U_{\text{obs}} \notin R$ não se rejeita a hip. H_0 para $\alpha = 2,5\%$, pelo que não existe evidência estatística de que a variância da população é inferior a 3.

Exercício 6

a) Rio Tejo (T)

$$s_{\text{Tejo}}^2 = \frac{10202,46}{6} - \left(\frac{246,6}{6}\right)^2 = 11,2 \text{ (mg/ml)}^2$$

$$s_{c, \text{Tejo}}^2 = \frac{6}{5} \times 11,2 = 13,44 \text{ (mg/ml)}^2 \text{ (não é pedido)}$$

Rio Mira (M)

$$s_{\text{Mira}}^2 = \frac{8266,41}{5} - \left(\frac{161,5}{5}\right)^2 = 9,992 \text{ (mg/ml)}^2$$

$$s_{c, \text{Mira}}^2 = \frac{5}{4} \times 9,992 = 12,49 \text{ (mg/ml)}^2 \text{ (não é pedido)}$$

b) $H_0: \mu_T = \mu_M$ contra $H_1: \mu_T \neq \mu_M$, $\alpha = 5\%$

ou $H_0: \mu_T - \mu_M = 0$ contra $H_1: \mu_T - \mu_M \neq 0$ Teste bilateral

• variável fulcral

parâmetros a estimar: $\mu_1 - \mu_2$

populações Normais

$n_T = 6 \leq 30$ e $n_M = 5 \leq 30$

σ_T^2 e σ_M^2 desconhecidos, mas supõe-se $\sigma_T^2 = \sigma_M^2$

$$T = \frac{(\bar{X}_T - \bar{X}_M) - (\mu_T - \mu_M)}{\sqrt{\frac{1}{n_T} + \frac{1}{n_M}} \sqrt{\frac{n_T \cdot s_T^2 + n_M \cdot s_M^2}{n_T + n_M - 2}}} \sim t_{n_T + n_M - 2}$$

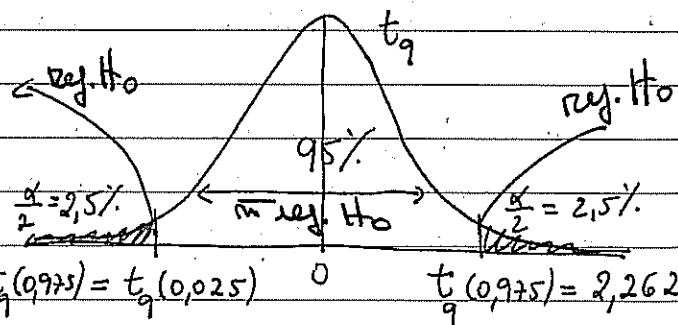
• valor observado da variável fulcral (supondo H_0 verdadeira)

$$T_{\text{obs}} = \frac{\left(\frac{246,6}{6} - \frac{161,5}{5}\right) - 0}{\sqrt{\frac{1}{6} + \frac{1}{5}} \sqrt{\frac{6 \times 11,2 + 5 \times 9,992}{6 + 5 - 2}}} \approx 4,028$$

• quantil de ordem apropriada da dist. t-Student com 9 g.l.
(consultar Tabela 5)

$$-2,262 = -t_9(0,975) = t_9(0,025)$$

$$t_9(0,975) = 2,262$$



• regra de rejeição $R = \{x^{(n)} \in \mathbb{R}^n : |T| > 2,262\}$ para $\alpha = 5\%$

$$= \{x^{(n)} \in \mathbb{R}^n : T < -2,262\} \cup \{x^{(n)} \in \mathbb{R}^n : T > 2,262\}$$

- conclusão estatística ($\alpha = 5\%$)

Como $t_{\text{observ}} \in \mathcal{R}$, rejeita-se H_0 , pelo que existe evidência estatística de que as médias populacionais dos rios Tejo e Mina são diferentes.

e) $H_0: \mu_T - \mu_M = 0$ contra $H_1: \mu_T - \mu_M \neq 0$ Teste Bilateral, $\alpha = 5\%$

- Escolha variável fulcral (a mesma da alínea b))
- cálculo das variáveis amostriais (alínea a))
- obtenção dos quantis (alínea b))

- Construção de I.E. para $\mu_T - \mu_M$ a 95% com base na amostra:

$$P\left(-2,262 \leq \frac{\left(\frac{246,6}{6} - \frac{161,5}{5}\right) - (\mu_T - \mu_M)}{\sqrt{\frac{1}{6} + \frac{1}{5}} \cdot \sqrt{\frac{6 \times 11,2 + 5 \times 9,92}{6+5-2}}} \leq 2,262\right) = 0,95 \Leftrightarrow$$

$$\Leftrightarrow P(-13,74193 \leq -(\mu_T - \mu_M) \leq -3,85807) = 0,95 \Leftrightarrow$$

$$\Leftrightarrow P(3,85807 \leq \mu_T - \mu_M \leq 13,74193) = 0,95$$

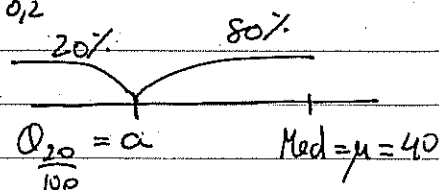
Assim $\mu_T - \mu_M \in [3,85807; 13,74193]$, a 95%.

(Conclusão estatística do Teste de hipóteses ($\alpha = 5\%$))

Como $\mu_T - \mu_M = 0$ (H_0) $\notin$ I.E. para $\mu_T - \mu_M$ a 95%, rejeita-se H_0 .
(igual conclusão b)).

d) X-v.a. que representa a concentração anual de estânio no rio Tejo.
 $X \sim \text{normal}(\mu = 40; \sigma^2 = 11)$

$$i) \text{ vigésimo percentil} = Q_{\frac{20}{100}} = Q_{\frac{2}{10}} = Q_{\frac{1}{5}} = Q_{0,2} = a$$



$$P(X \leq a) = 0,2 \Leftrightarrow P\left(Z \leq \frac{a-40}{\sqrt{11}}\right) = 0,2 \Leftrightarrow F\left(\frac{a-40}{\sqrt{11}}\right) = 0,2 \Leftrightarrow$$

z negativo

$$\Leftrightarrow 1 - F\left(-\frac{a-40}{\sqrt{11}}\right) = 0,2 \Leftrightarrow F\left(-\frac{a-40}{\sqrt{11}}\right) = 0,8 \Leftrightarrow$$

z positivo

$$\Leftrightarrow -\frac{a-40}{\sqrt{11}} \approx 0,84 \Leftrightarrow a \approx 37,214 \text{ mg/ml}$$

Interpretação: Em 20% das medições realizadas anualmente, a concentração de estômico tem o valor máximo de 37,214 mg/ml

iii) Para responder a esta alínea é necessário definir uma nova variável aleatória

Y - v.a. que representa o n.º de anos, em 100, em que se regista uma concentração anual de estômico entre 30 e 45 mg/ml no rio Tejo

$Y \sim \text{Binomial}(n=100; p=?)$

$$p = P(30 < X < 45) = P\left(\frac{30-40}{\sqrt{11}} < Z < \frac{45-40}{\sqrt{11}}\right) = P(-3,02 < Z < 1,51) =$$

$$= F(1,51) - F(-3,02) = F(1,51) - 1 + F(3,02) = 0,9345 - 1 + 0,9987 = 0,9332$$

Assim $Y \sim \text{Binomial}(n=100; p=0,9332) \approx N \sim \text{Normal}(\mu=93,32; \sigma^2=6,23)$

$\uparrow$
 $n > 20 \text{ e } np > 7$

$$P(Y \geq 72) \underset{\text{e.e.}}{\approx} P(N \geq 71,5) = P\left(Z \geq \frac{71,5 - 93,32}{\sqrt{6,234}}\right) = P(Z \geq -8,74) \approx 0,999$$

ou

$$P(Y \geq 72) = 1 - P(Y < 72) = 1 - P(Y \leq 71) =$$

$$= 1 - \text{binomcdf}(100, 0,9332, 71) = 1 - 9,12718 \times 10^{-12} \approx 1$$

mg.