33. Seja X – v.a. (população) que rep. a altura (em cm) de um estudante escolhido ao acaso na universidade em questão; sabe-se que $X \cap \mathcal{N}(\mu, \sigma)$, sendo μ e σ ambos desconhecidos.

Observada um amostra aleatória (a.a.) de 25 estudantes (escolheram-se aleatoriamente 25 estudantes no universo de todos os estudantes da universidade e mediram-se as suas alturas), obtiveram-se as características amostrais $\bar{x} = 170$ e $s = 5$ e, com base nestas, pretende-se determinar um intervalo para μ com 99% de confiança – $IC_{99\%}(\mu)$.

Sendo que:

- a distribuição da população em estudo é <u>Normal</u> e
- o desvio-padrão (equivalentemente, a variância) populacional, <u>σ</u>, <u>é desconhecido</u>,

o intervalo com $100(1-\alpha)\%$ de confiança para o valor esperado da população, com base numa amostra observada de dimensão n, é o seguinte:

$$IC_{100(1-\alpha)\%}(\mu) = \left(\bar{x} \mp t_{1-\alpha/2;n-1}\, s/\sqrt{n} \right),$$

onde $t_{1-\alpha/2;n-1}$ representa o quantil de probabilidade $1-\alpha/2$ da distribuição t de Student com $n-1$ graus de liberdade (g.l. ou d.f. do inglês *degrees of freedom*). No caso presente, pretende-se que $100(1-\alpha)\%$ seja igual a 99% e tem-se $n = 25$:

$100(1-\alpha) = 99 \Leftrightarrow (1-\alpha) = 0.99 \Leftrightarrow \alpha = 0.01 \Leftrightarrow \alpha/2 = 0.005 \Leftrightarrow 1-\alpha/2 = 0.995$

$t_{0.995;24} = 2.797$ (valor obtido na tabela de quantis da distribuição t de Student, nas calculadoras que têm essa funcionalidade ou numa folha de cálculo computacional)

$t_{1-\alpha/2;n-1}\, s/\sqrt{n} = 2.797 \times 5/\sqrt{25} = 2.797$

Então,

$$IC_{99\%}(\mu) = (170 - 2.797,\ 170 + 2.797) = (\underline{167.203},\ \underline{172.797})$$

34. Para um automobilista escolhido ao acaso, na população de automobilistas que adquiriram viatura há 12 meses (1 ano), considere-se o acontecimento que representa um "sucesso"

S: "o condutor percorreu mais de 10000Km"

Supondo que se pretende observar uma a.a. que inclui n condutores nas condições acima, então o n.º daqueles para os quais se observará a realização de S é uma v.a., que se pode definir como:

X – v.a. (população) que representa o n.º de condutores, em n, que constituem um sucesso.

Admitindo que os sucessos, de entre os n condutores, se realizam independentemente uns dos outros (o que faz sentido já que os condutores são escolhidos aleatoriamente na população), tem-se que

$$X \cap Bi(n, p)$$

onde $p = P(S)$, ou seja, p é a probabilidade de um condutor percorrer mais de 10000Km durante o primeiro ano após adquirir carro.

a) Já se sabe, pelo TLC, que a f.d. da v.a. X, centrada e reduzida, está tão mais próxima da f.d. da Normal padrão quanto mais elevado for n. Ou seja, uma vez que $E(X) = np$ e $Var(X) = np(1-p)$, para valores de n "suficientemente" grandes, tem-se

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \le x \right) \approx \Phi(x),$$

ou, equivalentemente (dividindo por n ambos os membros da desigualdade dentro da probabilidade),

$$P\left(\frac{X/n - p}{\sqrt{p(1-p)/n}} \le x \right) \approx \Phi(x).$$

Designando o quantil de probabilidade α da normal padrão por z_α (como habitualmente), verifica-se, então que

$$P\left(-z_{1-\alpha/2} \leq \frac{X/n - p}{\sqrt{p(1-p)/n}} \leq z_{1-\alpha/2}\right) \approx 1-\alpha.$$

Denotando $\hat{p} = X/n$, $\hat{p}$ é o <u>estimador</u> "natural" de p, já que p é a proporção de sucessos em toda a população de automobilistas nas condições em estudo e $\hat{p}$ é a proporção equivalente mas na a.a. de n automobilistas. Ora, se n for "suficientemente" grande para que se possa efectuar a aproximação anterior, também se espera que seja "suficientemente" grande para que as estimativas obtidas através de $\hat{p}$ já estejam próximas de p (quanto maior a dimensão da amostra mais perto se está de observar toda a população), o que leva a que

$$\sqrt{p(1-p)/n} \approx \sqrt{\hat{p}(1-\hat{p})/n}\ .$$

Tendo este raciocínio por base (que se demostra estar certo, do ponto de vista probabilístico, mas cuja demonstração está totalmente fora do âmbito desta U.C.), pode escrever-se então que

$$P\left(-z_{1-\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq z_{1-\alpha/2}\right) \approx 1-\alpha.$$

Resolvendo as inequações, entre parêntesis, em ordem a p, obtém-se

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 1-\alpha,$$

ou seja, o intervalo aleatório

$$\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

contém a probabilidade (ou proporção) p com uma probabilidade aproximadamente igual a $1-\alpha$.

Se x for um valor observado da v.a. X e, consequentemente, $\hat{p} = x/n$ uma <u>estimativa</u> de p, então um intervalo (assintótico) com $100(1-\alpha)\%$ de confiança (aproximadamente) para p será (utilizando a já habitual notação abreviada intervalos):

$$IC_{100(1-\alpha)\%}(p) = \left(\hat{p}\ \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Concretizando para a amostra recolhida pela Motorpress, com $n = 100$ e $x = 87$, onde se pretende uma confiança de 95%, tem-se:

$\hat{p} = x/n = 87/100 = 0.87$

$1-\alpha = 0.95 \Leftrightarrow \alpha = 0.05 \Leftrightarrow \alpha/2 = 0.025 \Leftrightarrow 1-\alpha/2 = 0.975 \Rightarrow z_{1-\alpha/2} = z_{0.975} = 1.96$

$z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = 1.96 \times \sqrt{0.87 \times 0.13}/10 \approx 0.066$

Então,

$$IC_{95\%}(p) = (\ 0.87 - 0.066, 0.87 + 0.066) = (\ \underline{0.804}, \underline{0.936})$$

b) *** Excluído do Programa ajustado ***

35. Para uma árvore escolhida ao acaso na região em questão, seja p a probabilidade de que o seu volume de madeira seja inferior a 12m^3 – "sucesso" (equivalentemente, p é a proporção de árvores da região que são sucessos). Uma vez que se recolheu uma a.a. de 40 árvores, com o intuito de estudar a verdadeira proporção de árvores da região que são sucessos, considere-se:

X – v.a. (população) que representa o n.º de árvores, em 40 seleccionadas ao acaso na região, com volume de madeira inferior a 12m^3 (n.º de sucessos em 40 provas). Admitindo que as árvores são sucessos, ou não, independentemente umas das outras (?!...), tem-se que X $\cap$ Bi(40, *p*).

Para uma a.a. de dimensão *n,* um estimador para *p* é dado por $\hat{p}$ = X/*n* e, sendo x um valor observado de X, uma estimativa de p será dada por $\hat{p}$ = x/*n*.
A amostra que foi observada conduz a um valor observado de X igual a x = 34, pelo que se estima a proporção de árvores que são sucessos por

$$\hat{p} = 34/40 = \underline{0.85}$$

Um intervalo com 100(1–α)% de confiança para *p* obtém-se a partir de:

$$\text{IC}_{100(1-\alpha)\%}(p) = \left(\hat{p} \mp z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Pretende-se um intervalo com 90% de confiança, sendo que de acordo com a amostra observada se tem *n* = 40 e $\hat{p}$ = 0.85:

$1-\alpha = 0.9 \Leftrightarrow \alpha = 0.1 \Leftrightarrow \alpha/2 = 0.05 \Leftrightarrow 1-\alpha/2 = 0.95 \Rightarrow z_{1-\alpha/2} = z_{0.95} = 1.645$

$z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = 1.645 \times \sqrt{0.85 \times 0.15/40} \approx 0.093$

Então,

$$\text{IC}_{90\%}(p) = (\,0.85 - 0.093,\ 0.85 + 0.093) = (\,\underline{0.757},\ \underline{0.943})$$

36. Seja X – v.a. (população) que rep. a produção anual de leite (em litros) de uma vaca da espécie em questão, escolhida ao acaso; sabe-se que X $\cap$ $\mathcal{N}(\mu,\ \sigma^2 = 123.21\text{litros}^2)$, sendo μ desconhecido.
Observada um a.a. de 100 vacas daquela espécie (as que estavam na vacaria) obteve-se uma média amostral $\overline{\text{x}}$ = 68. Pretende averiguar-se se a amostra observada fornece evidência estatística para afirmar que a produção anual média (ou esperada) de uma vaca daquela espécie é diferente de 66.6 litros ou, equivalentemente, se $\mu \neq 66.6$. Repare-se que averiguar se é <u>diferente</u> é equivalente a averiguar se é <u>menor ou maior</u>. **Porque se pretende averiguar <u>se existe diferença</u>**, vamos construir um **intervalo de confiança para μ**, o valor esperado da produção anual.
Sendo que:

- a distribuição da população em estudo é <u>Normal</u> e
- a variância (equivalentemente, o desvio-padrão) populacional, <u>σ^2, é conhecida</u>,

o intervalo com 100(1–α)% de confiança para o valor esperado da população, com base numa amostra observada de dimensão *n*, é o seguinte:

$$\text{IC}_{100(1-\alpha)\%}(\mu) = (\ \overline{\text{x}} \mp z_{1-\alpha/2}\ \sigma/\sqrt{n}\ ),$$

onde $z_{1-\alpha/2}$ representa o quantil de probabilidade 1–α/2 da distribuição Normal padrão. No caso presente, pretende-se que α seja igual a 5%, tem-se *n* = 100 e $\overline{\text{x}}$ = 68:

$\alpha = 0.05 \Leftrightarrow \alpha/2 = 0.025 \Leftrightarrow 1-\alpha/2 = 0.975 \Rightarrow z_{1-\alpha/2} = z_{0.975} = 1.96$

$z_{1-\alpha/2}\ \sigma/\sqrt{n} = 1.96 \times \sqrt{123.21}\ /10 = 2.1645$

Então,

$$\text{IC}_{95\%}(\mu) = (\ 68 - 2.1645,\ 68 + 2.1645) = (\ 65.8355,\ 70.1645),$$

ou seja, pode afirmar-se <u>com 95% de confiança</u> que $\mu \in$ (65.8355, 70.1645) e, como 66.6 pertence ao intervalo, <u>não há razões para excluir a possibilidade de que μ seja exactamente igual a 66.6</u>.
<u>Conclusão</u>: Com base nos dados, e considerando α = 5%, não se pode dizer que a produção anual média de leite das vacas da espécie em questão difere significativamente de 66.6 litros.