42. Seja X – v.a. (população) que representa o peso (em Kg) de um artigo escolhido ao acaso.
Sabe-se que X $\cap$ N(μ, σ). Observada uma amostra aleatória (a.a.) com n = 25 artigos, (x_1, ..., x_{25}), obteve-se $\sum x_i$ = 70 e $\sum x_i^2$ = 210.

a) Pretende-se um intervalo para μ com 95% de confiança – $IC_{95\%}(\mu)$
Sendo que:
- a distribuição da população em estudo é Normal e
- o desvio-padrão (equivalentemente, a variância) populacional, σ, é conhecido,

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:

$$IC_{100(1-\alpha)\%}(\mu) = (\ \bar{x} \mp z_{1-\alpha/2}\ \sigma/\sqrt{n}\ ),$$

onde $z_{1-\alpha/2}$ representa o quantil de probabilidade $1-\alpha/2$ da distribuição Normal padrão.
No caso presente, quer-se $100(1-\alpha)\%$ igual a 95%, com n = 25 e σ = 1.2:
$100(1-\alpha) = 95 \Leftrightarrow 1-\alpha = 0.95 \Leftrightarrow \alpha = 0.05 \Leftrightarrow \alpha/2 = 0.025 \Leftrightarrow 1-\alpha/2 = 0.975 \Rightarrow z_{1-\alpha/2} = z_{0.975} = 1.96$;
$\bar{x} = 70/25 = 2.8$;
$z_{1-\alpha/2}\ \sigma/\sqrt{n} = 1.96\times1.2/5 = 0.4704$.
Então,

$$IC_{95\%}(\mu) = (\ 2.8 - 0.4704,\ 2.8 + 0.4704) = (\ \underline{2.3296},\ \underline{3.2704})$$

b) Pretende-se um intervalo para μ com 95% de confiança – $IC_{95\%}(\mu)$
Sendo que:
- a distribuição da população em estudo é Normal e
- o desvio-padrão (equivalentemente, a variância) populacional, σ, é desconhecido,

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:

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

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, já foi visto na alínea anterior que $1-\alpha/2 = 0.975$:
$t_{0.975;24} = 2.0639$ (valor obtido na tabela de quantis da distribuição t de Student, nas calculadoras que têm essa funcionalidade ou numa folha de cálculo computacional);
$s = \sqrt{\frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)} = \sqrt{(210-25\times2.8^2)/24} = \sqrt{14/24}$;
$t_{1-\alpha/2;n-1}\ s/\sqrt{n} = 2.0639\times\sqrt{14/24}\ /5 \approx 0.3153$.
Então,

$$IC_{99\%}(\mu) = (\ 2.8 - 0.3153,\ 2.8 + 0.3153) = (\ \underline{2.4847},\ \underline{3.1153})$$

43. Seja X – v.a. (população) que representa o peso (em Kg) de um melão da variedade em questão, escolhido ao acaso. Nada se sabe sobre a distribuição de probabilidade de X. Observada uma a.a. com n = 230 melões, (x_1, ..., x_{230}), obteve-se $\bar{x}$ = 2.35Kg e s = 0.45Kg. Pretende-se um intervalo com 99% de confiança para o verdadeiro peso médio (ou esperado), μ, de um melão escolhido ao acaso – $IC_{99\%}(\mu)$.
Sendo que:
- a distribuição da população em estudo é desconhecida e
- a dimensão da amostra observada é "grande"

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:

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

onde $z_{1-\alpha/2}$ representa o quantil de probabilidade $1-\alpha/2$ da distribuição Normal padrão.
No caso presente, quer-se $100(1-\alpha)\%$ igual a 99%, com $n = 230$ e $\sigma = 1.2$:
$100(1-\alpha) = 99 \Leftrightarrow 1-\alpha = 0.99 \Leftrightarrow \alpha = 0.01 \Leftrightarrow \alpha/2 = 0.005 \Leftrightarrow 1-\alpha/2 = 0.995 \Rightarrow z_{1-\alpha/2} = z_{0.995} = 2.576$;
$z_{1-\alpha/2}\, s/\sqrt{n} = 2.576\times0.45/\sqrt{230} \approx 0.0764$.
Então,

$$IC_{99\%}(\mu) = (\ 2.35 - 0.0764,\ 2.35 + 0.0764) = (\ \underline{2.2736},\ \underline{3.4264})$$

44. ---------

45. Seja p a proporção de estudantes, na região em causa, que pretendem ingressar no ensino superior. De modo equivalente, pode definir-se p como sendo a probabilidade de um estudante, escolhido ao acaso na região em causa, pretender ingressar no ensino superior – "sucesso". O que se pretende é estudar a probabilidade de "sucesso" por entre os estudantes da região.
Supondo que se pretende observar uma a.a. que inclua n estudantes da região em causa, então o n.º de sucessos na amostra é uma v.a. que se pode definir como

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

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

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

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

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

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

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

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

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

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 para intervalos):

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

Concretizando para a amostra recolhida, onde se tem $n = 40$, $x = 30$ (n.º de "S" na amostra observada) e se pretende uma confiança de 90%, tem-se:

$\hat{p} = 30/40 = 0.75$

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

$z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} = 1.645\times\sqrt{0.75\times 0.25/40} \approx 0.1126$

Então,

$$\mathrm{IC}_{90\%}(p) = (\ 0.75 - 0.1126,\ 0.75 + 0.1126) = (\ \underline{0.6374},\ \underline{0.8626})$$