RESOLUÇÃO ASSISTIDA DE PROBLEMAS
CAPÍTULO 10 – ESTIMAÇÃO POR INTERVALO
PROBLEMA 10.1

APOIOS À RESOLUÇÃO

Apoio 1 (apenas o resultado)

[0.817, 0.883]. ■

Apoio 2 (sugestão)

Note-se que a amostra é suficientemente grande para que, qualquer que seja a distribuição original do tempo de reacção (variável *X*), a média amostral $\overline{X}$ siga uma distribuição aproximadamente Normal. Na secção 10.2 veja qual a expressão que define o intervalo de confiança para o valor esperado, numa situação em que se dispõe de uma amostra de grande dimensão. ■

Apoio 3 (resolução completa)

Sejam,

X: tempo de reacção dos condutores de camião, após o almoço

$N = 140$

$\bar{x} = 0.85$ [segundos]

$s = 0.20$ [segundos].

A amostra é suficientemente grande para que, de acordo com o teorema do limite central, independentemente da distribuição da variável original (*X*), $\overline{X}$ siga aproximadamente uma distribuição Normal.

Assim, $\overline{X} \sim N\left(\mu, \sigma^2 / N\right)$, onde μ e σ representam o valor esperado e a variância da população e *N* a dimensão da amostra. Padronizando, vem:

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

Uma vez que se admitiu que a amostra é grande, resulta

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

em que *S* representa o estimador desvio padrão amostral.

O intervalo de confiança para o valor esperado μ_X a $(1-\alpha)\cdot 100\%$ vem dado por:

$$\left[\overline{X} - z(\alpha/2)\cdot\frac{S}{\sqrt{N}},\ \overline{X} + z(\alpha/2)\cdot\frac{S}{\sqrt{N}}\right].$$

Como $z(\alpha/2 = 0.025) = 1.96$ (ver Tabela 3), resulta finalmente

$$\left[0.85 - 1.96\cdot\frac{0.20}{\sqrt{140}},\ 0.85 + 1.96\cdot\frac{0.20}{\sqrt{140}}\right] \equiv [0.817,\ 0.883].\ \blacksquare$$

PROBLEMA 10.2

APOIOS À RESOLUÇÃO

Apoio 1 (apenas o resultado)

$[4472, +\infty[$. ■

Apoio 2 (sugestão)

Uma vez que a amostra é de pequena dimensão, considera-se válida a hipótese de a tensão de rotura dos cabos do lote (variável *X*) seguir, aproximadamente, uma distribuição Normal. Nesta situação, a média amostral padronizada

$$\frac{\overline{X} - \mu}{S}$$

seguirá uma distribuição *t* de *Student*. Na secção 10.2 veja qual a expressão que define o intervalo de confiança para o valor esperado quando a amostra é de pequena dimensão. ■

Apoio 3 (resolução completa)

Sejam,

X: tensão de rotura de um cabo SCB-33R (pertencente ao lote)

$N = 10$

$\overline{x} = 4537$ [kg / cm^2]

$s = 112$ [kg / cm^2].

Admite-se que a variável X segue uma distribuição $N(\mu, \sigma^2)$. Nestas condições,

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

Uma vez que, por um lado, o desvio padrão é desconhecido e que, por outro, a amostra é de pequena dimensão, não se considera válida a aproximação $S \approx \sigma$. Assim, resulta que a variável

$\dfrac{\overline{X} - \mu}{S / \sqrt{N}}$ (em que S representa o estimador desvio padrão amostral)

segue uma distribuição $t_{N\text{-}1}$.

O intervalo de confiança, aberto à direita, para o valor esperado, μ_X, a $(1-\alpha)\cdot 100\%$ é, assim, dado por

$$\left[\overline{X} - t_{N-1}(\alpha)\cdot \frac{S}{\sqrt{N}},\ +\infty\right[.$$

Como t_9 (0.05) = 1.833 (ver Tabela 5), resulta

$$\left[4537 - 1.833\cdot \frac{112}{\sqrt{10}}, +\infty\right[\equiv \left[4472, +\infty\right[.$$

Note-se que, para o nível de confiança de 95%, o limite esquerdo deste intervalo corresponde ao menor valor possível para o valor esperado da tensão de rotura. Ou seja, este intervalo de confiança pode igualmente escrever-se da seguinte forma:

$\mu_X \geq 4472$ kg/cm^2, com 95% de confiança. ■

PROBLEMA 10.3

APOIOS À RESOLUÇÃO

Apoio 1 (apenas o resultado)

Intervalo de confiança a 90%: [0.059, 0.156]

Intervalo de confiança a 95%: [0.055, 0.174]

Intervalo de confiança a 99%: [0.047, 0.218]. ■

Apoio 2 (sugestão)

Na secção 10.2 veja qual a expressão que define o intervalo de confiança para a variância de uma população Normal. Note que o intervalo de confiança não é centrado em s^2, pelo facto de a distribuição do χ^2 (utilizada na definição daquele intervalo) não ser simétrica. ■

Apoio 3 (resolução completa)

Sejam

X: tempo necessário para a realização da operação de montagem

$N = 25$

$s^2 = 0.3^2 = 0.09$ [horas^2].

Admite-se que X segue aproximadamente uma distribuição $N(\mu, \sigma^2)$. Nestas condições, a variável

$(N-1)\cdot\dfrac{S^2}{\sigma^2}$ segue uma distribuição χ^2_{N-1}.

O intervalo de confiança para a variância de uma população Normal, σ_X^2, a $(1-\alpha)\cdot 100\%$ é dado por

$$\left[\frac{(N-1)\cdot S^2}{\chi^2_{N-1}(\alpha/2)}, \frac{(N-1)\cdot S^2}{\chi^2_{N-1}(1-\alpha/2)}\right].$$

Na tabela seguinte representam-se os valores de $\chi^2_{N-1}(\alpha/2)$ e de $\chi^2_{N-1}(1-\alpha/2)$, obtidos a partir da Tabela 4 para intervalos de confiança de 90, 95 e 99%.

1 - α	$\chi^2_{24}(1-\alpha/2)$	$\chi^2_{24}(\alpha/2)$
90%	13.85	36.4
95%	12.40	39.4
99%	9.89	45.6

Os intervalos pedidos são então:

$\left[\dfrac{24\cdot 0.09}{36.4}, \dfrac{24\cdot 0.09}{13.85}\right] \equiv [0.059,\ 0.156]$, a 90% de confiança;

$\left[\dfrac{24\cdot 0.09}{39.4}, \dfrac{24\cdot 0.09}{12.40}\right] \equiv [0.055,\ 0.174]$, a 95% de confiança;

$\left[\dfrac{24\cdot 0.09}{45.6}, \dfrac{24\cdot 0.09}{9.89}\right] \equiv [0.047,\ 0.218]$, a 99% de confiança.

Note-se que os intervalos não são centrados em $s^2 = \hat{\sigma}^2 = 0.09$ pelo facto de a distribuição do Qui-Quadrado não ser simétrica. ■

<u>PROBLEMA 10.4</u>

APOIOS À RESOLUÇÃO

<u>Apoio 1</u> (apenas o resultado)

[0.203, 0.297]. ■

Apoio 2 (sugestão)

É perfeitamente razoável admitir que a variável *Y*, que representa o número de operários que conhece suficientemente bem as normas de segurança, segue uma distribuição Hipergeométrica. Na secção 6.4.3 reveja a relação entre as distribuições Binomial e Hipergeométrica. Analise ainda em que condições a distribuição Binomial poder ser aproximada por uma Normal. Na secção 10.2 veja qual a expressão que define o intervalo de confiança da proporção Binomial (amostra de grande dimensão). ■

Apoio 3 (resolução completa)

Sejam,

Y: número de operários que, de entre 300 seleccionados aleatoriamente uma semana após a emissão das normas, as conhece suficientemente bem

p: proporção de operários que, de entre 4000 da empresa industrial, conhece suficientemente bem as normas de segurança

M: número de operários da empresa industrial = 4000

$N = 300$.

É razoável admitir que *Y* segue uma distribuição $H\left(M.p\,, M.(1-p), N\right)$.
Substituindo os valores de *M* e de *N*, vem

$Y \frown H\left(4000 \cdot p\,, 4000 \cdot (1 - p), 300\right)$, com

$$\mu_Y = N \cdot p = 300 \cdot p$$

$$\sigma_Y^2 = N \cdot p \cdot (1-p) \cdot \frac{M-N}{M-1} = 300 \cdot p \cdot (1-p) \cdot \frac{4000-300}{4000-1} = 277.6 \cdot p \cdot (1-p).$$

Uma vez que $M \geq 10 \cdot N$, esta distribuição Hipergeométrica pode ser aproximada por uma distribuição Binomial. Por sua vez, dado que $N \geq 20$ e partindo do pressuposto que p é suficientemente elevado para que $N \cdot p > 7$, a distribuição Binomial pode ser aproximada por uma distribuição Normal.

Assim, *Y* segue aproximadamente a distribuição Normal
$N\left[\mu = 300 \cdot p, \sigma^2 = 277.6 \cdot p \cdot (1-p)\right]$

Tome-se agora a variável proporção Binomial, $\frac{Y}{N}$. Denotando $\hat{P}$ como o estimador de p resulta

$$\hat{P} = \frac{Y}{N} = \frac{Y}{300}.$$

Nestas condições, $\hat{P} = \dfrac{Y}{300}$ segue aproximadamente a distribuição Normal

$$N\left[\mu = p,\ \sigma^2 = \frac{277.6}{300^2} \cdot p \cdot (1-p)\right].$$

Padronizando a variável Normal $\hat{P} = Y/300$, vem

$$Z = \frac{\frac{Y}{300} - p}{\sqrt{\frac{277.6}{300^2} \cdot p \cdot (1-p)}} \sim N(0,1).$$

A expressão do intervalo de confiança para a proporção binomial, p, a (1-α)·100% é então

$$\left[\frac{Y}{300} - z(\alpha/2) \cdot \sqrt{\frac{277.6}{300^2} \cdot p \cdot (1-p)}, \quad \frac{Y}{300} + z(\alpha/2) \cdot \sqrt{\frac{277.6}{300^2} \cdot p \cdot (1-p)} \right].$$

Uma vez que a proporção p é desconhecida, o seu valor nesta expressão terá de ser substituído pelo da sua estimativa,

$$\hat{p} = \frac{y}{N} = \frac{75}{300} = 0.25.$$

Assim, a estimativa do desvio padrão de p resulta

$$\hat{\sigma}_p = \sqrt{\frac{277.6}{300^2} \cdot \hat{p} \cdot (1-\hat{p})} = \sqrt{\frac{277.6}{300^2} \cdot 0.25 \cdot 0.75} = 0.0240.$$

Finalmente, como $z(\alpha/2 = 0.025) = 1.96$, o intervalo de confiança para a proporção binomial, p, a 95% é

$[\,0.25 - 1.96 \cdot 0.0240\,,\, 0.25 + 1.96 \cdot 0.0240\,] \equiv$ [0.203, 0.297] ■

PROBLEMA 10.5

APOIOS À RESOLUÇÃO

Apoio 1 (apenas o resultado)

Admitindo que as variâncias são iguais:

Intervalo de confiança a 90%: [4.64, 8.76]
Intervalo de confiança a 95%: [4.22, 9.18].

Admitindo que as variâncias são diferentes:

Intervalo de confiança a 90%: [4.51, 8.89]
Intervalo de confiança a 95%: [4.05, 9.35]. ■

Apoio 2 (sugestão)

Perante a (pequena) dimensão das amostras, admita que o comportamento das variáveis "saldos das contas à ordem", quer na agência A quer na B, não se afastam substancialmente de distribuições Normais. Procure caracterizar a variável $\overline{X}_A - \overline{X}_B$. Considere os casos das variâncias das duas populações serem e não serem iguais. Recorrendo a um intervalo de confiança para a razão das variâncias de populações Normais, verifique se a primeira hipótese é plausível. Na secção 10.2 veja quais as expressões que definem os intervalos de confiança para a diferença entre os valores esperados de duas populações. Compare e analise os resultados que obtiver. ■

Apoio 3 (resolução completa)

Sejam,

X_A : saldo de uma conta à ordem da agência A [em euros]

X_B : saldo de uma conta à ordem da agência B [em euros]

$N_A = 17$, $\bar{x}_A = 48.2$ euros, $s_A = 2.74$ euros

$N_B = 13$, $\bar{x}_B = 41.5$ euros, $s_B = 3.91$ euros.

Admite-se que X_A e X_B seguem distribuições Normais independentes, com variâncias não necessariamente iguais. Nestas condições, a diferença $\overline{X}_A - \overline{X}_B$ é ainda uma variável aleatória Normal e assim:

$$\overline{X}_A - \overline{X}_B \sim N\left(\mu_A - \mu_B,\ \frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}\right),$$

ou ainda

$$Z = \frac{\left(\overline{X}_A - \overline{X}_B\right) - \left(\mu_A - \mu_B\right)}{\sqrt{\dfrac{\sigma_A^2}{N_A} + \dfrac{\sigma_B^2}{N_B}}} \sim N(0,\ 1).$$

Uma vez que, por um lado, os desvios padrões (σ_A e σ_B) são desconhecidos e, por outro, não são válidas as aproximações $S_A \approx \sigma_A$ e $S_B \approx \sigma_B$ (dado que ambas as amostras são de pequena dimensão), resulta que

$$\frac{\left(\overline{X}_A - \overline{X}_B\right) - \left(\mu_A - \mu_B\right)}{\sqrt{\dfrac{S_A^2}{N_A} + \dfrac{S_B^2}{N_B}}} \sim t_{GL}.$$

Admita-se, em primeiro lugar, que as variâncias são iguais (ou seja que $\sigma_A^2 = \sigma_B^2 = \sigma^2$). Note-se que esta conjectura pode ser verificada construindo um intervalo de confiança (a 95%, por exemplo) para a razão das variâncias σ_A^2 / σ_B^2. Tal intervalo é dado por

$$\left[\frac{1}{F_{N_A-1,\,N_B-1}(0.025)} \cdot \frac{S_A^2}{S_B^2}\ ,\ \frac{1}{F_{N_A-1,\,N_B-1}(0.975)} \cdot \frac{S_A^2}{S_B^2}\right].$$

Ora,

$F_{16,12}(0.025) = 3.15$

e

$$F_{16,12}(0.975)=\frac{1}{F_{12,16}(0.025)}=\frac{1}{2.89}.$$

Substituindo, o intervalo de confiança para σ_A^2/σ_B^2 a 95% é então:

$$\left[\frac{1}{3.15}\cdot\frac{2.74^2}{3.91^2}\ ,\ 2.89\cdot\frac{2.74^2}{3.91^2}\right]\equiv\left[0.1559\ ,1.4192\right].$$

Como o valor $\sigma_A^2/\sigma_B^2=1$ está contido no intervalo, é plausível admitir que as variâncias são iguais (embora, com base apenas neste resultado, isso não se possa afirmar com segurança).

Admitindo que as variâncias são iguais, então a estimativa da variância comum, σ^2, é dada pela expressão:

$$\hat{\sigma}^2=s^2=\frac{(N_A-1)\cdot s_A^2+(N_B-1)\cdot s_B^2}{N_A+N_B-2}=\frac{16\cdot 2.74^2+12\cdot 3.91^2}{17+13-2}=3.29^2$$

Neste caso, como a diferença $\overline{X}_A-\overline{X}_B$ segue uma distribuição Normal, a variável

$$\frac{(\overline{X}_A-\overline{X}_B)-(\mu_A-\mu_B)}{S\cdot\sqrt{\dfrac{1}{N_A}+\dfrac{1}{N_B}}}$$

segue uma distribuição *t* de *student* com *GL* (que denota o número de graus de liberdade) igual a 28 (o mesmo número de graus de liberdade com que foi estimado *S*).

Nestas condições, os intervalos de confiança para $\mu_A-\mu_B$ vêm:

a 90%:

$$\left[(\overline{X}_A-\overline{X}_B)-t_{28}(0.05)\cdot S\sqrt{\frac{1}{N_A}+\frac{1}{N_B}},\ (\overline{X}_A-\overline{X}_B)+t_{28}(0.05)\cdot S\sqrt{\frac{1}{N_A}+\frac{1}{N_B}}\right]$$

$$\equiv\left[6.7-1.701\cdot 1.212,\ 6.7+1.701\cdot 1.212\right]\equiv\left[4.64,\ 8.76\right];$$

a 95%:

$$\left[(\overline{X}_A-\overline{X}_B)-t_{28}(0.025)\cdot S\sqrt{\frac{1}{N_A}+\frac{1}{N_B}},\ (\overline{X}_A-\overline{X}_B)+t_{28}(0.025)\cdot S\sqrt{\frac{1}{N_A}+\frac{1}{N_B}}\right]$$

$$\equiv\left[6.7-2.048\cdot 1.212,\ 6.7+2.048\cdot 1.212\right]\equiv\left[4.22,\ 9.18\right].$$

Admitindo agora que as variâncias são diferentes, o valor de *GL* é calculado de acordo com a expressão:

$$GL = \frac{\left(\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}\right)^2}{\frac{\left(S_A^2 / N_A\right)^2}{N_A - 1} + \frac{\left(S_B^2 / N_B\right)^2}{N_B - 1}} = \frac{2.617}{0.0122 + 0.1152} = 20.5 \approx 20$$

(note-se que é recomendado a adopção do inteiro imediatamente inferior, já que este procedimento conduz à definição de um intervalo com uma confiança maior do que a especificada inicialmente).

Intervalos de confiança para $\mu_A - \mu_B$:

a 90%:

$$\left[\left(\overline{X}_A - \overline{X}_B\right) - t_{20}(0.05) \cdot \sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}},\ \left(\overline{X}_A - \overline{X}_B\right) + t_{20}(0.05) \cdot \sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}}\right]$$

$$\equiv [6.7 - 1.725 \cdot 1.272,\ 6.7 + 1.725 \cdot 1.272] \equiv [4.51,\ 8.89];$$

a 95%:

$$\left[\left(\overline{X}_A - \overline{X}_B\right) - t_{20}(0.025) \cdot \sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}},\ \left(\overline{X}_A - \overline{X}_B\right) + t_{20}(0.025) \cdot \sqrt{\frac{S_A^2}{N_A} + \frac{S_B^2}{N_B}}\right]$$

$$\equiv [6.7 - 2.086 \cdot 1.272,\ 6.7 + 2.086 \cdot 1.272] \equiv [4.05,\ 9.35].$$

Note-se que, quer para 90% quer para 95% de confiança, estes intervalos têm uma amplitude superior à dos que foram calculados no pressuposto de que as variâncias eram iguais. Este facto resulta de, agora, se recorrer aos valores das caudas da distribuição *t* de *student* com 20 graus de liberdade, em vez da distribuição com *GL* = 28 que se tinha utilizado no primeiro caso. ■

PROBLEMA 10.6

APOIOS À RESOLUÇÃO

Apoio 1 (apenas o resultado)

[0.006, 0.194]. ■

Apoio 2 (sugestão)

É de admitir que o número de quadros superiores e o número de gestores que leram a última edição do semanário seguem distribuições Binomiais. Para definir o intervalo de confiança pretendido procure caracterizar o estimador $\hat{P}_G - \hat{P}_Q$ de $p_G - p_Q$. Como as amostras possuem grandes dimensões, pode admitir que este estimador segue uma distribuição Normal. ■

Apoio 3 (resolução completa)

Sejam,

Y_Q: número de quadros superiores que, de entre os 350 quadros entrevistados, leram a última edição do semanário

Y_G: número de gestores que, de entre os 325 gestores entrevistados, leram a última edição do semanário

P_Q: proporção de quadros que leram a última edição do semanário

P_G: proporção de gestores que leram a última edição do semanário

$Y_Q \frown B(N_Q = 350, p_Q)$

$Y_G \frown B(N_G = 325, p_G)$.

Dado que N_Q e $N_G \geq 20$ e partindo do pressuposto que p_Q e p_G são suficientemente elevados para que $N_Q \cdot p_Q > 7$ e $N_G \cdot p_G > 7$, ambas as distribuições Binomiais podem ser aproximadas por distribuições Normais. Os parâmetros destas distribuições são:

$\mathrm{E}(Y_Q) = N_Q \cdot p_Q$

$\mathrm{Var}(Y_Q) = N_Q \cdot p_Q \cdot (1 - p_Q)$

$\mathrm{E}(Y_G) = N_G \cdot p_G$

$\mathrm{Var}(Y_G) = N_G \cdot p_G \cdot (1 - p_G)$.

Assim, Y_Q e Y_G seguem aproximadamente distribuições Normais, respectivamente

$N\left[\mu = 350 \cdot p_Q, \sigma^2 = 350 \cdot p_Q \cdot (1 - p_Q)\right]$ e $N\left[\mu = 325 \cdot p_G, \sigma^2 = 325 \cdot p_G \cdot (1 - p_G)\right]$.

Por sua vez, $\hat{P}_Q = \frac{Y_Q}{N_Q} = \frac{Y_Q}{350}$ e $\hat{P}_G = \frac{Y_G}{N_G} = \frac{Y_G}{325}$ são estimadores, respectivamente, de P_Q e de P_G. Estes estimadores seguem (aproximadamente) as distribuições Normais

$$N\left[\mu = p_Q, \quad \sigma^2 = \frac{p_Q \cdot (1 - p_Q)}{350}\right] \text{ e } N\left[\mu = p_G, \quad \sigma^2 = \frac{p_G \cdot (1 - p_G)}{325}\right].$$

Admitindo que as amostras são independentes, a variável aleatória $\hat{P}_G - \hat{P}_Q$ segue também (aproximadamente) a distribuição Normal

$$N\left[\mu = p_G - p_Q, \quad \sigma^2 = \frac{p_G \cdot (1 - p_G)}{325} + \frac{p_Q \cdot (1 - p_Q)}{350}\right]$$

Padronizando resulta:

$$Z = \frac{\hat{P}_G - \hat{P}_Q - \left(p_G - p_Q\right)}{\sqrt{\frac{p_G \cdot (1 - p_G)}{325} + \frac{p_Q \cdot \left(1 - p_Q\right)}{350}}} \sim N(0,1).$$

O intervalo de confiança para a diferença de proporções binomiais, P_Q - P_G, a 99% é então:

$$\frac{Y_G}{N_G} - \frac{Y_Q}{N_Q} \pm z(0.005) \cdot \sqrt{\frac{\hat{p}_G \cdot (1 - \hat{p}_G)}{325} + \frac{\hat{p}_Q \cdot \left(1 - \hat{p}_Q\right)}{350}}.$$

Ora,

$$\frac{Y_G}{N_G} = \hat{p}_G = \frac{130}{325} = 0.4 \text{ e } \frac{Y_Q}{N_Q} = \hat{p}_Q = \frac{105}{350} = 0.3.$$

Substituindo estes valores na expressão do intervalo de confiança e atendendo a que $Z(0.005) = 2.575$, obtém-se:

$$0.1 \pm 2.575 \cdot 0.0366 \equiv 0.1 \pm 0.094 \equiv \left[0.006\ ,\ 0.194\right].$$ ■

PROBLEMA 10.7

APOIOS À RESOLUÇÃO

Apoio 1 (apenas o resultado)

No mínimo 37 ensaios. ■

Apoio 2 (sugestão)

Especifique um intervalo de confiança a 99% para o valor esperado. Note que ele é centrado na média amostral. A partir do intervalo procure expressar a sua amplitude em função da dimensão da amostra. ■

Apoio 3 (resolução completa)

Sejam

X: tensão de rotura do material utilizado nos tubos de ensaio

N: dimensão da amostra a partir da qual será estimado μ_X

$\overline{X}$: média amostral das observações de X

σ_X =70 [psi].

Admite-se que a dimensão da amostra (que se irá determinar) é suficientemente grande para que, de acordo com o teorema do limite central e independentemente da distribuição da variável original X, a média amostral, $\overline{X}$, siga aproximadamente uma distribuição Normal.

Nestas condições, e uma vez que a variância é conhecida, verifica-se que

$$Z = \frac{\overline{X} - \mu}{\sigma_X / \sqrt{N}} \sim N(0,1).$$

O intervalo de confiança para o valor esperado μ_X a $(1-\alpha)\cdot 100\%$ é dado por:

$$\left[\overline{X} - z(\alpha/2)\cdot \frac{\sigma_X}{\sqrt{N}},\ \overline{X} + z(\alpha/2)\cdot \frac{\sigma_X}{\sqrt{N}} \right].$$

A amplitude deste intervalo é $2 \cdot z(\alpha/2)\cdot \frac{\sigma_X}{\sqrt{N}}$.

Ora, como $Z(\alpha = 0.005) = 2.575$ (ver Tabela 3) resulta:

$$2 \cdot 2.575 \cdot \frac{70}{\sqrt{N}} \le 60 \ \Rightarrow\ N \ge \left(\frac{2 \cdot 2.575 \cdot 70}{60} \right)^2 = 36.1$$

Ou seja

$N \ge 37$.

De facto, a dimensão mínima da amostra, N =3 7, é suficientemente grande para se poder admitir a Normalidade de $\overline{X}$ (embora de forma aproximada). ■

<u>**PROBLEMA 10.8**</u>

APOIOS À RESOLUÇÃO

<u>Apoio 1</u> (apenas o resultado)

No mínimo 349 alunos. ■

<u>Apoio 2</u> (sugestão)

Note que o número Y de alunos que, de entre os N inquiridos, favorecem o horário de sábado de manhã, segue seguramente uma distribuição Hipergeométrica. Na secção 6.4.3 reveja a relação entre as distribuições Binomial e Hipergeométrica. Na secção 10.2, veja qual a expressão que define o intervalo de confiança para a proporção Binomial (amostra de grande dimensão). Note que o intervalo de confiança da proporção Binomial é centrado. A partir do intervalo, procure expressar a sua amplitude em função da dimensão da amostra. ■

Apoio 3 (resolução completa)

Sejam,

Y: número de alunos que, de entre os N inquiridos, favorecem o horário de sábado de manhã

p: proporção de alunos que, de entre os 3800 da Faculdade, favorecem o horário de sábado de manhã

M: número de alunos da Faculdade = 3800

N: dimensão da amostra.

Admite-se que a variável Y segue uma distribuição Hipergeométrica

$H(M.p\,, M.(1-p), N)$. Substituindo vem,

$Y \sim H(3800 \cdot p\,, 3800 \cdot (1 - p), N)$ com média e variância iguais a:

$$\mu_Y = N \cdot p$$

e

$$\sigma_Y^2 = N \cdot p \cdot (1-p) \cdot \frac{M-N}{M-1} = N \cdot p \cdot (1-p) \cdot \frac{3800-N}{3799}.$$

Admite-se que a dimensão da amostra, N, é pequena quando comparada com o número total de alunos da Faculdade, M, mas é suficientemente grande para que a distribuição de Y possa ser aproximada por uma distribuição Normal.

Nestas condições, é possível aproximar a distribuição de Y pela Normal

$$N\left[\mu = N \cdot p,\ \sigma^2 = N \cdot p \cdot (1-p) \cdot \frac{3800-N}{3799}\right].$$

Recorrendo agora à proporção amostral, $\hat{P} = \dfrac{Y}{N}$ (onde $\hat{P}$ representa o estimador de p), resulta que esta variável seguirá (ainda que aproximadamente) a distribuição Normal

$$N\left[\mu = p,\ \sigma^2 = \frac{p \cdot (1-p)}{N} \cdot \frac{3800-N}{3799}\right].$$

Padronizando, vem

$$Z = \frac{\dfrac{Y}{N} - p}{\sqrt{\dfrac{p \cdot (1-p)}{N} \cdot \dfrac{3800-N}{3799}}} \sim N(0,1).$$

O intervalo de confiança para a proporção binomial, p, a (1-α)·100% é então:

$$\left[\frac{Y}{N} - Z(\alpha/2) \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N} \cdot \frac{3800-N}{3799}},\ \frac{Y}{N} + Z(\alpha/2) \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N} \cdot \frac{3800-N}{3799}}\right].$$

A amplitude deste intervalo é

$$2 \cdot Z(\alpha / 2) \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N} \cdot \frac{3800-N}{3799}}\,.$$

Assim, para $\alpha = 0.05$ vem $Z(0.025) = 1.96$, donde

$$2 \cdot 1.96 \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{N} \cdot \frac{3800-N}{3799}} \leq 0.1\,.$$

Para resolver esta inequação e determinar N, é necessário dispor de uma estimativa de p, $\hat{p}$. Como, neste caso, isso é impossível, calcula-se N para o valor de p que maximiza a amplitude do intervalo (ou seja, o valor que maximiza $\hat{p} \cdot (1-\hat{p})$). Assim,

$$\frac{d[\hat{p} \cdot (1-\hat{p})]}{d\hat{p}} = 1 - 2\hat{p} = 0 \quad \Rightarrow \quad \hat{p} = 0.5\,.$$

Substituindo,

$$2 \cdot 1.96 \cdot \sqrt{\frac{0.5 \cdot (1-0.5)}{N} \cdot \frac{3800-N}{3799}} \leq 0.1 \;\Rightarrow\; 0.1 \geq 3.92 \cdot \sqrt{\frac{0.25}{N} \cdot \frac{3800-N}{3799}} \;\Rightarrow$$

$$\frac{3800-N}{3799 \cdot N} \leq \frac{0.1^2}{3.92^2 \cdot 0.25} = 0.002603 \;\Rightarrow\; 3800 - N \leq 9.888 \cdot N \;\Rightarrow$$

$$N \geq \frac{3800}{10.89} = 348.9$$

Finalmente, a dimensão mínima da amostra que satisfaz os critérios enunciados é:

$N_{\text{mín}} = 349$.

Perante este resultado, a hipótese anteriormente formulada (de que N era grande, mas pequeno comparada com M) justifica-se plenamente. ■