FICHA 10 – resolução

DPC

1. Os rendimentos de um processo químico têm uma distribuição normal com $\sigma = 3$. Nos últimos três dias, os ganhos têm sido: $91.6, 88.75, 90.8, 89.95$, e 91.3. Calcule o intervalo de confiança de 95% para a média.

Resolução:
Temos uma distribuição $N(\mu, \sigma^2)$ onde μ é desconhecido e σ é dado. Consideramos um processo de amostragem independente onde cada amostra aleatória X_i, $i \in \{1, 2, \cdots\}$, segue a distribuição dada. A variável aleatória que representa a média amostral obtida a partir de n amostras é:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

É possível demonstrar que a soma de duas variáveis aleatórias independentes e com distribuição normal resulta numa terceira variável aleatória que também segue uma distribuição normal. Assim, como $\bar{X}$ resulta da soma de n v.a. independentes e normalmente distribuídas, concluímos que $\bar{X}$ também segue uma distribuição normal. Resta-nos determinar a respetiva média e variância:

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu,$$

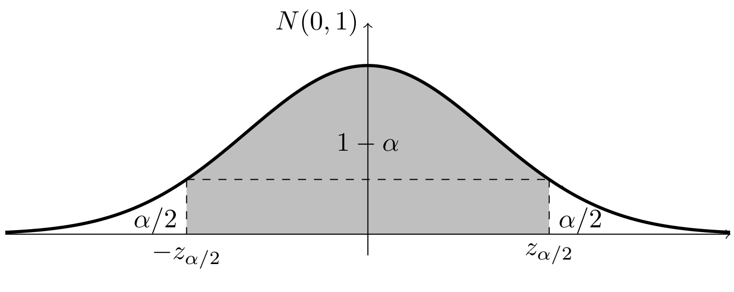
e, como as n amostras são independentes,

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{\sigma^2}{n}.$$

Concluímos assim que $\bar{X} \sim N(\mu, \sigma^2/n)$. Procuramos um intervalo $[a, b]$ centrado em μ e tal que $P(a \leq \bar{X} \leq b) = 1 - \alpha$. Note que, como o intervalo é centrado em μ, temos:

$$1 - \alpha = P(a \leq \bar{X} \leq b) = P\Bigg(-z_{\alpha/2} \leq \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{Z \sim N(0,1)} \leq z_{\alpha/2}\Bigg),$$

Pelo que, dado α, obtemos o valor de $z_{\alpha/2}$ através da tabela da distribuição normal, observando que $P(Z \geq z_{\alpha/2}) = \alpha/2$. O gráfico seguinte poderá ajudá-lo a compreender a situação:



Quando temos um conjunto de observações $\{x_1, x_2, \cdots, x_n\}$, podemos calcular $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$, que será uma observação da v.a. $\bar{X}$. Utilizando a média amostral observada $\bar{x}$ no lugar de $\bar{X}$, podemos estimar um intervalo de valores numéricos para μ (chamado *intervalo de confiança* para a média da distribuição normal):

$$-z_{\alpha/2} \leq \left.\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right|_{\bar{X}=\bar{x}} \leq z_{\alpha/2} \Leftrightarrow \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \mu \in \left[\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

No presente exercício, temos $\sigma = 3$, $n = 5$, $\bar{x} \approx 90.48$ e $1 - \alpha = 95\%$. Assim,

$$1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05 \implies z_{\alpha/2} = z_{0.025} \approx 1.960,$$

$$\mu \in \left[90.48 \pm 1.960 \frac{3}{\sqrt{5}}\right] \approx [87.86,\ 93.10].$$

Nota: Podemos ser tentados a escrever $P(87.86 \leq \mu \leq 93.10) = 0.95$, dizendo assim que a média μ tem uma probabilidade de 95% de pertencer ao intervalo obtido. No entanto, essa afirmação não é rigorosa. Note que (de acordo com a definição frequentista de probabilidade) μ não é uma variável aleatória, é simplesmente um valor desconhecido, mas fixo. Assim, não faz sentido atribuir probabilidades a valores de μ. A variável aleatória aqui presente é $\bar{X}$ (média amostral), que "desapareceu"quando considerámos a respetiva observação $\bar{x} = 90.48$. Como interpretamos então o intervalo obtido?

Note que a média amostral $\bar{X}$ é obtida a partir de n amostras aleatórias. Assim, se, em determinado momento, reunirmos n amostras, iremos obter um certo valor para a média amostral, chamemos-lhe $\bar{x}_1$. Utilizando este $\bar{x}_1$, podemos calcular um intervalo de confiança I_1 para μ (tal como acabámos de fazer). Noutro momento, podemos repetir todo o processo de amostragem, obtendo um novo conjunto n amostras, a partir do qual calculamos uma nova média amostral $\bar{x}_2$ e, finalmente, um novo intervalo de confiança I_2. Em teoria, podemos repetir este procedimento quantas vezes quisermos. Supondo que o fazemos m vezes, obtemos m intervalos $I_1, I_2, \cdots, I_m$. Se m for suficientemente grande ($m \to \infty$), a proporção desses m intervalos que contêm o valor real de μ irá aproximar-se de $1 - \alpha$.

Assim, o intervalo $I = [87.86,\ 93.10]$ que obtivemos é apenas um entre muitos (uma infinidade) que poderíamos ter obtido e, entre todos esses intervalos, sabemos que 95% deles contêm o valor real de μ, que desconhecemos. Evidentemente, desejamos que o intervalo que obtivemos faça parte destes 95% que contêm μ, mas não temos forma de o confirmar a partir dos dados de que dispomos.

2. Um teste de impacto foi realizado em 20 tubos de PVC. A média amostral observada foi de 1.25. Calcule o intervalo de confiança inferior de 99 % para a média. Assuma que a grandeza medida segue uma distribuição normal com $\sigma = 0.25$.

Resolução:
Desta vez, pede-se um "intervalo de confiança inferior", ou seja, um intervalo de confiança que apenas tem limite inferior, sendo aberto à direita. Esse intervalo será tal que:

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\right) = 1 - \alpha.$$

Seguindo um raciocínio idêntico ao do exercício anterior, o intervalo de confiança obtido é:

$$\mu \geq \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}.$$

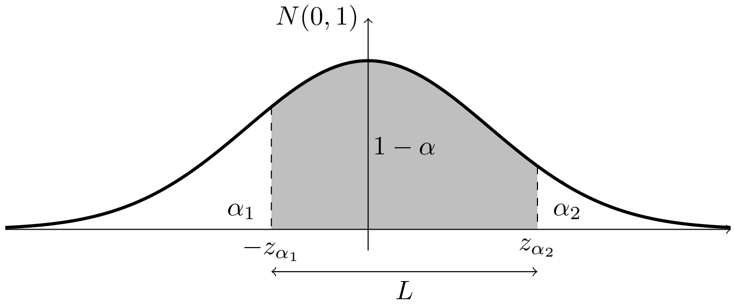
Temos $n = 20$, $\bar{x} = 1.25$, $1 - \alpha = 0.99$ e $\sigma = 0.25$, logo $z_\alpha = z_{0.01} \approx 2.326$ e $\mu \geq 1.12$.

3. A espessura das paredes de 25 garrafas de dois litros foi medida por um engenheiro de controle de qualidade. A média amostral foi 4.05 mm. Calcule o intervalo de confiança inferior de 95% da média μ. Assuma que a espessura das paredes segue uma distribuição normal com $\sigma = 0.08$ mm.

Solução:
$z_{0.05} = 1.645$, $\mu \geq 4.02$mm.

4. Para uma variável Z normal padrão, considere o intervalo de confiança para a média μ com σ conhecido: $\bar{x} - z_{\alpha_1}\sigma/\sqrt{n} \leq \mu \leq \bar{x} + z_{\alpha_2}\sigma/\sqrt{n}$, com $\alpha_1 + \alpha_2 = \alpha$. Se $\alpha_1 \neq \alpha_2$ o intervalo não é simétrico com respeito a μ. O comprimento do intervalo é $L = (z_{\alpha_1} + z_{\alpha_2})\sigma/\sqrt{n}$. Demonstrar que $\alpha_1 = \alpha_2$ minimiza L.

Resolução:
Observe o esboço abaixo:



A situação que considerámos no ex. 1 era um caso particular desta onde $z_{\alpha_1} = z_{\alpha_2} = z_{\alpha/2}$ e, portanto, $\alpha_1 = \alpha_2 = \alpha/2$ (ou seja, o intervalo era centrado na média). Neste exercício, pretende-se demonstrar que essa é a situação que minimiza a largura do intervalo obtido (denotada por L).

Assim, entre todos os pares de números reais $(z_{\alpha_1}, z_{\alpha_2})$ tais que $P(-z_{\alpha_1} \leq Z \leq z_{\alpha_2}) = 1-\alpha$, queremos encontrar aquele que minimiza $L(z_{\alpha_1}, z_{\alpha_2}) = (z_{\alpha_1} + z_{\alpha_2})\sigma/\sqrt{n}$. Isto traduz-se num problema de minimização com uma restrição:

$$\min_{z_{\alpha_1}, z_{\alpha_2}} L(z_{\alpha_1}, z_{\alpha_2}) \quad \text{sujeito a} \quad P(-z_{\alpha_1} \leq Z \leq z_{\alpha_2}) = 1 - \alpha.$$

Seja $g(z_{\alpha_1}, z_{\alpha_2}) = P(-z_{\alpha_1} \leq Z \leq z_{\alpha_2}) - 1 + \alpha$. O problema pode então ser reescrito como

$$\min_{z_{\alpha_1}, z_{\alpha_2}} L(z_{\alpha_1}, z_{\alpha_2}) \quad \text{sujeito a} \quad g(z_{\alpha_1}, z_{\alpha_2}) = 0,$$

e tem assim a forma típica de um problema a ser resolvido pelo método dos multiplicadores de Lagrange. Seja $(z^*_{\alpha_1}, z^*_{\alpha_2})$ o minimizante que procuramos. Pelo método referido, sabemos que existe um número real λ (chamado multiplicador de Lagrange) tal que:

$$\nabla L(z^*_{\alpha_1}, z^*_{\alpha_2}) = \lambda \nabla g(z^*_{\alpha_1}, z^*_{\alpha_2}).$$

Esta igualdade e a restrição $g(z^*_{\alpha_1}, z^*_{\alpha_2}) = 0$ permitem-nos encontrar $z^*_{\alpha_1}$, $z^*_{\alpha_2}$ e λ. Vamos então calcular os gradientes necessários:

$$\frac{\partial L}{\partial z_{\alpha_1}}(z_{\alpha_1}, z_{\alpha_2}) = \frac{\partial L}{\partial z_{\alpha_2}}(z_{\alpha_1}, z_{\alpha_2}) = \frac{\sigma}{\sqrt{n}}.$$

Note que

$$g(z_{\alpha_1}, z_{\alpha_2}) = P(-z_{\alpha_1} \leq Z \leq z_{\alpha_2}) - 1 + \alpha = \int_{-z_{\alpha_1}}^{z_{\alpha_2}} f_Z(z)\, dz - 1 + \alpha,$$

onde f_Z é a função densidade de probabilidade da distribuição normal padrão. Assim,

$$\frac{\partial g}{\partial z_{\alpha_1}}(z_{\alpha_1}, z_{\alpha_2}) = f_Z(-z_{\alpha_1}) = f_Z(z_{\alpha_1}), \quad \frac{\partial g}{\partial z_{\alpha_2}}(z_{\alpha_1}, z_{\alpha_2}) = f_Z(z_{\alpha_2}).$$

Finalmente,

$$\begin{cases} \nabla L(z^*_{\alpha_1}, z^*_{\alpha_2}) = \lambda \nabla g(z^*_{\alpha_1}, z^*_{\alpha_2}) \\ g(z^*_{\alpha_1}, z^*_{\alpha_2}) = 0 \end{cases} \Leftrightarrow \begin{cases} \frac{\partial L}{\partial z_{\alpha_1}}(z^*_{\alpha_1}, z^*_{\alpha_2}) = \lambda \frac{\partial g}{\partial z_{\alpha_1}}(z^*_{\alpha_1}, z^*_{\alpha_2}) \\ \frac{\partial L}{\partial z_{\alpha_2}}(z^*_{\alpha_1}, z^*_{\alpha_2}) = \lambda \frac{\partial g}{\partial z_{\alpha_2}}(z^*_{\alpha_1}, z^*_{\alpha_2}) \\ P(-z^*_{\alpha_1} \leq Z \leq z^*_{\alpha_2}) = 1 - \alpha \end{cases}$$

$$\Leftrightarrow \begin{cases} \sigma/\sqrt{n} = \lambda f_Z(z^*_{\alpha_1}) \\ \sigma/\sqrt{n} = \lambda f_Z(z^*_{\alpha_2}) \\ P(-z^*_{\alpha_1} \leq Z \leq z^*_{\alpha_2}) = 1 - \alpha \end{cases}.$$

As primeiras duas equações implicam $f_Z(z^*_{\alpha_1}) = f_Z(z^*_{\alpha_2})$, logo $z^*_{\alpha_2} = \pm z^*_{\alpha_1}$. Note que $z^*_{\alpha_2} = -z^*_{\alpha_1} \implies P(-z^*_{\alpha_1} \leq Z \leq z^*_{\alpha_2}) = 0 \neq 1 - \alpha$, logo esta solução não é válida. Assim, concluímos que L é mínimo quando $z^*_{\alpha_2} = z^*_{\alpha_1}$, ou seja, quando $\alpha_1 = \alpha_2$. $\square$

5. Um teste de impacto foi realizado em 20 tubos de PVC. A média amostral é 1.25 e o desvio padrão amostral é $s = 0.25$. Calcule o intervalo de confiança inferior de 99% para a média. Assuma que a grandeza medida segue uma distribuição normal.

Resolução:

Note que, agora, não conhecemos o desvio padrão da distribuição (σ), mas apenas o desvio padrão das amostras recolhidas (s). Como não conhecemos σ, não podemos utilizar o mesmo processo que anteriormente para determinar o intervalo de confiança. Considere então a v.a. $\bar{X}$, definida como habitualmente (média amostral), e a v.a. S^2, que representa a variância amostral, definida por:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

É possível demonstrar (teorema) que, se as amostras $X_1, X_2, \cdots, X_n$ forem independentes e se $X_i \sim N(\mu, \sigma^2)$, $i \in \{1, \cdots, n\}$, então:

- $\bar{X}$ e S^2 são v.a. independentes;
- $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, isto é, a v.a. $(n-1)S^2/\sigma^2$ segue uma distribuição qui-quadrado com $n-1$ graus de liberdade;
- $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, isto é, a v.a. $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ segue uma distribuição t de Student com $n-1$ graus de liberdade.

Vamos tirar partido do último ponto para resolver o presente exercício. A distribuição t de Student é simétrica em torno de 0, tal como a distribuição normal padrão, logo, para um intervalo centrado, temos:

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1-\alpha.$$

Agora, tal como nos exercícios anteriores, o intervalo de confiança obtém-se utilizando $\bar{x}$ e s no lugar de $\bar{X}$ e S, respetivamente, e reescrevendo as desigualdades em ordem a μ:

$$\begin{aligned} -t_{\alpha/2,n-1} \leq \left.\frac{\bar{X}-\mu}{S/\sqrt{n}}\right|_{\bar{X}=\bar{x},\, S=s} \leq t_{\alpha/2,n-1} &\Leftrightarrow \bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \\ &\Leftrightarrow \mu \in \left[\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]. \end{aligned}$$

Neste exercício, pede-se um intervalo de confiança inferior, que, de forma análoga aos casos anteriores, será dado por:

$$\mu \geq \bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}.$$

Temos $n = 20$, $\bar{x} = 1.25$, $s = 0.25$ e $1 - \alpha = 0.99$. Da tabela da distribuição t de Student, obtemos $t_{0.01,19} \approx 2.539$, logo $\mu \geq 1.11$ é o intervalo de confiança obtido.

<u>Nota:</u> Repare que os valores numéricos dados no enunciado deste exercício são exatamente iguais aos dados no ex. 2. Porém, desta vez obtivemos $\mu \geq 1.11$ e antes tínhamos $\mu \geq 1.12$. A única diferença nos enunciados é que, no ex. 2., conhecíamos o desvio padrão da distribuição (σ) e, neste exercício, temos apenas o desvio padrão amostral (s). Assim, em virtude de não conhecermos o valor exato de σ mas apenas uma estimativa do seu valor (s), a incerteza associada ao problema aumenta. Consequentemente, para o mesmo nível de confiança $1-\alpha$, o intervalo que obtivemos neste exercício é maior do que o obtido no ex. 2.

Isto é facilmente percetível comparando a distribuição t de Student com a distribuição normal padrão. Ambas as distribuições têm média 0. No entanto, a variância da distribuição normal padrão é igual a 1, enquanto a variância da distribuição t de Student com $n-1$ graus de liberdade é $(n-1)/(n-3) > 1$. Assim, a distribuição t de Student afasta-se mais de 0 e, por esse motivo, temos sempre $|t_{\alpha,n-1}| > |z_\alpha|$. Note, no entanto, que quando n aumenta, a variância da distribuição $t(n-1)$ diminui e, no limite quando $n \to \infty$, a variância é igual a 1, coincidindo com a variância da distribuição normal padrão. De facto, prova-se que a distribuição $t(n-1)$ converge para a distribuição $N(0,1)$ quando $n \to \infty$, pelo que $\lim_{n\to\infty} t_{\alpha,n-1} = z_\alpha$.

Este resultado é igualmente intuitivo: se a amostra recolhida for muito grande, o desvio padrão amostral tenderá a ser muito próximo do desvio padrão da distribuição (ou seja, $s \approx \sigma$), pelo que podemos tomá-lo como se fosse o verdadeiro σ e assim utilizar a distribuição normal padrão em vez da distribuição t de Student.

6. A espessura das paredes de 25 garrafas de dois litros foi medida por um engenheiro de controle de qualidade. A média amostral foi 4.05 mm, e o desvio padrão amostral foi de $s = 0.08$ mm. Calcule o intervalo de confiança inferior de 95% da média μ. Assuma que a espessura das paredes segue uma distribuição normal.

   Solução:

   $t_{0.05,24} = 1.711$, $\mu \geq 4.02$mm. (compare com o resultado obtido no ex. 3.)

7. Suponha que X tenha distribuição $N(\mu, \sigma^2)$. Uma amostra de tamanho 30 fornece os seguintes valores: $\sum_{i=1}^{30} x_i = 700.8$, $\sum_{i=1}^{30} x_i^2 = 16395.8$. Determine um intervalo de confiança de 95% (bilateral) para μ.

   Resolução:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{700.8}{30} = 23.36,$$

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) \\ &= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{2\bar{x}}{n-1}\sum_{i=1}^{n} x_i + \frac{n}{n-1}\bar{x}^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2 \\ &= \frac{16395.8}{29} - \frac{30}{29} \times 23.36^2 \approx 0.866 \implies s \approx 0.931, \\ & t_{0.025,29} \approx 2.045, \quad \mu \in [23.01,\ 23.71]. \end{aligned}$$

8. Uma amostra de tamanho 1000 de cancro do pulmão, 823 casos resultaram na morte do paciente nos seguintes 10 anos. Indique o tamanho da amostra para ter pelo menos 95% de confiança de que o erro de estimação seja menor do que 0.03.

   Resolução:

   Seja $X_i \in \{0, 1\}$ a v.a. que indica se o paciente i morreu ($X_i = 1$) ou sobreviveu ($X_i = 0$) nos 10 anos seguintes. O parâmetro $p = P(X_i = 1)$ é desconhecido, por isso recolheu-se uma amostra de tamanho $n = 1000$ para o estimar. Na amostra, observou-se que 823 pacientes morreram, logo $\sum_{i=1}^{n} x_i = 823$.

   Considere a v.a. $\bar{X}$ definida como habitualmente: $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$. Queremos determinar a distribuição de $\bar{X}$. Neste caso, como cada X_i não segue uma distribuição normal, a distribuição de $\bar{X}$ não será exatamente uma distribuição normal. Contudo, como a amostra é de grande dimensão e $\bar{X}$ resulta da soma dessas amostras independentes, o teorema do limite central garante-nos que a distribuição normal será uma boa aproximação da distribuição de $\bar{X}$. Resta-nos então determinar a respetiva média e variância (em função do parâmetro p). Observe que $E(X_i) = p$ e $V(X_i) = p(1-p)$, logo $E(\bar{X}) = p$ e, como se assume amostras independentes, $V(\bar{X}) = p(1-p)/n$. Assim, $\bar{X} \sim N(p, p(1-p)/n)$, pelo que $\frac{\bar{X}-p}{\sqrt{p(1-p)/n}} \sim N(0, 1)$. Como consequência da lei dos grandes números, $\bar{X}$ aproxima-se de p para n grande, tornando-se possível demonstrar que a distribuição da v.a. $\frac{\bar{X}-p}{\sqrt{\bar{X}(1-\bar{X})/n}}$ também se aproxima de $N(0, 1)$ quando n é grande. Agora, como habitualmente, temos:

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

   Para construir o intervalo de confiança para p, toma-se, no lugar de $\bar{X}$, a proporção $\hat{p} = (1/n)\sum_{i=1}^{n} x_i$, estimada a partir da amostra recolhida:

$$\begin{aligned} -z_{\alpha/2} \leq \left.\frac{\bar{X} - p}{\sqrt{\bar{X}(1-\bar{X})/n}}\right|_{\bar{X}=\hat{p}} \leq z_{\alpha/2} &\Leftrightarrow \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &\Leftrightarrow p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]. \end{aligned}$$

No presente exercício, temos $n = 1000$, $\hat{p} = 823/1000 = 0.823$ e $1 - \alpha = 0.95$, logo $z_{\alpha/2} = z_{0.025} = 1.960$ e $p \in [0.823 \pm 0.024] = [0.799,\ 0.847]$. Assim, o erro de estimação do intervalo obtido (que corresponde à semi-amplitude do intervalo) foi de 0.024. O exercício pede que determinemos o tamanho mínimo da amostra (ou seja, o mínimo valor de n) que garante um erro de estimação menor do que $\varepsilon = 0.03$. Assumindo que, na nova amostra, a proporção observada se manteria igual ($\hat{p} = 0.823$), temos:

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < \varepsilon \Leftrightarrow n > z_{\alpha/2}^2 \frac{\hat{p}(1-\hat{p})}{\varepsilon^2} \approx 622.$$

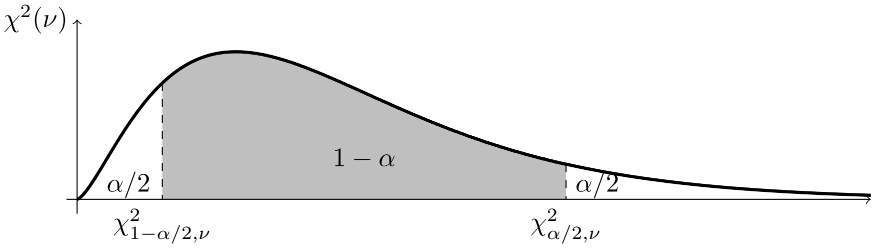
Se quisermos ser mais cautelosos, podemos remover o pressuposto de que $\hat{p}$ se mantém e considerar o pior caso, $\hat{p} = 0.5$ (que maximiza $\hat{p}(1-\hat{p})$). Nesse caso, teríamos $n > 1067$.

9. Uma centena de componentes foi ensaiada, e 93 deles funcionaram mais de 500 horas. Determine um intervalo de confiança a 95% (bilateral) para $p = P$(componente funcione mais de 500 horas).

   Resolução:
   Aplicam-se exatamente os resultados enunciados no ex. 8. Temos $n = 100$, $\hat{p} = 93/100 = 0.93$, $\alpha = 0.05$, logo $z_{\alpha/2} = z_{0.025} = 1.960$ e $p \in \left[0.93 \pm 1.960\sqrt{\frac{0.93(1-0.93)}{100}}\right] \approx [0.88,\ 0.98]$.

10. Suponha que X tenha distribuição $N(\mu, \sigma^2)$ onde μ e σ^2 são desconhecidos. Uma amostra de tamanho 15 forneceu os valores $\sum_{i=1}^{15} x_i = 8.7$ e $\sum_{i=1}^{15} x_i^2 = 27.3$. Determine um intervalo de confiança de 95% (bilateral) para σ^2.

    Resolução:
    Para obter um intervalo de confiança para a variância da distribuição normal (isto é, para σ^2), vamos recordar um resultado enunciado na resolução do ex. 5.: sendo as amostras $X_1, X_2, \cdots, X_n$ independentes e com distribuição $N(\mu, \sigma^2)$ (onde μ e σ são desconhecidos), temos $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$. Ao contrário das distribuições normal e t de Student, a distribuição do qui-quadrado não é simétrica, tal como pode observar na figura abaixo.

    Assim, teremos de procurar na tabela desta distribuição dois valores $\chi^2_{\alpha/2,n-1}$ e $\chi^2_{1-\alpha/2,n-1}$ tais que:

    $$P\left(\frac{(n-1)S^2}{\sigma^2} \geq \chi^2_{\alpha/2,n-1}\right) = \frac{\alpha}{2} \quad \text{e} \quad P\left(\frac{(n-1)S^2}{\sigma^2} \geq \chi^2_{1-\alpha/2,n-1}\right) = 1 - \frac{\alpha}{2}.$$

    Pode confirmar pela figura que estes valores garantem:

    $$P\left(\chi^2_{1-\alpha/2,n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}\right) = 1 - \alpha.$$

    O intervalo obtém-se substituindo S^2 pela variância da amostra observada, s^2:

    $$\chi^2_{1-\alpha/2,n-1} \leq \left.\frac{(n-1)S^2}{\sigma^2}\right|_{S^2=s^2} \leq \chi^2_{\alpha/2,n-1} \Leftrightarrow \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}.$$

    No presente exercício, temos $n = 15$, $\bar{x} = 0.58$, $s^2 \approx 1.59$ (ver resol. do ex. 7.), $1 - \alpha = 0.95$, logo $\chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,14} \approx 5.63$, $\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,14} \approx 26.12$ e $\sigma^2 \in [0.85,\ 3.95]$.

11. A espessura das paredes de 25 garrafas de dois litros foi medida por um engenheiro de controlo de qualidade. A média amostral foi 4.05 mm, e o desvio padrão amostral foi de $s = 0.08$ mm. Calcule o intervalo de predição de 90% para a amostra seguinte assumindo que a espessura é normalmente distribuida.

Resolução:

Observadas n amostras $x_1, x_2, \cdots, x_n$, pretende-se estimar um intervalo (chamado *intervalo de predição*) para uma amostra aleatória, X. Como se assume amostras independentes e normalmente distribuídas (com μ e σ desconhecidos), temos $\bar{X} \sim N(\mu, \sigma^2/n)$ e $X \sim N(\mu, \sigma^2)$. Então,

$$\bar{X} - X \sim N\left(\mu - \mu, \frac{\sigma^2}{n} + \sigma^2\right) = N\left(0, \sigma^2\left(1 + \frac{1}{n}\right)\right).$$

Utilizando um corolário do teorema enunciado no ex. 5., conclui-se que:

$$\frac{\bar{X} - X}{S\sqrt{1 + \frac{1}{n}}} \sim t(n-1).$$

O intervalo centrado será tal que:

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\bar{X} - X}{S\sqrt{1 + \frac{1}{n}}} \leq t_{\alpha/2,n-1}\right) = 1 - \alpha,$$

e, finalmente, o intervalo de predição para X obtém-se substituindo $\bar{X}$ e S pelos respetivos valores observados na amostra:

$$-t_{\alpha/2,n-1} \leq \left.\frac{\bar{X} - X}{S\sqrt{1 + \frac{1}{n}}}\right|_{\bar{X}=\bar{x},\, S=s} \leq t_{\alpha/2,n-1} \Leftrightarrow X \in \left[\bar{x} \pm t_{\alpha/2,n-1} s\sqrt{1 + \frac{1}{n}}\right].$$

No exercício dado, temos $n = 25$, $\bar{x} = 4.05$mm, $s = 0.08$mm e $1 - \alpha = 0.90$, logo $t_{\alpha/2,n-1} = t_{0.05,24} \approx 1.711$ e $X \in [3.91,\ 4.19]$ (mm).