Licenciatura em Engenharia Informática
Ano Letivo 2021/22

Métodos Estatísticos

Exercícios - Resolução

Deolinda M.L.D. Rasteiro, Luís M. Margalho, M. do Céu Marques

Instituto Superior de Engenharia de Coimbra
Departamento de Física e Matemática

1. Probabilidades

1. *Uma caixa contém 5 lâmpadas das quais 2 são defeituosas. Estas têm os números 3 e 5. Considere a experiência aleatória "extracção de duas lâmpadas ao acaso, uma a seguir à outra, sem reposição da primeira".*

   (a) *Construa o espaço de resultados associado a esta experiência aleatória.*

   (b) *Defina por extenso os acontecimentos:*

   $$\begin{aligned} A &= \{\textit{saída de lâmpada defeituosa na primeira tiragem}\}; \\ B &= \{\textit{saída de lâmpada defeituosa na segunda tiragem}\}; \\ C &= \{\textit{saída de duas lâmpadas defeituosas}\}; \\ D &= \{\textit{não sair qualquer lâmpada defeituosa}\}; \end{aligned}$$

   (c) *Se as lâmpadas forem extraídas ao acaso, os resultados possíveis são equiprováveis. Calcule a probabilidade dos acontecimentos A, B, C e D.*

   (d) *Resolva as alíneas anteriores supondo que, agora, que a experiência aleatória é "extracção de duas lâmpadas ao acaso, uma a seguir à outra,* ***com*** *reposição da primeira".*

2. Calcule a probabilidade de, ao lançar três vezes uma moeda equilibrada, obter:

   (a) duas caras;

   (b) pelo menos uma cara.

3. Lança-se simultaneamente um dado e uma moeda equilibrados.

   (a) Construa o espaço de resultados associado a esta experiência aleatória.

   (b) Defina por extenso os acontecimentos:

   $$\begin{aligned} A &= \{\text{saída de coroa e número par}\}; \\ B &= \{\text{saída de cara e número ímpar}\}; \\ C &= \{\text{saída de múltiplos de três}\}; \end{aligned}$$

   e determine as respectivas probabilidades.

4. Sejam A, B, e C acontecimentos de Ω tais que:

   $$A \cup B \cup C = \Omega, \quad P(A) = 0.3, \quad P(\overline{B}) = 0.7, \quad P(C) = 0.5 \;\text{ e }\; A \cap B = C \cap B = \emptyset.$$

   Determine $P(A \cap C)$.

5. Sejam A e B acontecimentos de um mesmo espaço de probabilidade Ω, tais que $P(A) = 0.7$, $P(B) = 0.6$ e $P(A \cup B) - P(A \cap B) = 0.3$. Calcule:

   (a) $P(\overline{B})$;

   (b) $P(A \cup B)$ e $P(A \cap B)$.

6. Supondo que A e B são acontecimentos independentes com probabilidade não nula prove que os acontecimentos A e $\overline{B}$, $\overline{A}$ e $\overline{B}$, $\overline{A}$ e B também são independentes.

7. Uma empresa fabrica aparelhos elétricos em duas cadeias de produção A e B. Sabe-se que a probabilidade de um desses artigos ser exportado é 0.2 se produzido pela cadeia A e 0.5 se produzido pela cadeia B. Além disso, a proporção de artigos provenientes da cadeia A é 52%. Escolhe-se, ao acaso, um artigo da produção desta empresa.

   (a) Determine a probabilidade do artigo ser exportado.

   (b) Sabendo que o artigo não foi exportado, qual a probabilidade dele ter sido produzido pela cadeia B?

8. A central telefónica do INEM de uma grande cidade recebe chamadas, umas genuínas e outras falsas, isto é, correspondentes ou não a verdadeiros acidentes. A central recebe na totalidade 2% de chamadas falsas. Destas, 20% são efectuadas durante o período da manhã, 40% durante o período da tarde e as restantes à noite. Das chamadas genuínas recebidas na central, 30% são feitas durante a manhã.

   (a) Mostre que a percentagem de chamadas recebidas na central durante o período da manhã é de 29.8%.

   (b) Considerando que a probabilidade de uma chamada, recebida na central, ser efectuada no período da tarde é de 40%, calcule a probabilidade de uma chamada ser feita durante a noite dado que é uma chamada genuína.

9. Em determinada linha de montagem 2% das peças ficam mal colocadas. Um programa para detectar falhas de montagem tem as seguintes propriedades:

   - se a peça está mal colocada, o programa indica essa falha com probabilidade 0.99;
   - se a peça está correctamente colocada, o programa indica falha com probabilidade 0.005.

   (a) Determine a probabilidade de, ao ser efectuado o referido teste, o programa indicar falha.

   (b) Se o teste indicar a existência de uma falha, qual a probabilidade de efectivamente existirem peças mal colocadas?

10. Uma empresa de fabrico de válvulas de televisão dispões de três sectores de produção: A, B e C. Sabe-se que:

    - a percentagem de válvulas da marca A é 50%;
    - a percentagem de válvulas defeituosas é 10%;
    - em C não há válvulas defeituosas;
    - 2% das válvulas provêm de B e são defeituosas.

Escolhe-se aleatoriamente uma válvula de televisão da produção da empresa.

(a) Mostre que a probabilidade da válvula ser defeituosa, sabendo que provém de A é 0.16.

(b) Calcule a probabilidade da válvula não provir de B sabendo que é defeituosa.

(c) Sabendo que, das válvulas não defeituosas 40% provêm de C, qual a probabilidade da válvula ser proveniente de C?

11. Dos utilizadores de telefones móveis duma determinada localidade, 50% estão ligados à rede A, 40% à rede B e 10% à rede C. Após um estudo de opinião de mercado conclui-se que:

- 70% dos utilizadores estão satisfeitos com o serviço;
- dos utilizadores ligados à rede A, 80% estão satisfeitos;
- dos utilizadores satisfeitos com o serviço, 10% estão ligados à rede C.

Determine a percentagem de utilizadores:

(a) da rede B que estão satisfeitos com o serviço;

(b) não satisfeitos com o serviço, sabendo que estes não estão ligados à rede C.

12. O fabrico de uma peça consta de duas operações. Inicialmente a peça é moldada numa máquina M e, em seguida, passa por uma de duas impressoras, I_1 ou I_2. A probabilidade de uma peça apresentar defeito de moldagem é 0.4 e 70% das peças são impressas em I_1. Além disso, a probabilidade de surgir um defeito de impressão é de 0.05 para I_1 e de 0.02 para I_2. Note que defeitos de moldagem e de impressão são independentes entre si.

No final de determinado dia de laboração, da produção total da fábrica retira-se uma peça ao acaso.

(a) Qual a probabilidade da peça ter defeitos de impressão?

(b) Qual a probabilidade da peça apresentar um qualquer defeito?

(c) Supondo que a peça apresenta defeito de impressão, calcule a probabilidade de ter sido impressa em I_1.

2. Variáveis Aleatórias e Distribuições de Probabilidade Discretas

9. *Uma loja quer vender rapidamente os 100 computadores portáteis que tem em armazém, pelo que realizou uma promoção com descontos oferecendo o sistema operativo. O processo de instalação do sistema operativo não é completamente fiável e 10 dos portáteis necessitarão de assistência complementar.*

Uma empresa comprou na loja 20 portáteis, seleccionados aleatoriamente no armazém.

(a) *Indique (justificando) a lei de probabilidade do número de portáteis comprados pela empresa, que necessitarão de assistência.*

Defina-se então a v.a.

$X =$ "Número de portáteis que necessitarão de assistência, no conjunto dos portáteis comprados pela empresa".

Identificam-se um número finito de $N = 100$ objetos (portáteis), dos quais $M = 10$ são do tipo *necessitarão de assistência* e os restantes, $N - M = 90$, de outro tipo, *não necessitarão de assistência*. Assumindo que o levantamento de cada portátil no armazém para vender à empresa, no total de $n = 20$, é feito ao acaso (e "sem reposição"), X segue a lei Hipergeométrica de parâmetros $n = 20$, $N = 100$ e $M = 10$, $X \sim \mathcal{H}(20, 100, 10)$, ou seja,

para $x \in S = \{0, 1, ..., 10\}$, $P(X = x) = \dfrac{\begin{pmatrix} 10 \\ x \end{pmatrix}\begin{pmatrix} 90 \\ 20-x \end{pmatrix}}{\begin{pmatrix} 100 \\ 20 \end{pmatrix}}$

(b) *Qual a probabilidade de nenhum dos portáteis apresentar problemas?*

$$P(X = 0) = \frac{\begin{pmatrix} 10 \\ 0 \end{pmatrix}\begin{pmatrix} 90 \\ 20 \end{pmatrix}}{\begin{pmatrix} 100 \\ 20 \end{pmatrix}} = 0.0951$$

Lembrete: $C_k^m \equiv \begin{pmatrix} m \\ k \end{pmatrix} = \dfrac{m!}{(m-k)!\,k!}$ e $0! = 1$.

(c) *Indique uma expressão matemática que dê a probabilidade de mais de 5 portáteis necessitarem de assistência.*

$$P(X > 5) = P(X \geq 6) = P(X = 6) + P(X = 7) + ... + P(X = 10)$$
$$= \sum_{x=6}^{10} P(X = x) = \sum_{x=6}^{10} \frac{C_x^{10}\, C_{20-x}^{90}}{C_{20}^{100}}$$

(d) *Cada um dos portáteis é vendido a* 1 520 *euros e a sua eventual assistência custará à loja* 55 *euros. Indique o lucro esperado da loja com a venda dos portáteis à empresa.*

Defina-se a v.a,

$L =$"Lucro da loja com a venda dos portáteis à empresa".

Sem mais informação, podemos definir a seguinte equação para L,

$$L = \underbrace{20 \times 1520}_{\text{parcela fixa}} - \underbrace{55 \times X}_{\text{parcela aleatória}} = 30\,400 - 55 \times X.$$

Pretende-se o lucro esperado, ou seja,

$E(L) = E(30\,400 - 55 \times X)$

$= 30\,400 - 55 \times E(X)$ (das propriedades do valor esperado)

$= 30\,400 - 55 \times 2$ (uma vez que $X \sim \mathcal{H}(20, 100, 10)$, $E(X) = \dfrac{20 \times 10}{100}$)

$= 30\,290$ (euros)

10. *Sabe-se que 40% dos alunos do IPC concordam com o sistema de avaliação intercalar por escolha múltipla. Escolhidos 10 alunos ao acaso,*

Para ambas as alíneas seja

$X =$ "Número de alunos do IPC que concordam com o sistema de avaliação intercalar por escolha múltipla, em 10 alunos selecionados ao acaso".

$X \dot{\sim} \mathcal{B}(10, 0.4)$

(a) *a probabilidade de menos de 3 concordarem com este sistema de avaliação é igual a*:

$P(X < 3) = P(X \leq 2) = 0.1673$ (**B**)

(b) a probabilidade de mais de 3 concordarem com este sistema de avaliação é igual a:

$P(X > 3) = 1 - P(X \leq 3) = 1 - 0.3823 = 0.6177$ (**C**)

14. *O número de chamadas que chegam num período de 5 minutos à central telefónica de uma empresa é uma variável aleatória com distribuição de* Poisson *de parâmetro 10.*

Do enunciado decorre de imediato

$X =$ "Número de chamadas que chegam num período de 5 minutos à central telefónica de uma empresa"

$X \sim \mathcal{P}(10)$.

E$(X) =$V$(X) = 10$

(a) *Calcule a probabilidade de, num período de 5 minutos,*

i. *chegarem exactamente 8 chamadas;*

$P(X = 8) = 0.1126$ (consulta direta na máquina)
ou, utilizando a tabela

$$\begin{aligned} P(X=8) &= P(X \leq 8) - P(X \leq 7) \\ &= 0.3328 - 0.2202 \\ &= 0.1126. \end{aligned}$$

ii. *chegarem menos de 5 chamadas;*

$$P(X < 5) = P(X \leq 4) = 0.0293.$$

iii. *chegarem, no mínimo, 3 chamadas;*

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - P(X \leq 2) \\ &= 1 - 0.0028 \\ &= 0.9972. \end{aligned}$$

iv. *não chegar alguma chamada.*

$$P(X = 0) = 0.0000454 \backsimeq 0.5 \times 10^{-4}.$$

(b) *Qual a probabilidade de chegarem à central telefónica 35 chamadas, num período de 10 minutos consecutivos? Justifique.*

Para determinar a probabilidade pretendida repare-se que o período de contagem das chamadas se alterou!!! Passou de 5 para 10 minutos!!!

Seja X_i = "Número de chamadas que chegam à central de uma empresa no $i^{\underline{esimo}}$ período de 5 minutos", $i = 1, 2$.

$X_1 + X_2$ = "Número de chamadas que chegam à central de uma empresa num período de 10 minutos consecutivos".

Admitindo que o número de chamadas em cada período de 5 minutos é independente, pela aditividade da *Poisson*, sabemos que

$$X_1 + X_2 \sim \mathcal{P}(10 + 10) = \mathcal{P}(20)$$

Assim,

$$P(X_1 + X_2 = 35) = 0.00069 \sim 0.0007.$$

16. *O número de visitantes que entra num* cibercafe *ao longo dos vários períodos diários segue uma lei de Poisson. No entanto, o número médio de visitantes varia consoante o período do dia: no período da manhã espera-se* 3 *visitantes e no período da tarde* 15*. Assuma independência entre o número de visitantes ao* cibercafe *nos dois períodos diários.*

Seja $X_M =$ "Número de visitantes que entra num *cybercafe* no período da manhã"e

$X_T =$ "Número de visitantes que entra num *cybercafe* no período da tarde"

$\mathrm{E}(X_M) = 3 \Rightarrow X_M \sim \mathcal{P}(3)$

$\mathrm{E}(X_T) = 15 \Rightarrow X_T \sim \mathcal{P}(15)$

X_M e X_T são independentes.

(a) *Qual a probabilidade de numa manhã de um dia qualquer, o número de visitantes ao* cibercafe *ser pelo menos cinco?*

$$\begin{aligned} P(X_M \geq 5) &= 1 - P(X_M < 5) \\ &= 1 - P(X_M \leq 4) \\ &= 1 - 0.8153 \\ &= 0.1847. \end{aligned}$$

(b) *Qual a probabilidade de, num dia qualquer, o número total de visitantes ao* cibercafe *nos períodos da manhã e da tarde ser menor que 31?*

Seja $N =$ "Número total de visitantes ao *cybercafe*"$= X_M + X_T$

Sendo X_M e X_T independentes, pela aditividade da distribuição de *Poisson*, $N \sim \mathcal{P}(3+15) = \mathcal{P}(18)$

$P(N < 31) = P(N \leq 30) = 0.9967.$

(c) *Qual a probabilidade de, num dia qualquer, o número de visitantes na manhã ser igual a 5 e na tarde ser igual a 20?*

$$\begin{aligned} P(X_M = 5 \bigcap X_T = 20) &= P(X_M = 5)P(X_T = 20) \\ &= 0.1008 \times 0.0418 \\ &= 0.0042. \end{aligned}$$

18. *O número de petroleiros que chegam em cada dia a determinada refinaria é uma variável aleatória com distribuição de* Poisson *de média 2. As actuais instalações do porto podem atender 3 petroleiros por dia; se acontecer que mais de 3 navios pretendam entrar no porto, os excedentes a 3 deverão seguir para outro destino.*

Seja $X =$ "Número de petroleiros que chegam em cada dia a determinada refinaria."

$\mathrm{E}(X) = 2 \Rightarrow X \sim \mathcal{P}(2)$

Capacidade de atendimento do porto = 3

Se $X > 3 \Rightarrow$ os excedentes($= X - 3$) vão para outro destino.

(a) *Em determinado dia, qual a probabilidade de se ter de mandar petroleiros para outro porto?*

$$\begin{aligned} P(X > 3) &= 1 - P(X \le 3) \\ &= 1 - 0.8571 \\ &= 0.1429. \end{aligned}$$

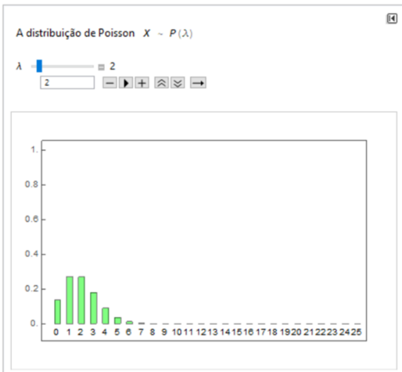
(b) *Qual o número esperado de petroleiros a chegarem por dia?*

O número esperado de petroleiros a chegarem por dia é 2.

(c) *Qual o número mais provável de petroleiros a chegarem por dia?*

O número mais provável de petroleiros é o(s) número(s) com maior probabilidade

$$\begin{gathered} P(X = 0) = 0.1353 \\ P(X = 1) = 0.2707 \\ P(X = 2) = 0.2707 \\ P(X = 3) = 0.1804 \\ \vdots \\ \vdots \\ S_X = \mathbb{N}_0 \ldots \text{será necessário vermos todos os valores? Não!} \end{gathered}$$



(d) *De quanto deverão ser aumentadas as actuais instalações do porto para permitir manobrar todos os petroleiros em 95% dos dias?*

Determinar o valor c - aumento do número de petroleiros tal que

$$P(X \leq 3 + c) = 0.95.$$

Se $3 + c = 4, P(X \leq 3 + c) = 0.9473 \backsimeq 0.95$, então $c = 1$.

As instalações deverão ser aumentadas em um petroleiro.

(e) *Deduza a lei de probabilidade do número de petroleiros a serem* atendidos *por dia.*

Seja Y = "Número de petroleiros a serem atendidos por dia".

A lei de probabilidade de Y é

y	0	1	2	3	c.c.
$P(Y = y)$	$P(X = 0) = 0.1353$	$P(X = 1) = 0.2707$	$P(X = 2) = 0.2707$	$P(X \geq 3) = 0.3233$	0

Nota: $P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.6767$

(f) *Qual o número esperado de petroleiros a serem atendidos por dia?*

$\mathrm{E}(Y) = 0 * 0.1353 + 1 * 0.2707 + 2 * 0.2707 + 3 * 0.3233 = 1.782.$

19. *O número de acidentes de trabalho,* por mês*, numa obra de construção civil é uma v.a. com distribuição de* Poisson *de valor médio 2.*

Seja X = "Número de acidentes de trabalho, por mês, numa obra de construção civil."

$\mathrm{E}(X) = 2 \Rightarrow X \sim \mathcal{P}(2)$

(a) *Determine a probabilidade de não ocorrerem acidentes num determinado mês.*

$P(X = 0) = 0.1353.$

(b) *Calcule a probabilidade de ocorrerem pelo menos 6 acidentes em* 3 meses*.*

X_i = "Número de acidentes de trabalho, no i$^{\underline{esimo}}$ mês, numa obra de construção civil." $i = 1, 2, 3$

$X_i \sim \mathcal{P}(2)$

$Y = X_1 + X_2 + X_3$ = "Número de acidentes de trabalho que ocorrem em 3 meses numa obra de construção civil."

Pela aditividade da distribuição de *Poisson*

$Y \sim \mathcal{P}(2 + 2 + 2) = \mathcal{P}(6)$

$$\begin{aligned} P(Y \geq 6) &= 1 - P(Y < 6) \\ &= 1 - P(Y \leq 5) \\ &= 1 - 0.4457 \\ &= 0.5543. \end{aligned}$$

(c) *Suponha que a obra foi observada durante 6 meses consecutivos. Qual a probabilidade de não ocorrerem acidentes em exactamente 4 meses?*

E= "Observar a obra durante um mês e registar caso não ocorram acidentes."

Repetir a experiência E, 6 vezes sempre nas mesmas condições.

$W =$ "Número de meses, em 6, em que não se registaram acidentes de trabalho numa obra de construção civil."

$W \sim \mathcal{B}(6, P(X=0)) = \mathcal{B}(6, 0.1353)$

$P(W = 4) = 0.0038.$

20. *Seja X a v.a. relativa ao número de defeitos encontrados numa unidade de determinado artigo e Y a v.a. que indica o número da fábrica que o produziu. A tabela seguinte representa a função de probabilidade conjunta do vector (X, Y), ou seja,*

$p_{XY}(x, y) = P(X = x \cap Y = y)$, *neste caso para todo o par* $(x, y) \in S = S_X \times S_Y$:

Y \ X	0	1	2	3
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{8}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$

(a) *Determine as leis de probabilidade marginais de X e Y.*

função de probabilidade marginal de X:

$p_X(x) = P(X = x) = \sum_{\mathbf{y} \in \mathbb{R}} p_{XY}(x, \mathbf{y}) = \sum_{\mathbf{y} \in \mathbb{R}} P(X = x \cap Y = \mathbf{y}), \ \forall x \in \mathbb{R}$

função de probabilidade marginal de Y:

$p_Y(y) = P(Y = y) = \sum_{\mathbf{x} \in \mathbb{R}} p_{XY}(\mathbf{x}, y) = \sum_{\mathbf{x} \in \mathbb{R}} P(X = \mathbf{x} \cap Y = y), \ \forall y \in \mathbb{R}$

Cálculos...

$Y \backslash X$	0	1	2	3	$p_Y(y)$
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{1}{2}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
$p_X(x)$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{5}{16}$	$\frac{6}{16}$	

Assim,

x	0	1	2	3	c.c
$p_X(x)$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{5}{16}$	$\frac{6}{16}$	0

y	1	2	c.c
$p_Y(y)$	$\frac{1}{2}$	$\frac{1}{2}$	0

(b) *Calcule $P(X=2)$, $P(X \geq 2)$, $P(X \leq Y)$ e $P(Y=3)$.*

$P(X=2) = p_X(2) = \frac{5}{16}$

$P(X \geq 2) = p_X(2) + p_X(3) = \frac{11}{16}$ $\left(\textbf{ou } 1 - P(X<2) = 1 - P(X \leq 1) = 1 - \frac{5}{16}\right)$

$P(X \leq Y) = \sum_{(x,y):x \leq y} p_{XY}(x,y) = p_{XY}(0,1) + p_{XY}(1,1) + p_{XY}(0,2) + p_{XY}(1,2) + p_{XY}(2,2) = \frac{7}{16}$

$P(Y=3) = p_Y(3) = 0$

(c) *Determine $E(X)$, $V(X)$, Cov (X,Y) e ρ_{XY}.* (**Nota:** E$(Y) = 1.5$, V$(Y) = 0.25$).

$$E(X) = \sum_x x \cdot p_X(x) = \frac{30}{16} = 1.875,$$

$$E(X^2) = \sum_x x^2 \cdot p_X(x) = \frac{76}{16} = 4.75,$$

$$V(X) = E(X^2) - E^2(X) = \frac{76}{16} - \left(\frac{30}{16}\right)^2 = \frac{79}{64} = 1.234$$

$cov(X,Y) = E(XY) - E(X)E(Y)$, com

$$E(XY) = \sum_x \sum_y x \cdot y \cdot p_{XY}(x,y) = 1 \cdot 1 \cdot \frac{1}{16} + 1 \cdot 2 \cdot \frac{1}{16} + ... + 3 \cdot 2 \cdot \frac{1}{4} = \frac{47}{16}$$

Então, $cov(X,Y) = \frac{47}{16} - \frac{30}{16} \times 1.5 = \frac{1}{8} = 0.125$

e $\rho_{XY} = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{0.125}{\sqrt{1.234 \times 0.25}} = 0.225$ (correlação linear positiva, mas fraca)

(d) *O número de defeitos que um artigo apresenta é independente da fábrica que o produziu?*

As variáveis X e Y são independentes se e só se

$$\underline{\forall (x,y)} \in \mathbb{R}^2,\ p_{XY}(x,y) = p_X(x) \times p_Y(y)$$

Assim, uma vez que $p_{XY}(0,1) \neq p_X(0) \times p_Y(1)$, $\frac{1}{8} \neq \frac{3}{16} \times \frac{1}{2}$, por exemplo, concluímos que X e Y não são independentes.

(e) *Sabendo que determinado artigo foi produzido pela fábrica 2, qual a probabilidade de apresentar defeitos?*

$$P(X > 0/Y = 2) = \frac{P(X > 0 \cap Y = 2)}{P(Y = 2)} = \frac{p_{XY}(1,2) + p_{XY}(2,2) + p_{XY}(3,2)}{p_Y(2)} = \frac{7/16}{1/2} = \frac{7}{8}$$

ou

$$P(X > 0/Y = 2) = 1 - P(X \leq 0/Y = 2) = 1 - \frac{P(X \leq 0 \cap Y = 2)}{P(Y = 2)} = 1 - \frac{1/16}{1/2} = 1 - \frac{1}{8} = \frac{7}{8}$$

21. *A tabela seguinte indica a função de probabilidade conjunta das variáveis aleatórias X e Y:*

Y / X	-1	0	1
-1	0	p	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

(a) *Determine o valor de p, justificando a sua resposta.*

Das propriedades da função de probabilidade conjunta,

$$0 \leq p_{XY}(x,y) \leq 1$$

de onde $0 \leq p \leq 1$.

Além disso,

$$\sum_x \sum_y p_{XY}(x,y) = 1$$

o que equivale a

$$p + \frac{3}{4} = 1 \Leftrightarrow p = \frac{1}{4}$$

(b) *Determine as funções de probabilidade marginais de X e Y.*

Da definição de função de probabilidade marginal,

$$p_X(x) = \sum_y p_{XY}(x,y)\ ,\forall x \in \mathbb{R}$$

Assim, para $x \in S_X$

$$p_X(-1) = \sum_{y=-1}^{1} p_{XY}(-1, y) = 0 + \frac{1}{4} + 0 = \frac{1}{4}$$

$$p_X(0) = \sum_{y=-1}^{1} p_{XY}(0, y) = \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}$$

$$p_X(1) = \sum_{y=-1}^{1} p_{XY}(1, y) = 0 + \frac{1}{4} + 0 = \frac{1}{4}$$

e

x	-1	0	1	c.c
$p_X(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0

De forma equivalente,

$$p_Y(y) = \sum_{x \in S_X} p_{XY}(x, y) \ , \forall y \in \mathbb{R}$$

Para $y \in S_Y$

$$p_Y(-1) = \sum_{x=-1}^{1} p_{XY}(x, -1) = 0 + \frac{1}{4} + 0 = \frac{1}{4}$$

$$p_Y(0) = \sum_{x=-1}^{1} p_{XY}(x, 0) = \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}$$

$$p_Y(1) = \sum_{x=-1}^{1} p_{XY}(x, 1) = 0 + \frac{1}{4} + 0 = \frac{1}{4}$$

e

y	-1	0	1	c.c
$p_Y(y)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0

(c) *Calcule $P(X = x/Y = 0)$.*

Da definição de função de probabilidade condicionada,

$$\forall x \in \mathbb{R},\ P(X = x/Y = 0) = \frac{P(X = x \cap Y = 0)}{P(Y = 0)} = \frac{p_{XY}(x, 0)}{p_Y(0)}$$

Então, para $x \in S_X$,

$$P(X = -1/Y = 0) = \frac{p_{XY}(-1, 0)}{p_Y(0)} = \frac{1/4}{1/2} = \frac{1}{2}$$

$$P(X=0/Y=0)=\frac{p_{XY}(0,0)}{p_Y(0)}=\frac{0}{1/2}=0$$
$$P(X=1/Y=0)=\frac{p_{XY}(1,0)}{p_Y(0)}=\frac{1/4}{1/2}=\frac{1}{2}$$

e tem-se

x	-1	1	c.c.
$P(X=x/Y=0)$	$\frac{1}{2}$	$\frac{1}{2}$	0

(d) *Mostre que $cov(X,Y)=0$ mas as variáveis aleatórias X e Y não são independentes.* Uma forma possível para calcular $cov(X,Y)$ é

$$cov(X,Y)=E(XY)-E(X)\cdot E(Y)$$

onde

$$E(XY)=\sum_{x\in S_X}\sum_{y\in S_Y} x\cdot y\cdot p_{XY}(x,y)$$

Assim,

$$\begin{aligned}\text{E(XY)} &= (-1)\cdot(-1)\cdot 0+(-1)\cdot 0\cdot\frac{1}{4}+(-1)\cdot 1\cdot 0\\ &\quad + 0\cdot(-1)\cdot\tfrac{1}{4}+0\cdot 0\cdot 0+0\cdot 1\cdot\tfrac{1}{4}\\ &\quad + 1\cdot(-1)\cdot 0+1\cdot 0\cdot\tfrac{1}{4}+1\cdot 1\cdot 0\\ &= 0\end{aligned}$$

$$\begin{aligned}\text{E(X)} = \text{E(Y)} &= (-1)\cdot\frac{1}{4}+0\cdot\frac{1}{2}+1\cdot\frac{1}{4}\\ &= 0\end{aligned}$$

e

$$cov(X,Y)=0-0\cdot 0=0$$

Por definição, as variáveis X e Y são independentes se e só se

$$\forall (x,y)\in \mathbb{R}^2,\ p_{XY}(x,y)=p_X(x)\cdot p_Y(y) \quad (1)$$

Verifica-se, por exemplo, que

$$p_{XY}(-1,-1)=0$$

e

$$p_X(-1)\cdot p_Y(-1)=\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$$

ou seja, há pelo menos um par de valores (x,y) para o qual a condição (1) não se verifica, isto é, as variáveis X e Y não são independentes.

22. *Seja* $f(x,y)=\frac{x+y}{32}$, $x=1,\ 2$ *e* $y=1,2,3,4$ *a função de probabilidade conjunta do par de variáveis aleatórias* (X,Y).

Repare que a função de probabilidade conjunta pode ser representada na forma

Y / X	1	2	3	4
1	$\frac{1+1}{32}$	$\frac{1+2}{32}$	$\frac{1+3}{32}$	$\frac{1+4}{32}$
2	$\frac{2+1}{32}$	$\frac{2+2}{32}$	$\frac{2+3}{32}$	$\frac{2+4}{32}$

isto é,

Y / X	1	2	3	4
1	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{4}{32}$	$\frac{5}{32}$
2	$\frac{3}{32}$	$\frac{4}{32}$	$\frac{5}{32}$	$\frac{6}{32}$

(a) *Deduza as funções de probabilidade marginais de* X *e* Y.

Sabe-se que

$$p_X(x)=\sum_y p_{XY}(x,y)\ ,\forall x\in\mathbb{R}$$

Assim, para $x\in S_X$

$$p_X(1)=\sum_{y=1}^{4}p_{XY}(1,y)=\frac{2}{32}+\frac{3}{32}+\frac{4}{32}+\frac{5}{32}=\frac{14}{32}$$

$$p_X(2)=\sum_{y=1}^{4}p_{XY}(2,y)=\frac{3}{32}+\frac{4}{32}+\frac{5}{32}+\frac{6}{32}=\frac{18}{32}$$

e

x	1	2	c.c
$p_X(x)$	$\frac{14}{32}$	$\frac{18}{32}$	0

De forma equivalente,

$$p_Y(y)=\sum_{x\in S_X}p_{XY}(x,y)\ ,\forall y\in\mathbb{R}$$

Para $y\in S_Y$

$$p_Y(1)=\sum_{x=1}^{2}p_{XY}(x,1)=\frac{2}{32}+\frac{3}{32}=\frac{5}{32}$$

$$p_Y(2)=\sum_{x=1}^{2}p_{XY}(x,2)=\frac{3}{32}+\frac{4}{32}=\frac{7}{32}$$

$$p_Y(3) = \sum_{x=1}^{2} p_{XY}(x,3) = \frac{4}{32} + \frac{5}{32} = \frac{9}{32}$$

$$p_Y(4) = \sum_{x=1}^{2} p_{XY}(x,4) = \frac{5}{32} + \frac{6}{32} = \frac{11}{32}$$

e

y	1	2	3	4	c.c
$p_Y(y)$	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{9}{32}$	$\frac{11}{32}$	0

(b) *Calcule* $P(X > Y), P(Y = 2X), P(X + Y = 3), P(X \leq 3 - Y),\ P(X \geq 1)$ *e* $P(0 \leq Y \leq 3)$.

Conhecida a função de probabilidade conjunta, podem ser identificados os pares de valores que satisfazem cada uma das condições:

$$\begin{aligned}
P(X > Y) &= P(X = 2 \cap Y = 1) \\
&= p_{XY}(2,1) \\
&= \frac{3}{32} \\
P(Y = 2X) &= P\left((X = 1 \cap Y = 2) \cup (X = 2 \cap Y = 4)\right) \\
&= p_{XY}(1,2) + p_{XY}(2,4) \\
&= \frac{3}{32} + \frac{6}{32} = \frac{9}{32} \\
P(X + Y = 3) &= P\left((X = 1 \cap Y = 2) \cup (X = 2 \cap Y = 1)\right) \\
&= p_{XY}(1,2) + p_{XY}(2,1) \\
&= \frac{3}{32} + \frac{3}{32} = \frac{6}{32} \\
P(X \leq 3 - Y) &= P\left((X = 1 \cap Y = 1) \cup (X = 1 \cap Y = 2) \cup (X = 2 \cap Y = 1)\right) \\
&= p_{XY}(1,1) + p_{XY}(1,2) + p_{XY}(2,1) \\
&= \frac{2}{32} + \frac{3}{32} + \frac{3}{32} = \frac{8}{32} \\
P(X \geq 1) &= P(\Omega) \\
&= 1 \\
P(0 \leq Y \leq 3) &= 1 - P(Y = 4) \\
&= 1 - \frac{11}{32} \\
&= \frac{21}{32}
\end{aligned}$$

(c) *Calcule* $P(Y = y/X = 2),\ \forall y$.

Da definição de função de probabilidade condicionada,

$$\forall y \in \mathbb{R},\ P(Y = y/X = 2) = \frac{P(X = 2 \cap Y = y)}{P(X = 2)} = \frac{p_{XY}(2,y)}{p_X(2)}$$

Então, para $y \in S_Y$,

$$P(Y = 1/X = 2) = \frac{p_{XY}(2,1)}{p_X(2)} = \frac{3/32}{18/32} = \frac{3}{18}$$

$$P(Y=2/X=2)=\frac{p_{XY}(2,2)}{p_X(2)}=\frac{4/32}{18/32}=\frac{4}{18}$$
$$P(Y=3/X=2)=\frac{p_{XY}(2,3)}{p_X(2)}=\frac{5/32}{18/32}=\frac{5}{18}$$
$$P(Y=4/X=2)=\frac{p_{XY}(2,4)}{p_X(2)}=\frac{6/32}{18/32}=\frac{6}{18}$$

e tem-se

y	1	2	3	4	c.c.
$P(Y=y/X=2)$	$\frac{3}{18}$	$\frac{4}{18}$	$\frac{5}{18}$	$\frac{6}{18}$	0

(d) *As variáveis aleatórias X e Y são independentes? Justifique.*

Por definição, as variáveis X e Y são independentes se e só se

$\forall (x,y)\in \mathbb{R}^2,\ p_{XY}(x,y)=p_X(x)\cdot p_Y(y)$ (1)

Verifica-se, por exemplo, que

$$p_{XY}(1,1)=\frac{2}{32}=0.0625$$

e

$$p_X(1)\cdot p_Y(1)=\frac{14}{32}\cdot\frac{5}{32}=\frac{70}{1024}\simeq 0.06836$$

ou seja, há pelo menos um par de valores (x,y) para o qual a condição (1) não se verifica, isto é, as variáveis X e Y não são independentes.

23. *Numa dada loja de componentes para computadores, as vendas diárias de discos rígidos das marcas 1 e 2, respetivamente X_1 e X_2, têm a seguinte função de probabilidade conjunta:*

X_2 \ X_1	0	1	2
0	0.12	0.25	0.13
1	0.05	0.30	0.01
2	0.03	0.10	0.01

(a) *Calcule as funções de probabilidade marginais de X_1 e X_2.*

Cálculos...

X_2 \ X_1	0	1	2	$p_{X_2}(y)$
0	0.12	0.25	0.13	0.5
1	0.05	0.30	0.01	0.36
2	0.03	0.10	0.01	0.14
$p_{X_1}(x)$	0.2	0.65	0.15	

Então,

x	0	1	2	c.c.
$p_{X_1}(x)$	0.2	0.65	0.15	0

y	0	1	2	c.c.
$p_{X_2}(y)$	0.5	0.36	0.14	0

(b) *Compare o número médio de vendas diárias de discos das duas marcas.*

Pretende-se comparar os valores dos parâmetros $E(X_1)$ e $E(X_2)$.
$E(X_1) = \sum_x x\, p_{X_1}(x) = 0.95$ e $E(X_2) = \sum_y y\, p_{X_2}(y) = 0.64$.
O número médio de vendas diárias de discos da marca 1 é *ligeiramente* superior ao número médio da marca 2.

(c) *Calcule a probabilidade de, num dia, a marca* 1 *ser a mais vendida.*

$P($"num dia, a marca 1 ser a mais vendida"$) =$

$$= P(X_1 > X_2) = \sum_{(x,y):x>y} p_{X_1 X_2}(x,y) =$$
$$= p_{X_1 X_2}(1,0) + p_{X_1 X_2}(2,0) + p_{X_1 X_2}(2,1) = 0.25 + 0.13 + 0.01 = 0.39$$

(d) *Calcule a função de probabilidade de X_2, nos dias em que não há vendas de discos da marca 1.*

Pretende-se a função de probabilidade condicionada de X_2 dado $X_1 = 0$, ou seja,

$$p_{X_2/X_1=0}(x) = P(X_2 = x/X_1 = 0) = \frac{P(X_1 = 0 \cap X_2 = x)}{P(X_1 = 0)} = \frac{p_{X_1 X_2}(0,x)}{p_{X_1}(0)}, \forall x \in \mathbb{R}.$$

Cálculos auxiliares:

$$P(X_2 = 0/X_1 = 0) = \frac{p_{X_1 X_2}(0,0)}{p_{X_1}(0)} = \frac{0.12}{0.2} = 0.6$$
$$P(X_2 = 1/X_1 = 0) = \frac{p_{X_1 X_2}(0,1)}{p_{X_1}(0)} = \frac{0.05}{0.2} = 0.25$$
$$P(X_2 = 2/X_1 = 0) = \frac{p_{X_1 X_2}(0,2)}{p_{X_1}(0)} = \frac{0.03}{0.2} = 0.15$$

Então,

x	0	1	2	c.c
$P(X_2 = x/X_1 = 0)$	0.6	0.25	0.15	0

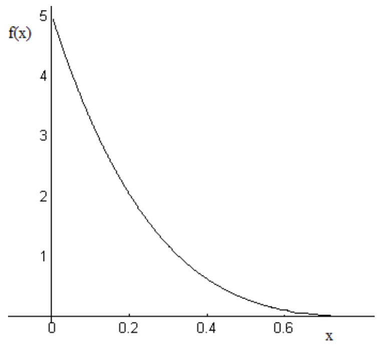
(e) *As vendas diárias de discos das duas marcas são independentes?*

Vimos que $P(X_2 = 0) = 0.5$ e $P(X_2 = 0/X_1 = 0) = 0.6$.
Assim, como $P(X_2 = 0) \neq P(X_2 = 0/X_1 = 0)$, pode concluir-se que X_1 e X_2 não são independentes.

3. Variáveis Aleatórias e Distribuições de Probabilidade Contínuas

1. *Uma estação de serviço enche os seus depósitos uma vez por semana. A quantidade de combustível procurada por semana é uma variável aleatória (v.a.) X com função densidade f, definida por*

$$f(x) = \begin{cases} 5(1-x)^4 & \text{se} \quad 0 < x \leq 1 \\ 0 & \text{se} \quad x \leq 0 \vee x > 1 \end{cases} .$$



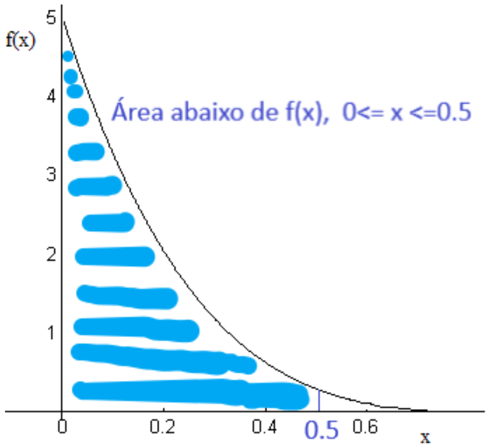
Esboço do gráfico de f

(a) *Qual a probabilidade de, numa semana qualquer, a quantidade de combustível procurada naquela estação de serviço não exceder* 0.5 *(*unidades de medida*)? Interprete geometricamente o resultado obtido.*

X = "Quantidade de combustível procurada por semana."

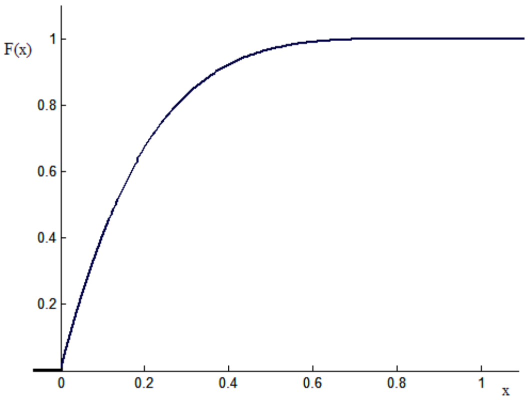
P("a quantidade de combustível procurada numa semana não exceder 0.5")

$$\begin{aligned} &= P(X \leq 0.5) \\ &= \int_{-\infty}^{0.5} f(x)\, dx \\ &= \int_0^{0.5} 5(1-x)^4\, dx \ , \quad \text{pois} \quad f(x) = 0, \forall x \notin]0,1[\\ &= (-1) \times 5 \int_0^{0.5} (-1)(1-x)^4\, dx \\ &= \left[-(1-x)^5\right]_0^{0.5} \\ &= -(1-0.5)^5 + 1 \\ &= 0.96875 \end{aligned}$$



(b) *A função distribuição de X é dada por*

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ 1-(1-x)^5 & , \quad 0 \leq x < 1 \\ 1 & , \quad x \geq 1 \end{cases} .$$



Esboço do gráfico de F

Usando F, determine:

i. $P(0.2 < X < 0.5)$;

$$\begin{aligned} P(0.2 < X < 0.5) &= P(X < 0.5) - P(X \leq 0.2) \\ & \quad X \text{ é contínua logo } P(X = a) = 0, \forall a \in \mathbb{R} \\ &= P(X \leq 0.5) - P(X \leq 0.2) \\ &= F(0.5) - F(0.2) \\ &= 1-(1-0.5)^5 - [1-(1-0.2)^5] \\ &= 0.29643 \end{aligned}$$

ii. *a capacidade dos depósitos por forma a que a probabilidade de se esvaziarem numa determinada semana seja de 5%.*

? Capacidade = c tal que

$P($"os depósitos se esvaziarem numa determinada semana"$) = 0.05$

$$\begin{array}{lll} & P(X \geq c) & = & 0.05 \\ \Leftrightarrow & 1 - P(X < c) & = & 0.05 \\ \Leftrightarrow & P(X \leq c) & = & 0.95 \\ \Leftrightarrow & 1 - (1-c)^5 & = & 0.95 \\ \Leftrightarrow & c & = & 1 - \sqrt[5]{0.05} \\ \Leftrightarrow & c & = & 0.4507 \quad \text{(Este valor nunca poderia ser superior a 0.5, porquê?)} \end{array}$$

2. *O tempo diário (em horas) de acesso à* internet *por uma determinada pessoa é representado por uma v.a. X com função densidade e função distribuição dadas, respetivamente, por*

$$f(x) = \begin{cases} \frac{1}{25}\,x & \text{se} \quad 0 \le x \le 5 \\ \frac{1}{25}\,(10-x) & \text{se} \quad 5 < x \le 10 \\ 0 & \text{se} \quad x < 0 \lor x > 10 \end{cases} \qquad \text{e} \qquad F(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{x^2}{50} & , \quad 0 \le x \le 5 \\ 1 - \frac{(10-x)^2}{50} & , \quad 5 \le x \le 10 \\ 1 & , \quad x > 10 \end{cases}.$$

Assuma que $E(X) = 5$ e $E(X^2) = \frac{175}{6}$.

(a) *Considere a v.a. $Y = 2X - 5$. Calcule $E(Y)$ e $V(Y)$.*

Por aplicação de propriedades do valor esperado e da variância, tem-se que

$$\begin{aligned} E(Y) &= E(2X-5) \\ &= 2E(X) - 5 \\ &= 2 \times 5 - 5 \\ &= 5 \end{aligned}$$

e

$$\begin{aligned} V(Y) &= V(2X-5) \\ &= 2^2\,V(X) + 0 \\ &= 4 \times \left(E(X^2) - E^2(X)\right) \\ &= 4 \times \left(\frac{175}{6} - 5^2\right) \\ &= \frac{50}{3} \end{aligned}$$

(b) *Considere os acontecimentos: $A = \{X \ge 5\}$, $B = \{X < 5\}$ e $C = \{2.5 \le X < 7.5\}$.*

i. *Calcule $P(A)$, $P(A/B)$ e $P(A/C)$.*

$$\begin{aligned} P(A) &= P(X \ge 5) \\ &= 1 - P(X < 5) \\ &= \underbrace{1 - F(5)}_{X \text{ é contínua, pelo que } P(X<5) = P(X \le 5)} \\ &= 1 - \frac{5^2}{50} \\ &= 0.5 \end{aligned}$$

$$\begin{aligned} P(A/B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(\emptyset)}{P(B)} \quad \text{(note que } B = \overline{A}\text{)} \\ &= 0 \end{aligned}$$

$$
\begin{aligned}
P(A/C) &= \frac{P(A\cap C)}{P(C)} \\
&= \frac{P(X\geq 5\cap 2.5\leq X<7.5)}{P(2.5\leq X<7.5)} \\
&= \frac{P(5\leq X<7.5)}{P(2.5\leq X<7.5)} \\
&= \frac{F(7.5)-F(5)}{F(7.5)-F(2.5)} \\
&= \frac{\big(1-(10-7.5)^2/50\big)-5^2/50}{(1-(10-7.5)^2/50)-2.5^2/50} \\
&= 0.5
\end{aligned}
$$

ii. *Verifique se A e B são independentes.*

Por definição, A e B são independentes **sse** $P(A\cap B)=P(A)\cdot P(B)$.

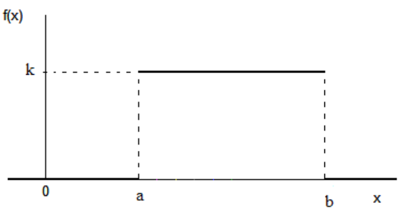
Da alínea anterior, $P(A)=0.5$ e $P(A\cap B)=0$.

Porque $B=\overline{A}$, $P(B)=1-P(A)=0.5$.

Assim,

$P(A)\cdot P(B)=0.5\times 0.5=0.25\neq P(A\cap B)$, pelo que os acontecimentos A e B não são independentes.

3. *Seja X uma v. a. com função densidade f, cujo esboço gráfico é apresentado na figura seguinte, onde k, a, e b são constantes reais. Note que $f(x)=0$, $\forall x\in \mathbb{R}\backslash[a,b]$.*



Determine:

(a) *o valor da constante k;*

Como f é uma função densidade,

i. $f(x) \geq 0, \forall x \in \mathbb{R} \;\Leftrightarrow k \geq 0$;

ii. $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1 \Leftrightarrow \int_a^b k\,dx = 1 \Leftrightarrow k[x]_a^b = 1 \Leftrightarrow k(b-a) = 1 \Leftrightarrow k = \frac{1}{b-a}$.

Concluímos que $f(x) = \begin{cases} \dfrac{1}{b-a} & \text{se} \quad x \in [a,b] \\ 0 & \text{se} \quad x < a \,\vee\, x > b \end{cases}$, ou seja que $X \sim \mathcal{U}_{[a,b]}$.

(b) *o valor esperado, variância e desvio padrão de X;*

Como $X \sim \mathcal{U}_{[a,b]}$,

$E(X) = \dfrac{a+b}{2}, \quad V(X) = \dfrac{(b-a)^2}{12} \quad$ e $\quad \sigma_X = \sqrt{\dfrac{(b-a)^2}{12}} = \dfrac{b-a}{\sqrt{12}}$ (ver tabelas).

(c) *a função distribuição de X.*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} \int_{-\infty}^{x} 0\,dt & \text{se} \quad x < a \\ \displaystyle\int_a^x \frac{1}{b-a}\,dt & \text{se} \quad x \in [a,b] \\ \displaystyle\int_a^b \frac{1}{b-a}\,dt + \int_b^x 0\,dt & \text{se} \quad x > b \end{cases}$$

$$= \begin{cases} 0 & \text{se} \quad x < a \\ \dfrac{x-a}{b-a} & \text{se} \quad x \in [a,b] \\ 1 & \text{se} \quad x > b \end{cases}.$$

4. *Alguns cabos utilizados nas instalações de telefones podem ser reaproveitados. Assume-se que o comprimento dos cabos segue uma lei* Uniforme *no intervalo de 1 a 15, $\mathcal{U}_{[1,15]}$, polegadas. Para um cabo escolhido ao acaso, calcule:*

(a) *o seu comprimento médio e mediana;*

Como X = *"comprimento, em polegadas, de um cabo"* é uma variável aleatória tal que $X \sim \mathcal{U}_{[1,15]}$, sabe-se que

$f(x) = \begin{cases} \dfrac{1}{14} & \text{se} \quad x \in [1,15] \\ 0 & \text{se} \quad x < 1 \,\vee\, x > 15 \end{cases}$ e $F(x) = \begin{cases} 0 & \text{se} \quad x < 1 \\ \dfrac{x-1}{14} & \text{se} \quad 1 \leq x \leq 15 \\ 1 & \text{se} \quad x > 15 \end{cases}$

Então, o comprimento médio de um cabo é

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\,dx = \frac{1+15}{2} = 8$$

e, representando o comprimento mediano por Md,

$$F(Md) = 0.5 \Leftrightarrow \frac{Md-1}{14} = 0.5 \Leftrightarrow Md - 1 = 7 \Leftrightarrow Md = 8$$

(b) *a sua variância e desvio padrão;*

$$V(X) = \frac{(15-1)^2}{12} = \frac{196}{12} \simeq 16.33 \quad \text{e } \sigma_X = \sqrt{\frac{196}{12}} = \frac{14}{\sqrt{12}} \simeq 4.04$$

(c) *a probabilidade de que o seu comprimento seja superior a 5 polegadas;*

$$\begin{aligned} P(X > 5) &= 1 - P(X \leq 5) \\ &= 1 - F(5) \\ &= 1 - \frac{5-1}{14} \\ &\simeq 0.7143 \end{aligned}$$

(d) *a probabilidade de que o seu comprimento se situe entre 0 e 8 polegadas.*

$$\begin{aligned} P(0 < X < 8) &= F(8) - F(0) \\ &= \frac{8-1}{14} - 0 \\ &= 0.5 \end{aligned}$$

5. *A duração de vida, em milhares de horas, de uma componente de certo tipo de aparelho de radar é uma v. a. X com distribuição* Exponencial *de parâmetro* 0.1*, isto é, com função densidade e distribuição:*

$$f(x) = \begin{cases} 0 & \text{se} \quad x \leq 0 \\ 0.1\; e^{-0.1x} & \text{se} \quad x > 0 \end{cases} \qquad F(x) = \begin{cases} 0 & \text{se} \quad x \leq 0 \\ 1 - \; e^{-0.1x} & \text{se} \quad x > 0 \end{cases}.$$

(a) *Determine a probabilidade de uma componente, escolhida ao acaso, durar menos de 4 mil horas.*

$P(X < 4) = P(X \leq 4) = F(4) = 1 - e^{-0.1\times 4} = 0.3297.$

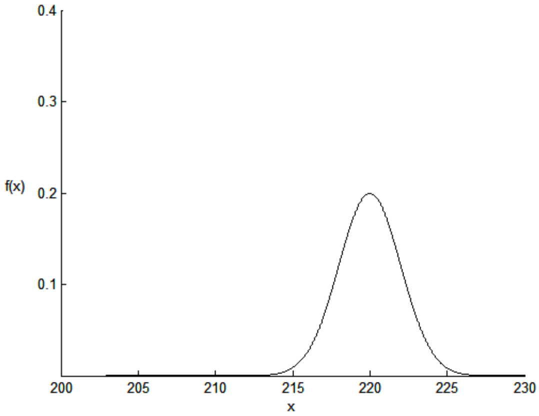
(b) *Indique a duração média de vida de uma componente e o respetivo desvio padrão.*

Se $X \sim \mathcal{E}(0.1)$ então, pelas tabelas sabemos que

$E(X) = \dfrac{1}{0.1} = 10;$

$V(X) = \dfrac{1}{0.1^2} = 100 \Rightarrow \sigma(X) = \sqrt{100} = 10.$

6. *A tensão de corrente X numa instalação elétrica tem distribuição* Normal *de média* $220\,V$ *e desvio padrão* $2\,V$*;* $X \sim \mathcal{N}(220, 2)$.



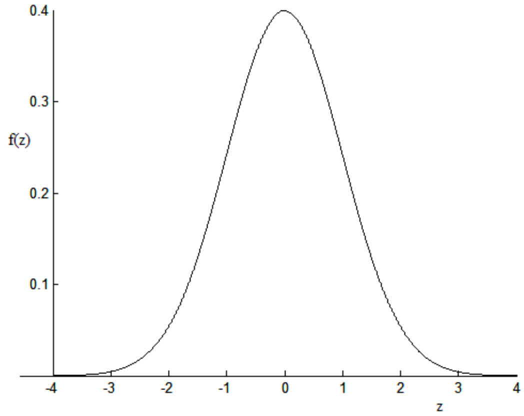
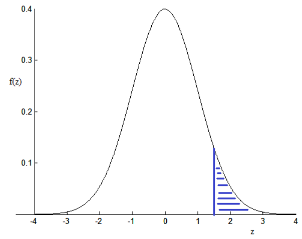


Gráfico da densidade da lei $\mathcal{N}(220, 2)$ *Gráfico da densidade da lei* $\mathcal{N}(0, 1)$

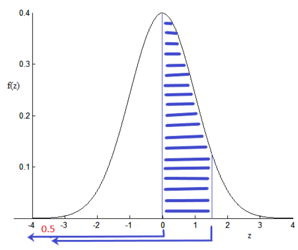
Recorrendo à lei Normal standard, *calcule:*

(a) $P(X > 223)$; (b) $P(220 < X < 223)$; (c) $P(X < 218)$; (d) $P(X \leq 223 / X > 221)$.

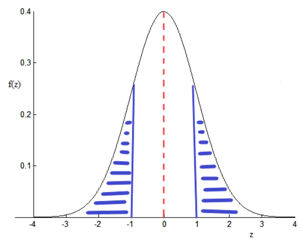
$$\text{(a) } P(X > 223) = P\Bigg(\underbrace{\frac{X-220}{2}}_{Z \sim N(0,1)} > \frac{223-220}{2}\Bigg)$$
$$= P(Z > 1.5)$$
$$= 1 - P(Z \leq 1.5)$$
$$= 1 - 0.9332$$
$$= 0.0668$$



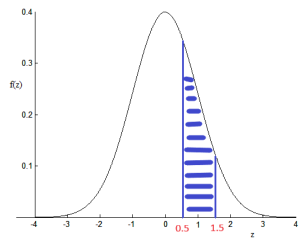
$$\text{(b) } P(220 < X < 223) = P\left(\frac{220-220}{2} < \frac{X-220}{2} < \frac{223-220}{2}\right)$$
$$= P(0 < Z < 1.5)$$
$$= P(Z \leq 1.5) - P(Z \leq 0)$$
$$= 0.9332 - 0.5$$
$$= 0.4332$$



$$\text{(c) } P(X < 218) = P\left(\frac{X-220}{2} < \frac{218-220}{2}\right)$$
$$= P(Z < -1)$$
simetria de Z relativamente a $z = 0$
$$= P(Z > 1)$$
$$= 1 - P(Z \leq 1)$$
$$= 1 - 0.8413 = 0.1587$$



$$\begin{aligned}
\text{(d) } P(X \leq 223/X > 221) &= \frac{P(X < 223 \bigcap X > 221)}{P(X > 221)} \\
&= \frac{P(Z < 1.5 \bigcap Z > 0.5)}{P(Z > 0.5)} \\
&= \frac{P(0.5 < Z < 1.5)}{1 - P(Z \leq 0.5)} \\
&= \frac{P(Z \leq 1.5) - P(Z \leq 0.5)}{1 - P(Z \leq 0.5)} \\
&= \frac{0.9332 - 0.6915}{1 - 0.6915} \\
&= \frac{0.2417}{0.3085} \\
&= 0.7835
\end{aligned}$$



7. *Uma empresa fabrica parafusos cujo comprimento é uma v. a. X com distribuição normal de média 0.25 cm e desvio padrão 0.02 cm. Considera-se defeituoso um parafuso cujo comprimento não pertença ao intervalo* $]0.2, 0.28[$. *Calcule a proporção de parafusos defeituosos.*

Sabe-se que $X =$ "comprimento de um parafuso, em *cm*", com $X \sim \mathcal{N}(0.25, 0.02)$.

Defina-se o acontecimento $D = \{\text{parafuso defeituoso}\}$. Pretende-se $P(D)$.

$P(D) = P(X \notin]0.2, 0.28[) = 1 - P(X \in]0.2, 0.28[)$ (uma possibilidade)

$$P(\overline{D}) = P(X \in]0.2, 0.28[) = P(0.2 < X < 0.28) = P\left(\frac{0.2 - 0.25}{0.02} < \frac{X - 0.25}{0.02} < \frac{0.28 - 0.25}{0.02}\right)$$

$= P(-2.5 < Z < 1.5), \quad \text{com } Z \sim \mathcal{N}(0, 1)$

$= P(Z < 1.5) - P(Z < -2.5) = P(Z < 1.5) - P(Z > 2.5)$ (uma possibilidade)

$= P(Z < 1.5) - (1 - P(Z \leq 2.5)) = 0.9332 - (1 - 0.9938) = 0.927$

Então, $P(D) = 0.073$ e pode concluir-se que 7.3% dos parafusos são defeituosos.

8. *O erro de medição do comprimento do raio de um circulo, em* mm, *é uma variável aleatória X com distribuição normal de média zero e desvio padrão σ.*

   (a) *Calcule σ de modo a que* 9.85% *das medições apresentem erros superiores a 6.45* mm.

   Sabe-se que $X =$ "erro de medição do comprimento do raio de um circulo, em *mm*", com $X \sim \mathcal{N}(0, \sigma)$.

Pretende-se σ : $P(X > 6.45) = 0.0985$.

$$P(X > 6.45) = 0.0985 \Leftrightarrow P\left(\frac{X}{\sigma} > \frac{6.45}{\sigma}\right) = 0.0985 \quad \text{(notando que } X \text{ já é centrada)}$$
$$\Leftrightarrow P\left(Z > \frac{6.45}{\sigma}\right) = 0.0985, \quad \text{com } Z \sim \mathcal{N}(0,1)$$
$$\Leftrightarrow P\left(Z \leq \frac{6.45}{\sigma}\right) = 1 - 0.0985$$
$$\Leftrightarrow F_Z\left(\frac{6.45}{\sigma}\right) = 0.9015$$
$$\Leftrightarrow \frac{6.45}{\sigma} = F_Z^{-1}(0.9015)$$
$$\Leftrightarrow \frac{6.45}{\sigma} = 1.29 \quad \text{(consulta da tabela ou da função inversa da normal } \textit{standard} \text{ nas máquinas)}$$
$$\Leftrightarrow \sigma = \frac{6.45}{1.29} = 5.0$$

(b) *Determine a percentagem de medições cujo erro varia entre* -1 *e* 1 mm.

Sabendo já que $X \sim \mathcal{N}(0,5)$ e $Z = \frac{X}{5} \sim \mathcal{N}(0,1)$,

$$\begin{aligned} P(-1 < X < 1) &= P(-\frac{1}{5} < Z < \frac{1}{5}) = P(-0.2 < Z < 0.2) \\ &= 2 \times P(0 < Z < 0.2) \quad \text{(uma possibilidade)} \\ &= 2 \times (\, P(Z < 0.2) - P(Z \leq 0)\,) = 2 \times (0.5793 - 0.5) = 0.1586 \end{aligned}$$

9. *Determinada empresa opera no mercado da União Europeia na área da distribuição de encomendas. A entrega de encomendas é executada em duas etapas. O tempo de entrega duma encomenda na primeira etapa tem distribuição normal de média 24h e desvio-padrão 4h, enquanto que o tempo de entrega na segunda etapa, que leva finalmente a encomenda ao destinatário, segue distribuição normal de média 8h e desvio-padrão 3h. Os tempos nas duas etapas são independentes.*

   (a) *Calcule a probabilidade do tempo de entrega duma encomenda na primeira etapa exceder 12h.*

   Seja $T_1 =$ "Tempo, em horas, de entrega duma encomenda na primeira etapa"

   $T_2 =$ "Tempo, em horas, de entrega duma encomenda na segunda etapa"

   Sabemos, do enunciado, que $T_1 \sim \mathcal{N}(24, 4)$; $T_2 \sim \mathcal{N}(8, 3)$ e que T_1 e T_2 são independentes.

$$\begin{aligned} P(T_1 > 12) &= P\left(\frac{T_1 - 24}{4} > \frac{12 - 24}{4}\right) \\ &= P(Z > -3), \text{ em que } Z = \frac{T_1 - 24}{4} \sim \mathcal{N}(0, 1) \\ &\quad \text{simetria da normal centrada e reduzida relativamente a 0} \\ &= P(Z < 3) \\ &= P(Z \leq 3), \ Z \text{ é contínua} \\ &= 0.99865 \end{aligned}$$

   (b) *Sabendo que na primeira etapa uma encomenda demorou mais de 24h, qual a probabilidade de ser entregue ao destinatário durante as próximas 8h?*

$$\begin{aligned} P(T_2 < 8/T_1 > 24) &= P(T_2 < 8), \ T_1 \text{ e } T_2 \text{ são independentes} \\ &= 0.5 \end{aligned}$$

   note que $E(T_2) = 8$ e T_2, tendo distribuição normal, é simétrica relativamente à sua média.

   (c) *Qual a probabilidade duma encomenda ser entregue ao destinatário num período superior a dois dias após o seu envio?*

   Para que a encomenda seja entregue ao destinatário num período superior a dois dias (48 horas) a soma dos tempos de entrega nas duas etapas tem que ser superior a 48 horas, isto é, é-nos pedida a $P(T_1 + T_2 > 48)$.

   Para determinar este valor de probabilidade é necessário identificarmos a distribuição da variável aleatória $T_1 + T_2$ que é uma combinação linear de variáveis aleatórias com distribuição normal independentes.

   Sabemos que $T_1 \sim \mathcal{N}(24, 4)$; $T_2 \sim \mathcal{N}(8, 3)$ e que T_1 e T_2 são independentes. Então, pela *estabilidade da lei normal* podemos afirmar que

   $T_1 + T_2 \sim \mathcal{N}\left(24 + 8, \sqrt{4^2 + 3^2}\right) = \mathcal{N}(32, 5)$

   e, deste modo,

$$
\begin{aligned}
P(T_1+T_2>48) &= P\left(\frac{T_1+T_2-32}{5} > \frac{48-32}{5}\right)\\
&= P(Z>3.2),\ \text{em que } Z\sim\mathcal{N}(0,1)\\
&= 1-P(Z\leq 3.2)\\
&= 1-0.999313 = 0.000687\approx 0.
\end{aligned}
$$

10. *O tempo de combustão de uma fita de magnésio de diâmetro* A *é normalmente distribuído de média* $\mu_A = 420\,seg$ *e desvio padrão* $\sigma_A = 80\,seg$*; para outra fita de diâmetro* B*, o tempo de combustão é também normalmente distribuído mas com* $\mu_B = 280\,seg$ *e* $\sigma_B = 45\,seg$*. Admitindo independência entre os tempos de combustão das fitas tipos* A *e* B*, calcule:*

    (a) *a probabilidade de o tempo de combustão de uma fita tipo* A *variar entre* 400 *e* 480 *seg;*

    Seja X_A = "tempo de combustão de uma fita de magnésio de diâmetro A, em *segundos (seg)*", com $X_A \sim \mathcal{N}(420, 80)$.

    $$
    \begin{aligned}
    P(400<X_A<480) &= P\left(\frac{400-420}{80} < \frac{X_A-420}{80} < \frac{480-420}{80}\right)\\
    &= P(-0.25<Z<0.75),\quad \text{com } Z\sim\mathcal{N}(0,1)\\
    &= P(Z<0.75)-P(Z<-0.25) = P(Z<0.75)-P(Z>0.25)\\
    &= P(Z<0.75)-(1-P(Z\leq 0.25))\\
    &= 0.7734-(1-0.5987) = 0.3721
    \end{aligned}
    $$

    (b) *a probabilidade de o tempo de combustão de uma fita de diâmetro* B *ser superior ao de uma fita de diâmetro* A*.*

    Defina-se também X_B = "tempo de combustão de uma fita de magnésio de diâmetro B, em *segundos*", com $X_B \sim \mathcal{N}(280, 45)$.
    Pretende-se $P(X_B > X_A)$.

    $P(X_B > X_A) = P(X_B - X_A > 0) = P(Y > 0)$, com $Y = X_B - X_A$.
    Qual é a distribuição/lei de probabilidade de Y?

    Uma vez que $Y = X_B - X_A = X_B + (-1)\times X_A$, isto é, Y é uma combinação linear de v.a.'s independentes com distribuições normais, com $X_A \sim \mathcal{N}(420, 80)$ e $X_B \sim \mathcal{N}(280, 45)$, da *Estabilidade da lei normal* concluímos que Y tem também distribuição normal, $Y \sim \mathcal{N}(E(Y), \sigma(Y))$
    .

    Em termos de parâmetros,
    $E(Y) = E(X_B - X_A) = E(X_B) - E(X_A) = 280 - 420 = -140$,
    $V(Y) = V(X_B - X_A) = V(X_B) + V(X_A) = 45^2 + 80^2 = 8425$.

    Assim, $Y \sim \mathcal{N}(-140, \sqrt{8425})$ e

$$P(Y>0) = P\left(\frac{Y-(-140)}{\sqrt{8425}} > \frac{0-(-140)}{\sqrt{8425}}\right) = P(Z>1.53), \text{ com } Z \sim \mathcal{N}(0,1)$$
$$= 1 - P(Z \le 1.53) = 1 - 0.9370 = 0.0630$$

11. *O peso de uma peça, produzida com determinado material, segue uma distribuição normal de média* $140\,g$ *e variância* $625\,g^2$.

   (a) *Determine a probabilidade de uma peça ter peso superior a* $120\,g$*, sabendo que o seu peso não excede* $150\,g$.

   Seja X = "Peso de uma peça, em gramas, produzida com determinado material."

   Do enunciado sabemos que,

   $X \sim \mathcal{N}(140, \sqrt{625}) = \mathcal{N}(140, 25)$

   então

$$\begin{aligned}
P(X>120/X\le 150) &= \frac{P(X>120 \cap X \le 150)}{P(X \le 150)} \\
&= \frac{P(120 < X \le 150)}{P(X \le 150)} \\
&= \frac{P(\frac{120-140}{25} < \frac{X-140}{25} \le \frac{150-140}{25})}{P(\frac{X-140}{25} \le \frac{150-140}{25})} \\
&= \frac{P(-0.8 < Z \le 0.4)}{P(Z \le 0.4)}, \text{ com } Z \sim \mathcal{N}(0,1) \\
&= \frac{P(Z \le 0.4) - P(Z \le -0.8)}{P(Z \le 0.4)} \\
&= \frac{0.6554 - P(Z \ge 0.8)}{0.6554} \\
&= \frac{0.6554 - [1 - P(Z < 0.8)]}{0.6554} \\
&= \frac{0.6554 - 1 + 0.7881}{0.6554} \\
&= \frac{0.4435}{0.6554} \\
&= 0.6767.
\end{aligned}$$

   (b) *As peças deste tipo são embaladas em caixas contendo 50 unidades. O peso de cada caixa também é aleatório, normalmente distribuído de média* $1\,kg$ *e desvio-padrão* $20\,g$*. Determine a probabilidade de o peso de uma caixa* completa *exceder* $8.5\,kg$.

1 caixa contém 50 unidades (peças);

Y = "Peso de uma caixa vazia"; $Y \sim \mathcal{N}(1000, 20)$

P("*peso de uma caixa completa* " $> 8500) = ?$

Seja W ="peso de uma caixa completa"$= \sum_{i=1}^{50} X_i + Y$ em que $X_i, i = 1, \ldots, 50$ são variáveis aleatórias identicamente distribuídas com X.

Os pesos das peças são independentes entre si e independentes do peso da caixa. Pela *estabilidade da lei normal*

$$\sum_{i=1}^{50} X_i \sim \mathcal{N}\left(\sum_{i=1}^{50} 140, \sqrt{\sum_{i=1}^{50} 625}\right) = \mathcal{N}\left(7000, \sqrt{31250}\right) = \mathcal{N}(7000, 176.78)$$

e

$$W = \sum_{i=1}^{50} X_i + Y \sim \mathcal{N}\left(7000 + 1000, \sqrt{31250 + 400}\right) = \mathcal{N}(8000, 177.90)$$

Assim,

$$\begin{aligned} P(\text{"}\textit{peso de uma caixa completa}\text{"} > 8500) &= P(W > 8500) \\ &= 1 - P\left(Z \leq \frac{8500 - 8000}{177.90}\right) \\ &= 1 - P(Z \leq 2.8106) \\ &= 1 - 0.9975 \\ &= 0.0025. \end{aligned}$$

(c) *Qual a probabilidade de, numa caixa completa, no máximo uma das peças ter peso superior a 150 g?*

$\mathcal{E}$ ="Pesar uma peça e registar caso o seu peso seja superior a 150g"

Repetir $\mathcal{E}$, 50 vezes, sempre nas mesmas condições.

V ="Número de peças, em 50 (uma caixa completa), que têm peso superior a 150g ".

Sendo as repetições de $\mathcal{E}$ efetuadas nas mesmas condições, $V \sim \mathcal{B}(50, P(X > 150)) = \mathcal{B}(50, 0.3446)$

então $P(V \leq 1) = 0.0.000000018252 \approx 0$

Uma resolução alternativa obtendo, contudo, apenas a probabilidade aproximada é

Como $n > 20$ e $p = 0.3446 \in]0.1, 0.9[$ podemos aproximar a distribuição binomial à distribuição normal tendo-se

$$V \sim \mathcal{B}(50, 0.3446) \dot{\sim} \mathcal{N}\left(50 \times 0.3446, \sqrt{50 \times 0.3446 \times (1 - 0.3446)}\right) = \mathcal{N}(17.23, 3.36)$$

Deste modo,

$$\begin{aligned} P(V \le 1) &= P\left(Z \le \frac{1 - 17.23}{3.36}\right) \\ &= P(Z \le -4.83) = P(Z \ge 4.83) = 1 - P(Z \le 4.83), \ Z \text{ é contínua} \\ &= 0.000000683575 \approx 0 \end{aligned}$$

12. *Suponha que o consumo de água num dado dia da semana, numa determinada localidade, segue uma distribuição normal de média 200* m^3 *e desvio padrão 10* m^3*. A capacidade do reservatório que abastece a localidade (e apenas esta) é de* 4240 m^3*. Sempre que o nível de água no reservatório cai 10% abaixo da sua capacidade é acionado um sistema de alarme.*

   (a) *Qual a probabilidade de o consumo de água, num dado dia da semana, estar compreendido entre 200 e 210* m^3*?*

   Seja X a variável aleatória que representa a "quantidade de água consumida por dia, em m^3". Tem-se que $X \sim \mathcal{N}(200, 10)$.

$$\begin{aligned} P(200 \le X \le 210) &= P(X \le 210) - P(X \le 200), \ Z \text{ é contínua} \\ &= P\left(\frac{X - 200}{10} \le \frac{210 - 200}{10}\right) - P\left(\frac{X - 200}{10} \le \frac{200 - 200}{10}\right) \\ &= P(Z \le 1) - P(Z \le 0) \\ &= 0.8413 - 0.5 \\ &= 0.3413 \end{aligned}$$

   (b) *Suponha que num dado dia a quantidade de água no reservatório se situa 5 % abaixo da sua capacidade; o abastecimento do referido reservatório processa-se, nesse dia, a uma taxa que segue uma distribuição normal de média 100* m^3 *e desvio padrão 30* m^3*. Supondo que consumo e abastecimento são independentes, qual a probabilidade de o alarme ser acionado?*

   De acordo com as condições propostas, o alarme é acionado sempre que a quantidade de água presente no reservatório for inferior a 90% da capacidade do reservatório, isto é, inferior a $3816m^3$.

   A "quantidade de água disponível no reservatório em qualquer instante, em m^3", no dia referido, é uma variável aleatória, por exemplo T, tal que

$$T = 95\% \times 4240 - X + A$$

em que A é a variável aleatória que representa a "taxa de abastecimento do reservatório, em m^3", com $A \sim \mathcal{N}(100, 30)$.

Da *Estabilidade da lei normal* concluímos que T tem também distribuição normal, $T \sim \mathcal{N}(E(T), \sigma(T))$.

$E(T) = E(4028 - X + A) = 4028 - E(X) + E(A) = 4028 - 200 + 100 = 3928,$

$V(T) = V(4028 - X + A) = V(X) + V(A) = 10^2 + 30^2 = 1000.$

$$\begin{aligned}
P(T < 3816) &= P\left(\frac{T - 3928}{\sqrt{1000}} < \frac{3816 - 3928}{\sqrt{1000}}\right) \\
&= P\left(Z < -\frac{112}{\sqrt{1000}}\right) \\
&= P(Z < -3.54) \\
&= 1 - P(Z \leq 3.54) \\
&= 1 - 0.9998 \\
&= 0.0002
\end{aligned}$$

13. *Um inspetor de controlo de qualidade rejeita qualquer lote de rolamentos esféricos se 3 ou mais defeituosos são encontrados num lote de 20 testados. Admita que a probabilidade de um rolamento ser defeituoso é 20%.*

    (a) *Determine a probabilidade de o lote ser rejeitado.*

    Pretende-se identificar o número de rolamentos defeituosos, em 20 rolamentos testados. Tem-se que

    $X =$"número de rolamentos defeituosos, em 20"é tal que $X \sim \mathcal{B}(20, 0.2)$

$$\begin{aligned}
P(X \geq 3) &= 1 - P(X \leq 2) \\
&= 1 - 0.2061 \\
&= 0.7939
\end{aligned}$$

    (b) *Qual o número esperado de rolamentos defeituosos num lote?*

    Sabe-se que, se $X \sim \mathcal{B}(n, p)$, então $E(X) = n \cdot p$.

$$\begin{aligned}
E(X) &= 20 \cdot 0.2 \\
&= 4
\end{aligned}$$

(c) *Admitindo que vai analisar um lote de 100 rolamentos, calcule um valor* aproximado *para a probabilidade de encontrar pelo menos 24 defeituosos.*

Em relação a (a), apenas é alterado o tamanho do lote de rolamentos a analisar, isto é, deve ser considerado

$X =$"número de rolamentos defeituosos, em 100"com $X \sim \mathcal{B}(100, 0.2)$

Sabe-se que a distribuição Binomial resulta de uma soma de variáveis aleatórias independentes com distribuição de Bernoulli. Porque, neste exemplo em particular, o número de parcelas nessa soma é suficientemente grande, o **Teorema Limite Central** é aplicável. Então,

$X \sim \mathcal{B}(100, 0.2) \dot{\sim} \mathcal{N}(100 \cdot 0.2, \sqrt{100 \cdot 0.2 \cdot (1 - 0.2)})$
isto é,
$X \dot{\sim} \mathcal{N}(20, 4)$

$$\begin{aligned} P(X \geq 24) &\simeq P\left(\frac{X-20}{4} \geq \frac{24-20}{4}\right) \\ &= 1 - P(Z \leq 1) \ , \quad Z \text{ é contínua} \\ &= 1 - 0.8413 \\ &= 0.1587 \end{aligned}$$

14. *O número de vírus detetados por mês por um departamento de informática segue uma lei de Poisson de média 5.*

(a) *Se num determinado mês se detetaram menos de 5 vírus, qual a probabilidade de terem sido detetados exactamente 4 vírus?*

Seja $X =$ "número de vírus detetados por mês por um departamento de informática". Segundo o enunciado, $X \sim \mathcal{P}(5)$.

$$P(X = 4 | X < 5) = \frac{P(X = 4 \cap X < 5)}{P(X < 5)} = \frac{P(X = 4)}{P(X \leq 4)} = \frac{0.1755}{0.4405} = 0.3984 \ .$$

(b) *Identifique a distribuição do número de vírus detetados durante um ano (12 meses consecutivos). Para esse período de tempo, calcule um valor aproximado para a probabilidade de se detetarem pelo menos 40 vírus.*

Definindo-se

$X_i =$ "número de vírus detetados no $\text{i}^{\underline{\text{ésimo}}}$mês por um departamento de informática", para $i = 1, 2, ..., 12$, seja

$X_T = \sum_{i=1}^{12} X_i =$ "número de vírus detetados durante um ano".

Assumindo que $X_1, X_2, ..., X_{12}$ são v.a.'s independentes, e como $X_i \sim \mathcal{P}(5), i = 1, 2, ..., 12$, da aditividade da *Poisson* pode concluir-se que $X_T \sim \mathcal{P}(60)$.

Pretende-se também uma aproximação para $P(X_T \geq 40)$.

Como $E(X_T) = 60 > 20$ (suficientemente grande), pode efetuar-se a aproximação $X_T \dot{\sim} \mathcal{N}(60, \sqrt{60})$ e,

$$\begin{aligned} P(X_T \geq 40) \simeq & P\left(\frac{X_T - 60}{\sqrt{60}} \geq \frac{40-60}{\sqrt{60}}\right) \\ = & P(Z \geq -2.58) \\ = & P(Z \leq 2.58) \\ = & 0.9951 \end{aligned}$$

4. Amostragem e Distribuições Amostrais

1. *A quantidade de chuva que cai por dia, expressa em litros por metro quadrado, pode ser descrita por uma v. a. X com distribuição contínua, admitindo a densidade*

$$f(x) = \begin{cases} \frac{21}{8192\times 10^7}\left(40x^5 - x^6\right) & \text{se} \quad 0 \leq x \leq 40 \\ 0 & \text{se} \quad \text{caso contrário} \end{cases} .$$

Admita que $E(X) = 30$ e $V(X) = 33.33$.

Sejam $X_1, X_2, \ldots, X_{100}$ uma amostra aleatória de X, com X_i a quantidade de chuva, em litros por metro quadrado, que cai no i-ésimo dia, $i = 1, ..., 100$.

(a) *Indique as propriedades de que as v. a.'s que constituem a amostra aleatória de X gozam.*

$X_1, X_2, \ldots, X_{100}$ são v.a.'s independentes e têm a mesma distribuição de probabilidade de X (i.i.d. com X)

(b) Considere a v.a. $\overline{X}_{100} = \dfrac{1}{100}\displaystyle\sum_{i=1}^{100} X_i$.

i. Calcule o valor médio e a variância de $\overline{X}_{100}$.

$$E(\overline{X}_{100}) = E\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{100}\sum_{i=1}^{100}\underbrace{E(X_i)}_{E(X)} = \frac{1}{100}\,100\,E(X) = E(X) = 30$$

e

$$V(\overline{X}_{100}) = V\left(\frac{1}{100}\sum_{i=1}^{100} X_i\right) = \frac{1}{100^2}\sum_{i=1}^{100}\underbrace{V(X_i)}_{V(X)} = \frac{1}{100^2}\,100\;V(X)$$

$$= \frac{V(X)}{100} = \frac{33.33}{100}$$

ii. Justifique que $\dfrac{\overline{X}_{100} - 30}{0.577} \stackrel{.}{\sim} \mathcal{N}(0,1)$.

Como a dimensão da amostra aleatória é suficientemente grande, $n = 100 > 30$, do teorema do limite central (T.L.C.) podemos, de facto, avançar que

$$\frac{\overline{X}_{100} - E(\overline{X}_{100})}{\sigma(\overline{X}_{100})} = \frac{\overline{X}_{100} - 30}{\sqrt{\frac{33.33}{100}}} = \frac{\overline{X}_{100} - 30}{0.577} \stackrel{.}{\sim} \mathcal{N}(0,1).$$

iii. Calcule uma aproximação para o valor de $P\left(28.5 < \overline{X}_{100} \leq 31.5\right)$. Interprete o resultado.

$$P\left(28.5 < \overline{X}_{100} \leq 31.5\right) = P\left(\frac{28.5-30}{0.577} < \frac{\overline{X}_{100} - 30}{0.577} \leq \frac{31.5-30}{0.577}\right)$$
$$= P(-2.6 < Z \leq 2.6), \quad Z \stackrel{.}{\sim} \mathcal{N}(0,1)$$
$$= 2 \times P(0 < Z < 2.6) \simeq 2 \times 0.4953 = 0.9906$$

A probabilidade da quantidade média de chuva a cair durante 100 dias variar entre 28.5 e 31.5 é de aproximadamente 99.06%.

2. *A energia em* Joules (J) *de qualquer partícula de um sistema é uma variável aleatória com distribuição* Exponencial *de parâmetro* 2. *A energia do sistema é a soma da energia das suas partículas que são independentes. Admita que determinado sistema* β *contém 1600 partículas.*

(a) *Indique, justificando, a lei aproximada da energia do sistema* β.

Seja $X =$"Energia, em *Joules*, de uma partícula de um sistema"

$X \sim Exp(2)$; O sistema β tem 1600 partículas;

Seja $Y =$"Energia do sistema β"$= \sum_{i=1}^{1600} X_i =$, onde $X_i =$"Energia, em *Joules*, da i$^{\underline{ésima}}$ partícula do sistema $\beta, i = 1, \ldots, 1600$".

Sendo $X_i, i = 1, \ldots, 1600$, variáveis aleatórias independentes e identicamente distribuídas com $X \sim Exp(2)$ e o número de partículas do sistema β suficientemente grande, $n = 1600 > 30$, pelo teorema do Limite Central (T.L.C.)

$$\frac{\overline{X}_{1600} - E\left(\overline{X}_{1600}\right)}{\sigma\left(\overline{X}_{1600}\right)} \ \dot{\sim}\ \mathcal{N}(0,1)$$

Note-se que

$$E\left(\overline{X}_{1600}\right) = E\left(\frac{1}{1600}\sum_{i=1}^{1600} X_i\right) = \frac{1}{1600}\sum_{i=1}^{1600} E(X_i)$$

Como X_i são i.i.d com $X \sim Exp(2)$ sabemos que $E(X_i) = \dfrac{1}{2}$ logo,

$$E\left(\overline{X}_{1600}\right) = \frac{1}{1600}\sum_{i=1}^{1600}\frac{1}{2} = \frac{1}{1600} \times 1600 \times \frac{1}{2} = \frac{1}{2}.$$

e

$$V\left(\overline{X}_{1600}\right) = V\left(\frac{1}{1600}\sum_{i=1}^{1600} X_i\right) = \left(\frac{1}{1600}\right)^2 \sum_{i=1}^{1600} V(X_i)$$

uma vez que $X_i, i = 1, \ldots, 1600$ são independentes. Como X_i são identicamente distribuídas com $X \sim Exp(2)$, sabemos que $V(X_i) = \dfrac{1}{2^2} = \dfrac{1}{4}$ logo,

$$V\left(\overline{X}_{1600}\right) = \left(\frac{1}{1600}\right)^2 \sum_{i=1}^{1600}\frac{1}{4} = \left(\frac{1}{1600}\right)^2 \times 1600 \times \frac{1}{4} = \frac{1}{4 \times 1600}.$$

Assim,

$$\frac{\overline{X}_{1600} - E\left(\overline{X}_{1600}\right)}{\sigma\left(\overline{X}_{1600}\right)} \ \dot{\sim}\ \mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{\frac{1}{1600}\sum_{i=1}^{1600} X_i - \frac{1}{2}}{\sqrt{\frac{1}{4\times 1600}}} \ \dot{\sim}\ \mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{1600 \times \left(\frac{1}{1600}\sum_{i=1}^{1600} X_i - \frac{1}{2}\right)}{1600 \times \sqrt{\frac{1}{4\times 1600}}} \;\dot{\sim}\; \mathcal{N}(0,1)$$

$$\Leftrightarrow \frac{\sum_{i=1}^{1600} X_i - 800}{20} \;\dot{\sim}\; \mathcal{N}(0,1)$$

$$\sum_{i=1}^{1600} X_i \;\dot{\sim}\; \mathcal{N}(800,20).$$

(b) *Calcule a probabilidade (aproximada) da energia do sistema β variar entre 780 e 840 J.*

$$\begin{aligned} P\left(780 < \sum_{i=1}^{1600} X_i < 840\right) &= P\left(\frac{780-800}{20} < \frac{\sum_{i=1}^{1600} X_i - 800}{20} < \frac{840-800}{20}\right) \\ &= P(-1 < Z < 2), \quad Z \dot{\sim}\; \mathcal{N}(0,1) \\ &= P(Z<2) - P(Z<-1) = P(Z<2) - P(Z>1) \\ &= P(Z<2) - 1 + P(Z \le 1) \simeq 0.9772 - 1 + 0.8413 \\ &= 0.8185 \end{aligned}$$

3. *O erro de medição do comprimento do raio de um circulo, em* mm*, é uma variável aleatória X com distribuição normal de média zero e desvio padrão 5. Considere uma amostra aleatória $X_1, X_2, ..., X_{10}$, daquela população, e as seguintes estatísticas:*

$$T_1 = \frac{1}{10}\sum_{i=1}^{10} X_i \qquad T_2 = \frac{5X_1 + 5X_{10}}{10}$$

(a) *Indique (justificando) as distribuições amostrais de T_1 e de T_2.*

Como $X_1, \ldots, X_{10}$ é amostra aleatória de X, podemos afirmar que $X_1, \ldots, X_{10}$ são i.i.d. com X, logo pela estabilidade da lei normal

$$\begin{aligned} T_1 = \frac{1}{10}\sum_{i=1}^{10} X_i &\sim \mathcal{N}\left(\frac{1}{10}\sum_{i=1}^{10} E(X_i), \sqrt{\left(\frac{1}{10}\right)^2 \sum_{i=1}^{10} V(X_i)}\right) \\ &= \mathcal{N}\left(\tfrac{1}{10} \times 10 \times 0, \sqrt{\left(\tfrac{1}{10}\right)^2 \times 10 \times 5^2}\right) \\ &= \mathcal{N}\left(0, \sqrt{2.5}\right) \end{aligned}$$

e

$$\begin{aligned} T_2 = \frac{5X_1 + 5X_{10}}{10} = \frac{1}{2}(X_1 + X_{10}) &\sim \mathcal{N}\left(\tfrac{1}{2}\left(E(X_1) + E(X_{10})\right), \sqrt{\left(\tfrac{1}{2}\right)^2 \left(V(X_1) + V(X_{10})\right)}\right) \\ &= \mathcal{N}\left(\tfrac{1}{2}(0+0), \sqrt{\tfrac{1}{4}(5^2+5^2)}\right) \\ &= \mathcal{N}\left(0, \sqrt{12.5}\right). \end{aligned}$$

(b) *Calcule e interprete* $P(T_1 > 3)$.

$$\begin{aligned} P(T_1 > 3) &= P\left(\frac{T_1 - 0}{\sqrt{2.5}} > \frac{3-0}{\sqrt{2.5}}\right) \\ &= P(Z > 1.8974),\ \text{com } Z \ \dot\sim\ \mathcal{N}(0,1) \\ &= 1 - P(Z \leq 1.8974) = 1 - 0.9711 = 0.0289. \end{aligned}$$

Numa amostra de dimensão 10, a probabilidade do erro médio de medição do comprimento do raio de um círculo ser superior a 3 *mm* é 0.0289.

4. *Seja* X *uma medida aleatória de valor esperado* $\frac{2}{3}$ *e variância* $\frac{8}{9}$. *Considere uma amostra aleatória de* X, $X_1, X_2, \ldots, X_n$, *e a média dessa amostra* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) *Para* $n \in \mathbb{N}$, *suficientemente grande, qual a lei de* $\overline{X}_n$? *Justifique.*

Uma vez que temos uma amostra aleatória, de uma dada população X, de dimensão n suficientemente grande, o T.L.C. permite concluir que a média amostral $\overline{X}_n$ tem lei aproximadamente normal,

$$\overline{X}_n \ \dot\sim\ \mathcal{N}(E(\overline{X}_n), \sigma(\overline{X}_n)) \Rightarrow Z = \frac{\overline{X}_n - E(\overline{X}_n)}{\sigma(\overline{X}_n)} = \sqrt{n}\ \frac{\overline{X}_n - 2/3}{\sqrt{8/9}} \ \dot\sim\ \mathcal{N}(0,1).$$

(b) *Qual deverá ser a dimensão mínima da amostra para que* $P(\overline{X}_n > 0.8) \leq 0.0119$?

Pretende-se o mínimo $n:\ P(\overline{X}_n > 0.8) \leq 0.0119$.

$$P(\overline{X}_n > 0.8) \leq 0.0119 \quad \Leftrightarrow \quad P\left(\sqrt{n}\ \frac{\overline{X}_n - 2/3}{\sqrt{8/9}} > \sqrt{n}\ \frac{0.8 - 2/3}{\sqrt{8/9}}\right) \leq 0.0119$$

$$\Leftrightarrow \quad P(Z > \sqrt{n}\ 0.1414) \leq 0.0119, \quad Z \dot{\sim} \mathcal{N}(0,1)$$

$$\Leftrightarrow \quad 1 - P(Z \leq \sqrt{n}\ 0.1414) \leq 0.0119$$

$$\Leftrightarrow \quad P(Z \leq \sqrt{n}\ 0.1414) \geq 1 - 0.0119$$

$$\Leftrightarrow \quad P(Z \leq \sqrt{n}\ 0.1414) \geq 0.9881$$

$$\Leftrightarrow \quad F_Z\left(\sqrt{n}\ 0.1414\right) \geq 0.9881$$

$$\Leftrightarrow \quad \sqrt{n}\ 0.1414 \geq \underbrace{F_Z^{-1}(0.9881)}_{2.26}, \text{ consultar tabela ou máquinas}$$

$$\Leftrightarrow \quad \sqrt{n}\ 0.1414 \geq 2.26$$

$$\Leftrightarrow \quad \sqrt{n}\ \geq \frac{2.26}{0.1414} \simeq 15.98$$

$$\Rightarrow \quad n\ \geq 15.98^2 \simeq 255.46$$

Assim, como $n \in \mathbb{N}, n\ \geq 256$ (e o mínimo $n = 256$).

5. *As normas ambientais em vigor exigem que a concentração diária de certo poluente não exceda* 120 ng/m^3 *(*nanogramas por metro cúbico*). Admita que essa concentração segue uma lei normal de valor esperado* 100 *e desvio padrão 9.71* ng/m^3*, e que as concentrações em dias distintos são independentes.*

   (a) *Mostre que em* 1.97% *dos dias as normas ambientais não são cumpridas.*

   Representando por X a variável aleatória que mede, em nanogramas por metro cúbico, a concentração diária desse poluente, tem-se que $X \sim \mathcal{N}(100, 9.71)$. Pretende-se identificar a probabilidade de, num qualquer dia, as normas ambientais não estarem a ser cumpridas, isto é, $P(X > 120)$.

$$
\begin{aligned}
P(X > 120) =& 1 - P(X \leq 120) \\
=& 1 - P\left(Z \leq \frac{120 - 100}{9.71}\right) \\
=& 1 - P(Z \leq 2.06) \\
=& 1 - 0.9803 \\
=& 0.0197
\end{aligned}
$$

(b) *Qual a probabilidade de a concentração média em* 15 *dias, escolhidos aleatoriamente, exceder* 120 ng/m^3*?*

Sendo observada uma amostra aleatória de X, de dimensão $n = 15$, a propriedade de estabilidade da lei normal permite concluir que a média amostral $\overline{X}_{15}$ também tem distribuição normal,

$$
\overline{X}_{15} \sim \mathcal{N}(E(\overline{X}_{15}), \sigma(\overline{X}_{15})) \Rightarrow Z = \frac{\overline{X}_{15} - E(\overline{X}_{15})}{\sigma(\overline{X}_{15})} = \sqrt{15}\ \frac{\overline{X}_{15} - 100}{9.71} \sim \mathcal{N}(0, 1).
$$

$$
\begin{aligned}
P(\overline{X}_{15} > 120) =& 1 - P(\overline{X}_{15} \leq 120) \\
=& 1 - P\left(Z \leq \sqrt{15}\ \frac{120 - 100}{9.71}\right) \\
=& 1 - P(Z \leq 7.98) \\
\simeq& 1 - 1 \\
=& 0
\end{aligned}
$$

5. Estimação

1. Uma fábrica produz cabos cujo diâmetro X (em milímetros) segue uma lei *Uniforme* no intervalo de 5 a $5+\theta$, $X \sim \mathcal{U}_{[5,\,5+\theta]}$, onde θ é um parâmetro real desconhecido.

Considere uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de X $(n \in \mathbb{N})$.

(a) Considere o estimador de θ, dado por

$$\hat{\Theta}_n = 2\left(\frac{1}{n}\sum_{i=1}^{n} X_i - 5\right).$$

Prove que $\hat{\Theta}_n$ é um estimador cêntrico de θ.

(b) Selecionaram-se aleatoriamente **20** cabos da produção da fábrica e registaram-se os respetivos diâmetros. Estes foram posteriormente classificados como indicado no quadro seguinte:

classes	$]5, 5.2]$	$]5.2, 5.4]$	$]5.4, 5.6]$	$]5.6, 5.8]$	$]5.8, 6]$
efectivos	4	3	5	4	4

i. Determine a média e a variância desta amostra.

ii. Indique uma estimativa cêntrica para θ, com base nesta amostra.

2. Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de dimensão n, $n \in \mathbb{N}$, de uma população X cuja lei de probabilidade é caracterizada pela seguinte função densidade

$$f_\theta(x) = \begin{cases} \dfrac{2}{\theta}\left(1 - \dfrac{x}{\theta}\right) & \text{se} \quad 0 < x < \theta \\ 0 & \text{se} \quad \text{caso contrário} \end{cases} \quad \text{, onde} \quad \theta \in \mathbb{R}^+.$$

Assuma que $E(X) = \frac{\theta}{3}$.

(a) Mostre que $\hat{\Theta}_1 = \sum_{i=1}^{n} X_i$ é um estimador enviesado de θ.

(b) Construa, a partir de $\hat{\Theta}_1$, um outro estimador de θ, $\hat{\Theta}_2$, que seja cêntrico.

(c) Suponha que se recolheu, ao acaso, uma amostra de X, de dimensão 100, $(x_1, x_2, \ldots, x_{100})$, para a qual se constatou que $\overline{x} = 20.2$. Indique uma estimativa cêntrica de θ. Sugira uma estimativa para $E(X)$.

3. *Uma máquina de parafusos está regulada para produzir em série peças de diâmetro médio 150 mm. Admite-se que os diâmetros são normalmente distribuídos. Uma amostra aleatória de 20 parafusos, extraída da população, forneceu os seguintes valores:*

$$\sum_{i=1}^{20} x_i = 2900; \qquad \sum_{i=1}^{20} x_i^2 = 432500.$$

Em face destes valores e utilizando um grau de confiança de 0.95, verifique se será de admitir irregularidade de produção na máquina supondo:

(a) *σ conhecido e igual a 25 mm;*

Seja $X =$"diâmetro de um parafuso"; $X \sim \mathcal{N}(\mu, \sigma)$

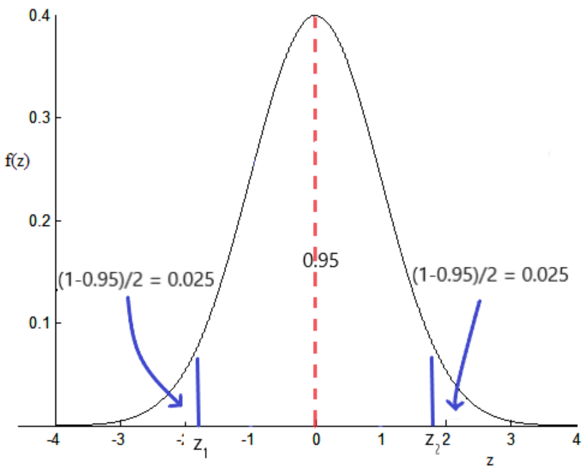
Dimensão da amostra $= 20$; $\sum_{i=1}^{20} x_i = 2900$; $\sum_{i=1}^{20} x_i^2 = 432500$

Pede-se um intervalo de confiança para μ ao grau $= 1-\alpha = 0.95$ com $\sigma = 25\ mm$ (conhecido).

- Passo 1: Escolha da variável fulcral - consultar as tabelas

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

- Passo 2: Determinar $z_1, z_2 \in \mathbb{R} : P(z_1 < Z < z_2) = 1 - \alpha \Leftrightarrow P(z_1 < Z < z_2) = 0.95$
  Como $Z \sim \mathcal{N}(0,1)$ que é simétrica relativamente a zero,



podemos considerar $z_2 = z$ e assim $z_1 = -z$ logo, determinar os valores de z_1 e z_2 nas condições acima referidas é equivalente a determinar

$$z \in \mathbb{R} : P(-z < Z < z) = 0.95.$$
$$\Leftrightarrow P(Z < z) = 0.95 + 0.025 = 0.975$$
$$\Leftrightarrow z = F_{\mathcal{N}}^{-1}(0.975) = 1.96$$

- Passo 3: Procurar L_1 e L_2 funções da amostra tais que $P(L_1 < \mu < L_2) = 0.95$ Do passo 2,

$$-1.96 < Z < 1.96$$

$$-1.96 < \sqrt{n}\frac{\overline{X} - \mu}{\sigma} < 1.96$$

$$-\frac{1.96\sigma}{\sqrt{n}} < \overline{X} - \mu < \frac{1.96\sigma}{\sqrt{n}}$$

$$-\overline{X} - \frac{1.96\sigma}{\sqrt{n}} < -\mu < -\overline{X} + \frac{1.96\sigma}{\sqrt{n}}$$

$$\overline{X} - \frac{1.96\sigma}{\sqrt{n}} < \mu < \overline{X} + \frac{1.96\sigma}{\sqrt{n}}$$

Assim, com 95% de confiança, o intervalo aleatório para μ é

$$\left]\overline{X} - \frac{1.96\sigma}{\sqrt{n}}, \overline{X} + \frac{1.96\sigma}{\sqrt{n}}\right[$$

- Passo 4: Concretização para uma amostra particular
  Para a amostra observada sabemos que
  $n = 20; \quad \overline{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{2900}{20} = 145$
  logo, um intervalo de confiança para μ ao grau 0.95 será

$$\left]145 - \frac{1.96 \times 25}{\sqrt{20}}, 145 + \frac{1.96 \times 25}{\sqrt{20}}\right[=]134.0433, 155.9567[\,.$$

Como 150 (medida segundo a qual o diâmetro médio do parafuso é considerado regular) pertence ao intervalo de confiança a 95%, podemos afirmar, com essa confiança, que não é de admitir irregularidade de produção na máquina.

(b) σ *desconhecido.*

Seja $X =$"diâmetro de um parafuso"; $X \sim \mathcal{N}(\mu, \sigma)$

Dimensão da amostra $= 20; \quad \sum_{i=1}^{20} x_i = 2900; \quad \sum_{i=1}^{20} x_i^2 = 432500$

Pede-se um intervalo de confiança para μ ao grau $= 1 - \alpha = 0.95$ com σ desconhecido.

- Passo 1: Escolha da variável fulcral - consultar as tabelas

$$Z = \sqrt{n}\frac{\overline{X} - \mu}{S} \sim t_{n-1} = t_{19}$$

- Passo 2: Determinar $z_1, z_2 \in \mathbb{R} : P(z_1 < Z < z_2) = 1 - \alpha \Leftrightarrow P(z_1 < Z < z_2) = 0.95$
  Como $Z \sim t_{19}$ que é simétrica relativamente a zero, podemos, tal como na alínea a), considerar $z_2 = z$ e assim $z_1 = -z$ logo, determinar os valores de z_1 e z_2 nas condições acima referidas é equivalente a determinar

$$\begin{aligned} &z \in \mathbb{R} : P(-z < Z < z) = 0.95. \\ &\Leftrightarrow P(Z < z) = 0.95 + 0.025 = 0.975 \\ &\Leftrightarrow z = F_t^{-1}(0.975) = 2.093 \end{aligned}$$

- Passo 3: Procurar L_1 e L_2 funções da amostra tais que $P(L_1 < \mu < L_2) = 0.95$ Do passo

2,

$$
\begin{aligned}
-2.093 &< Z < 2.093 \\
-2.093 &< \sqrt{n}\frac{\overline{X}-\mu}{S} < 2.093 \\
-\frac{2.093\ S}{\sqrt{n}} &< \overline{X}-\mu < \frac{2.093\ S}{\sqrt{n}} \\
-\overline{X}-\frac{2.093\sigma}{\sqrt{n}} &< -\mu < -\overline{X}+\frac{2.093\ S}{\sqrt{n}} \\
\overline{X}-\frac{2.093\ S}{\sqrt{n}} &< \mu < \overline{X}+\frac{2.093\sigma}{\sqrt{n}}
\end{aligned}
$$

Assim, com 95% de confiança, o intervalo aleatório para μ é

$$\left]\overline{X}-\frac{2.093\ S}{\sqrt{n}}, \overline{X}+\frac{2.093\ S}{\sqrt{n}}\right[$$

- Passo 4: Concretização para uma amostra particular

  Para a amostra observada sabemos que

  $n = 20;\quad \overline{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{2900}{20} = 145;$

  e que $s^2 = \frac{1}{19}\sum_{i=1}^{20} x_i^2 - \left(\frac{20}{19}\right)\overline{x}^2 = \frac{1}{19}432500 - \frac{20}{19}145^2 = 631.5789$

  logo, $s = \sqrt{631.5789} = 25.1312$

  e, um intervalo de confiança para μ ao grau 0.95 será

$$\left]145-\frac{2.093\times 25.1312}{\sqrt{20}}, 145+\frac{2.093\times 25.1312}{\sqrt{20}}\right[=]133.2384, 156.7616[\,.$$

Como 150 (medida segundo a qual o diâmetro médio do parafuso é considerado regular) pertence ao intervalo de confiança a 95%, podemos afirmar, com essa confiança, que não é de admitir irregularidade de produção na máquina.

4. *Para estudar a tensão de rutura de certo tipo de algodão fizeram-se 10 observações com os seguintes resultados, em Kg:*

   7.4; 7.8; 7.1; 6.9; 7.3; 7.6; 7.3; 7.4; 7.7; 7.3

   Admitindo que a tensão de rutura segue uma distribuição normal com variância 0.08, determine:

   (a) *uma estimativa para a tensão de rutura média;*

   Designando por X a variável aleatória que representa a "tensão de rutura", tem-se $X \sim \mathcal{N}(\mu, \sqrt{0.08})$.

Para estimar o valor esperado da população, μ, um "bom"estimador é a média de uma amostra aleatória recolhida dessa população.

Em relação à amostra observada, verifica-se que $\sum_{i=1}^{10} x_i = 73.8$

$$\begin{aligned}\widehat{\mu} =& \overline{x} \\ =& \frac{1}{10}\sum_{i=1}^{10} x_i \\ =& \frac{1}{10}\cdot 73.8 \\ =& 7.38\end{aligned}$$

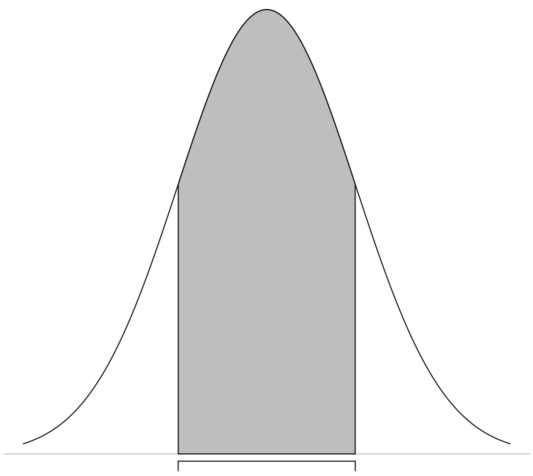
(b) *um intervalo com 95% de confiança para a tensão de rutura média;*

i) Parâmetro a estimar: μ

ii) Variável fulcral e respetiva distribuição: $Z = \sqrt{n}\,\frac{\overline{X}-\mu}{\sigma} \sim N(0,1)$

ivii) $P(-z < Z < +z) = 1-\alpha \Leftrightarrow \cdots \Leftrightarrow P\left(\overline{X} - z\,\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z\,\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$

iv) Grau de confiança: $1-\alpha = 0.95$ o que traduz que $z = F^{-1}(0.975) = 1.96$



Tem-se, então

$$I.C._{\mu} = \left]7.38 - 1.96\,\frac{\sqrt{0.08}}{\sqrt{10}}, 7.38 + 1.96\,\frac{\sqrt{0.08}}{\sqrt{10}}\right[=]7.205, 7.555[$$

(c) *se pretender que o erro dessa estimativa não ultrapasse 0.07, em* 95 % *dos casos, quantos elementos deveria incluir na amostra?*

O erro da aproximação de μ pelo Intervalo Aleatório de Confiança a $1-\alpha = 0.95$, $\left]\overline{X} - z\,\frac{\sigma}{\sqrt{n}}, \overline{X} + z\,\frac{\sigma}{\sqrt{n}}\right[$, é dado por

$$\varepsilon = z\,\frac{\sigma}{\sqrt{n}}$$

$$\begin{aligned}
\varepsilon \leq 0.07 &\Leftrightarrow z\,\frac{\sigma}{\sqrt{n}} \leq 0.07 \\
&\Leftrightarrow \sqrt{n} \geq z\,\frac{\sigma}{0.07} \\
&\Rightarrow n \geq \left(z\,\frac{\sigma}{0.07}\right)^2 \\
&\Leftrightarrow n \geq \left(1.96\,\frac{\sqrt{0.08}}{0.07}\right)^2 \\
&\Leftrightarrow n \geq 62.72
\end{aligned}$$

de onde se conclui que o menor valor é $n = 63$.

5. Sabe-se que as classificações X de determinado curso são normalmente distribuídas. Foi recolhida uma amostra de 42 classificações para os quais se obteve

$$\sum_{i=1}^{42} x_i = 588 \qquad e \qquad \sum_{i=1}^{42} x_i^2 = 8400.$$

   (a) Determine estimativas cêntricas para a média e para a variância da população.

   (b) Qual o grau de confiança que permite afirmar que o verdadeiro valor da média se encontra no interior de um intervalo de amplitude 1.224?

6. Mediu-se uma grandeza X 10 vezes, em condições idênticas, tendo-se obtido os seguintes resultados:

125.3; 124.8; 124.8; 125.1; 125.0; 125.1; 124.7; 125.4; 125.2; 125.0

Admitindo a normalidade da população, calcule:

   (a) estimativas da média e do desvio-padrão da população;

   (b) um intervalo de confiança para a média da população ao grau 0.98;

   (c) um intervalo de confiança para o desvio padrão da população ao grau 0.95.

7. As medidas dos diâmetros de uma amostra aleatória de 200 rolamentos esféricos apresentam uma média de 0.824 polegadas e desvio padrão de 0.042 polegadas. Determine um intervalo com 99% confiança para o valor médio dos diâmetros.

8. Um ecologista ao pretender investigar o nível de poluição por mercúrio em determinado lago, retirou aleatoriamente 20 peixes do referido lago e mediu a concentração de mercúrio nos mesmos. A amostra recolhida foi resumida no seguinte quadro:

classes	$]0,1]$	$]1,2]$	$]2,3]$	$]3,4]$	$]4,5]$	$]5,6]$
efectivos	1	4	6	4	3	2

   (a) Construa o histograma, e calcule a média e a variância da amostra.

   (b) Admitindo a normalidade da população, determine um intervalo de confiança para a variância da população em estudo, ao grau de confiança de 0.95. Que conclusão pode tirar sobre a variação do nível de poluição de peixe para peixe relativamente ao valor médio deste nível?

9. Num laboratório registaram-se os seguintes pontos de fusão de chumbo (em $^\circ C$) numa amostra proveniente de determinado fornecedor:

   329; 345; 330; 328; 342; 334; 337; 341; 343

   Assumindo a normalidade dos pontos de fusão:

   (a) determine estimativas centradas para o ponto de fusão médio do chumbo e variância;

   (b) o fornecedor garante que em 98% dos casos o ponto de fusão médio pode ser considerado 335; através da determinação de um intervalo de confiança conveniente, o que pode dizer acerca da garantia do fornecedor?

   (c) indique um intervalo de confiança a 99% para o desvio padrão da população.

10. Seja X uma v.a. com distribuição normal $N(\mu, \sigma)$ e $(X_1, X_2, ..., X_9)$ uma amostra aleatória da população X. Uma realização desta amostra conduziu ao seguinte intervalo de confiança para μ, a 90%: $]11.728, 12.472[$.

    (a) Determine estimativas para μ e σ^2.

    (b) Como varia a amplitude do intervalo de confiança para μ se:

       i. aumentarmos apenas o grau de confiança?

       ii. aumentarmos apenas a dimensão da amostra?

    (c) Determine um intervalo de confiança a 90% para o desvio padrão σ da população .

11. Certa empresa opera recentemente no mercado da União Europeia na área da distribuição de encomendas. O tempo de entrega duma encomenda, X, através desta nova empresa ainda não está bem caracterizado.

    (a) Considere duas amostras aleatórias, $X_1, X_2, ..., X_{\mathbf{n}}$ e $X'_1, X'_2, ..., X'_{\mathbf{m}}$, com $\mathbf{m} < \mathbf{n}$, de X.

       i. Defina amostra aleatória.

ii. Mostre que as médias amostrais, respetivamente $\overline{X}_{\mathbf{n}}$ e $\overline{X}_{\mathbf{m}}$, são estimadores centrados para o tempo médio de entrega duma encomenda, e compare-os em termos de eficiência.

(b) A empresa garante que todas as encomendas chegam ao seu destinatário, **em média**, em menos de 48 horas, com uma **variabilidade** máxima de 8 horas.

Para avaliar este desempenho, foram recolhidos os tempos (em horas) relativos a uma amostra de 51 encomendas, tendo-se obtido os seguintes resultados:

$$\sum_{i=1}^{51} x_i = 3250; \quad \sum_{i=1}^{51} x_i^2 = 277500.$$

O que pode concluir, **com confiança 95**%, sobre o desempenho da empresa relativamente às **condições** de entrega **referidas**?

12. De uma população ativa de 500 pessoas, de certa região, foram encontrados 41 desempregados. Determine um intervalo de confiança a 95% para a taxa de desempregados dessa região.

13. Um grupo de cientistas defende a tese de que a taxa de mortalidade devida a certa doença é aproximadamente 10%.

(a) Supondo verdadeira a tese daqueles cientistas, calcule a probabilidade de em 10 pessoas, observadas ao acaso entre as afetadas pela referida doença, haver pelo menos uma que acabe por falecer devido à mesma.

(b) Com o objetivo de tirar conclusões sobre a veracidade daquela tese, recolheu-se uma amostra de 500 pessoas afetadas pela doença, das quais faleceram 60. Determine um intervalo real que contenha, com uma confiança de 0.9, a proporção de indivíduos que faleceram com tal doença. Terão os cientistas razão?

14. Seja $(X_1, X_2, ..., X_n)$ uma amostra aleatória de dimensão n de uma variável aleatória real X de média μ e variância σ^2.

(a) Prove que $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ é um estimador cêntrico de μ.

(b) Com o objetivo de estudar a duração de vida média de determinado tipo de peças, recolheu-se uma amostra de 110 elementos que se resumiu no quadro seguinte:

Duração (milhares de horas)	]4,4.5]	]4.5,5]	]5,5.5]	]5.5,6]
número de peças	25	35	30	20

i. Construa o histograma da amostra e calcule estimativas cêntricas para a média e para a variância da população em estudo.

ii. Supondo que a duração de vida das referidas peças é normalmente distribuída (será razoável?), determine um intervalo de confiança para a sua média, ao grau 0.95.

6. Testes de Hipóteses Paramétricos tem erros

1. Uma empresa garante que, se os seus pneus forem utilizados em condições normais, têm um tempo médio de vida superior a 40000 Km. Uma amostra constituída por 31 pneus, utilizados em condições normais, proporcionou os seguintes dados: $\overline{x} = 43200$ e $s_{31} = 8000$ km.

   Teste, ao nível de significância de 5%, se os pneus têm a vida média que a empresa reivindica.

   i) Identificação da população e distribuição:

   Considerando a variável aleatória X ="tempo de vida de um pneu, em km", tem-se que $X \sim N(\mu, \sigma)$

   Parâmetro a testar: μ

   ii) Formulação de hipóteses:

   $$\mathrm{H}_0 : \mu = 40000 \qquad vs \qquad \mathrm{H}_1 : \mu > 40000$$

   iii) Nível de significância: $\alpha = 0.05$

   iv) Estatística de Teste: $Z = \sqrt{n}\,\dfrac{\overline{X} - 40000}{S} \dot{\sim} N(0,1)$

   v) Dedução da Região Crítica (ou de Rejeição): $R.C. = [+z, +\infty[$ onde $P(Z \in R.C.) = \alpha$

   $$P(Z \in R.C.) = \alpha \Leftrightarrow P\left(Z \geq +z\right) = \alpha$$

   $$\Leftrightarrow P\left(Z \leq +z\right) = 1 - \alpha$$

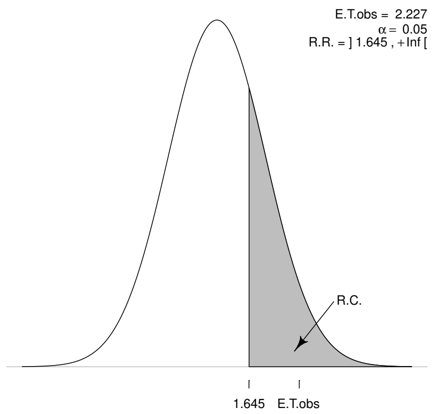
   Consultando as tabelas da lei normal, $z = F_Z^{-1}(0.95) = 1.645$ e, então,

   $R.C. = [1.645, +\infty[$

   vi) Valor da estatística de teste:

   $$Z_{obs} = \sqrt{31}\,\frac{43200 - 40000}{8000} = 2.2271$$

   vii) Tomada de Decisão: porque $Z_{obs} \in R.C.$ a hipótese nula deve ser rejeitada, com $\alpha = 0.05$.



2. Um molde de injeção tem produzido peças de um determinado material isolante térmico com uma resistência à compressão de valor médio 5.18 kg/cm^2 e variância 0.0625 $(kg/cm^2)^2$. As últimas 12 peças produzidas nesse molde foram recolhidas e ensaiadas, tendo-se obtido para a resistência média à compressão o valor 4.95 kg/cm^2. Assuma que a resistência à compressão tem distribuição normal.

   Poder-se-á afirmar, ao nível de significância de 0.05, que as peças produzidas recentemente são menos resistentes do que o habitual?

   i) Identificação da população e distribuição:

   Considerando a variável aleatória $X =$"resistência de uma peça à compressão, em kg/cm^2" , tem-se que $X \sim N(\mu, \sqrt{0.0625})$

   Parâmetro a testar: μ

   ii) Formulação de hipóteses:

   $\mathrm{H}_0 : \mu = 5.18 \qquad vs \qquad \mathrm{H}_1 : \mu < 5.18$

   iii) Nível de significância: $\alpha = 0.05$

   iv) Estatística de Teste: $Z = \sqrt{n}\,\dfrac{\overline{X} - 5.18}{\sqrt{0.0625}} \sim N(0, 1)$

   v) Dedução da Região Crítica: $R.C. =]-\infty, -z]$ onde $P(Z \in R.C.) = \alpha$

$$P(Z \in R.C.) = \alpha \Leftrightarrow P\left(Z \leq -z\right) = \alpha$$

$$\Leftrightarrow P\left(Z \geq +z\right) = \alpha$$

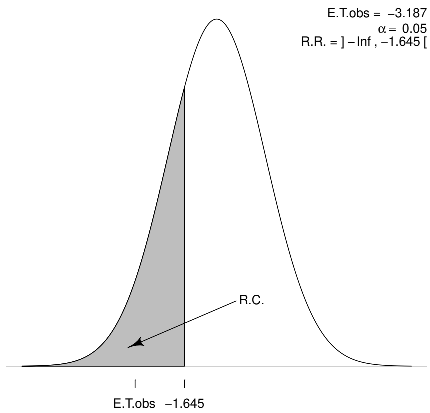
$$\Leftrightarrow P\left(Z \leq +z\right) = 1 - \alpha$$

Consultando tabelas da lei normal, $z = F_Z^{-1}(0.95) = 1.645$ e, então,

$R.C. =]-\infty, -1.645]$

vi) Valor da estatística de teste:

$$Z_{obs} = \sqrt{12}\,\frac{4.95 - 5.18}{\sqrt{0.0625}} = -3.187$$



vii) Tomada de Decisão: porque $Z_{obs} \in R.C.$ a hipótese nula deve ser rejeitada com $\alpha = 0.05$.

3. Considere uma fábrica que produz cabos elétricos cujos diâmetros são normalmente distribuídos com valor médio μ e desvio padrão $\sigma > 0$. Selecionaram-se aleatoriamente 20 cabos da produção da fábrica e registaram-se os seguintes valores:

$$\sum_{i=1}^{20} x_i = 130.27; \qquad \sum_{i=1}^{20} x_i^2 = 849.98\,.$$

Com base na informação anterior, teste ao nível de significância de 1%:

(a) $H_o : \mu = 6.3$ **vs** $\mu \neq 6.3$;

i) Identificação da população e distribuição:

Considerando a variável aleatória X =”diâmetro de um cabo elétrico” , tem-se que $X \sim N(\mu, \sigma)$

Parâmetro a testar: μ

ii) Formulação de hipóteses:

$\text{H}_0 : \mu = 6.3 \qquad vs \qquad \text{H}_1 : \mu \neq 6.3$

iii) Nível de significância: $\alpha = 0.01$

iv) Estatística de Teste: $Z = \sqrt{n}\,\dfrac{\overline{X} - 6.3}{S} \sim t_{20-1}$

v) Dedução da Região Crítica: $R.C. =]-\infty, -z] \cup [+z, +\infty[$ onde $P(Z \in R.C.) = \alpha$

$$P(Z \in R.C.) = \alpha \Leftrightarrow P\left(Z \leq -z \vee Z \geq +z\right) = \alpha$$

$$\Leftrightarrow P\left(Z \geq +z\right) = \frac{\alpha}{2}$$

$$\Leftrightarrow P\left(Z \leq +z\right) = 1 - \frac{\alpha}{2}$$

Consultando tabelas da lei t-Student, $z = F^{-1}_{t_{19}}(0.995) = 2.861$ e, então,

$R.C. =]-\infty, -2.861] \cup [2.861, +\infty[$

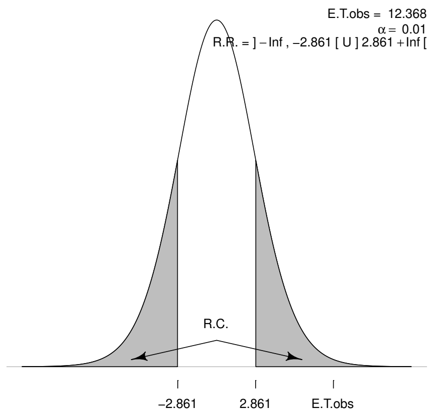
vi) Valor da estatística de teste:

$$\overline{x} = \frac{130.27}{20} = 6.5135 \quad \text{e} \quad s^2 = \frac{849.98}{19} - \frac{20}{19} \cdot 6.5135^2 = 0.0772$$

$$Z_{obs} = \sqrt{20}\,\frac{6.5135 - 6.3}{\sqrt{0.0772}} = 3.436$$

vii) Tomada de Decisão: porque $Z_{obs} \in R.C.$ a hipótese nula deve ser rejeitada para $\alpha = 0.01$.

(b) $H_o : \sigma = 0.5 \qquad$ **vs** $\qquad \sigma = 1.$



i) Identificação da população e distribuição:

Considerando a variável aleatória $X =$"resistência de uma peça à compressão, em kg/cm^2" , tem-se que $X \sim N(\mu, \sigma)$

Parâmetro a testar: σ

ii) Formulação de hipóteses:

$\mathrm{H}_0 : \sigma = 0.5$
$\mathrm{H}_1 : \sigma > 0.5$

o que é equivalente a

$\mathrm{H}_0 : \sigma^2 = 0.5^2$
$\mathrm{H}_1 : \sigma^2 > 0.5^2$

iii) Nível de significância: $\alpha = 0.01$

iv) Estatística de Teste: $Z = \dfrac{(n-1)S^2}{0.5^2} \sim \chi^2_{20-1}$

v) Dedução da Região Crítica: $R.C. = [c, +\infty[$ onde $P(Z \in R.C.) = \alpha$

$$P(Z \in R.C.) = \alpha \Leftrightarrow P\left(Z \geq c\right) = \alpha$$

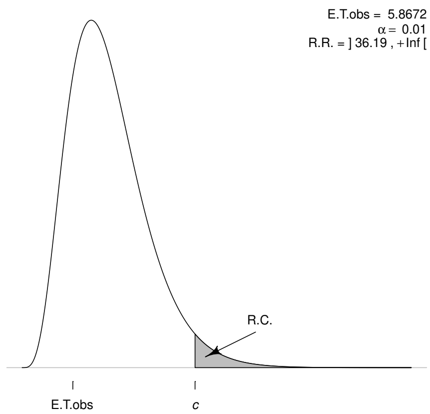
$$\Leftrightarrow P\left(Z \leq c\right) = 1 - \alpha$$

Consultando tabelas da lei Qui-quadrado, $z = F_Z^{-1}(0.99) = 36.19$ e, então,

$R.C. = [36.19, +\infty[$

vi) Valor da estatística de teste:

$$Z_{obs} = \frac{(20-1)\cdot 0.0772}{0.5^2} = 5.8672$$



vii) Tomada de Decisão: porque $Z_{obs} \notin R.C.$ a hipótese nula não deve ser rejeitada, com $\alpha = 0.01$.

Em caso de rejeição de H_0, ter-se-ia de realizar, ao mesmo nível, um segundo teste de hipóteses
$\mathrm{H}_0 : \sigma = 1$
$\mathrm{H}_1 : \sigma \neq 1$

o que é equivalente a

$\mathrm{H}_0 : \sigma^2 = 1^2$
$\mathrm{H}_1 : \sigma^2 \neq 1^2$
Neste caso não será necessária a realização do segundo teste de hipóteses acima mencionada.

4. Certo equipamento de empacotamento automático encontra-se regulado para encher embalagens de 1000 gramas de certo produto. O seu deficiente funcionamento origina prejuízo para a empresa. Aceita-se da experiência passada que o peso das embalagens se comporta normalmente com desvio

padrão de 12 gramas. Para verificar a afinação do equipamento, selecionaram-se aleatoriamente 9 embalagens com os resultados: $\overline{x} = 993.78\,gr$ e $s_9 = 11.29\,gr$.

Teste, ao nível de significância de 10%, se a máquina está a encher corretamente ou não as embalagens.

5. Numa fábrica de automóveis existe uma secção destinada à produção de determinado tipo de peças, cujo comprimento médio deverá ser aproximadamente de 2.5 cm. A secção de controlo de qualidade da referida fábrica afirma que as peças apresentam comprimentos inferiores aos exigidos.

   Com o objetivo de avaliar a veracidade da afirmação proferida pela secção de controlo de qualidade, selecionou-se ao acaso uma amostra de 26 peças na produção de um dia, tendo sido obtido os resultados seguintes:

$$\sum_{i=1}^{26} x_i = 52; \qquad \sum_{i=1}^{26} (x_i - \overline{x})^2 = 13\,.$$

   Admitindo a normalidade da população subjacente aos dados, teste, ao nível de significância de 5%, se secção de controlo de qualidade tem razão.

6. Sabe-se que o tempo diário (em horas) de utilização de um determinado terminal de computador é normalmente distribuído. Foram observados os tempos de utilização durante 10 dias consecutivos, tendo sido obtido os resultados seguintes:

$$\sum_{i=1}^{10} x_i = 56; \qquad \sum_{i=1}^{10} (x_i - \overline{x})^2 = 129.6\,.$$

   (a) Estime pontualmente o tempo médio diário (em horas) de utilização do referido terminal e a variância.

   (b) Determine intervalos de confiança a 95% para a média e variância da população em estudo.

   (c) Teste, ao nível de significância de 5%:

   (i) se o tempo médio diário de utilização do terminal é superior a 6 horas;

   (ii) se o desvio padrão excede as 8 horas.

7. Um agente de compras de um determinado supermercado testou uma amostra aleatória de 100 latas de conserva na própria fábrica de enlatados. O peso médio (em decagramas) encontrado por lata foi de 15.97 com $s_{100} = 0.15$. O fabricante afirma que o peso líquido médio por lata era de 16. Pode esta afirmação ser rejeitada? (use $\alpha = 0.1$)

8. Uma determinada pessoa, interessada em alugar uma loja, é informada que a renda média na área é de 750 euros. Suponha que, para o tipo de zona em questão, é possível dizer que as rendas têm distribuição aproximadamente normal com desvio padrão 50 euros. Para uma amostra aleatória de 15 lojas, a renda média foi de 800 euros.

A pessoa em causa está convencida de que o valor de 750 euros para a renda média está desatualizada. Terá a pessoa razão? Justifique convenientemente a sua resposta, utilizando um teste adequado a 2% de significância.