ELEMENTOS DE PROBABILIDADES E ESTATÍSTICA | 21037

Atividades Formativas (Exercícios e Resoluções)

Catarina S. Nunes
Secção de Matemática, Departamento de Ciências e Tecnologia
Universidade Aberta

UAb

ATIVIDADE FORMATIVA 1

Temas:

- Dados Estatísticos
- Probabilidades

Objetivos

- Amostra. População.
- Gráfico de caule-e-folhas
- Distribuição de frequências. Histogramas
- Características numéricas: média e desvio padrão
- Características numéricas: estatísticas de ordem
- Correlação
- Experiência aleatória e Espaço amostral
- Definição de probabilidade
- Probabilidade Condicional. Teorema de Bayes
- Independência de acontecimentos

Descrição da Atividade (Exercícios a Resolver)

1. Numa determinada empresa existe uma base de dados com informações relativas a cada trabalhador, nomeadamente, sexo, idade, estado civil, número de filhos, habilitações literárias, antiguidade na empresa, categoria profissional, salário actual, avaliação obtida no ano transacto. Classifique a natureza de cada item.

2. Complete o quadro seguinte usando a informação indicada:

x_i	n_i	N_i	f_i	F_i
A		12		
B			0.36	
C		40		0.80
D				

3. Os resultados parciais de um inquérito realizado para saber quantas vezes por ano os utentes inscritos num determinado centro de saúde recorrem ao seu médico de família constam da tabela seguinte:

x_i	n_i	N_i	f_i	F_i
0	100			
1				0.50
2			0.24	
3	64			
4	40			

3.1 Quantas pessoas foram inquiridas neste inquérito?

3.2 Complete a tabela.

3.3 Determine a mediana de visitas anuais ao médico de família.

3.4 Calcule o 1º e o 3º quartis de visitas por ano ao médico de família.

4. Com o intuito de sumariamente averiguar a escolaridade dos habitantes das freguesias A e B de uma mesma autarquia, em cada freguesia foram inquiridas 15 pessoas, com idades compreendidas entre os 30 e os 50 anos, sobre o nível de escolaridade de possuiem. Os resultados obtidos foram os seguintes:

Freguesia A

1º Ciclo	2º Ciclo	Secundário	Licenciatura	Mestrado
3º Ciclo	3º Ciclo	Mestrado	Secundário	1ºCiclo
2º Ciclo	3ºCiclo	Licenciatura	1ºCiclo	2º Ciclo

Freguesia B

2º Ciclo	Licenciatura	Licenciatura	Licenciatura	Licenciatura
2º Ciclo	1º Ciclo	Secundário	Doutoramento	Secundário
Licenciatura	Mestrado	3ºCiclo	Secundário	Doutoramento

4.1 Classifique o item em estudo.

4.2 De acordo com a amostra, em cada freguesia qual é a moda de escolaridade da população na faixa etária dos 30-50 anos de idade?

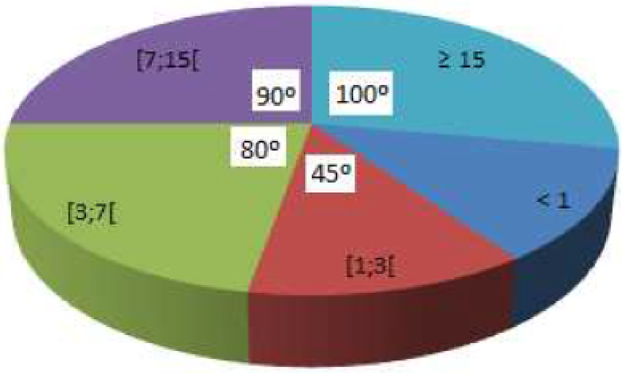
4.3 Para cada freguesia, determine a mediana da escolaridade dos seus habitantes com idades compreendidas entre os 30 e os 50 anos de idade.

4.4 Será que faz sentido calcular a média? Justifique.

4.5 Para cada freguesia desenhe um gráfico que melhor traduza os dados recolhidos no inquérito.

4.6 Para a faixa etária em análise, em qual das freguesias é maior o grau de escolaridade? Justifique.

5. A antiguidade dos 400 funcionários de uma empresa com 25 anos de actividade está identificada no gráfico seguinte:



5.1 Construa a tabela de frequências absolutas e de frequências relativas, que inclua as respectivas frequências acumuladas.

5.2 Calcule a antiguidade média dos funcionários da empresa.

5.3 Identifique a classe modal.

5.4 Determine a mediana.

6. Os dados seguintes são o resultado de um teste clínico realizado a 24 pessoas para análise de um determinado parâmetro sanguíneo:

2.12	2.10	3.20	5.40	5.60	1.75	2.30	2.80	3.10	4.20	4.40	4.30
1.95	1.85	1.83	5.80	2.76	5.60	5.40	2.65	2.98	1.98	2.30	2.30

6.1 Calcule a média.

6.2 Suponha que aos 24 dados chega um 25º exame clínico ao mesmo parâmetro sanguíneo e que a média baixa para 3.23. Qual foi o resultado desse 25º exame?

7. Na tabela seguinte apresentam-se os dados relativos à produção anual de bolota e ao peso médio dos suinos criados numa determinada propriedade agricola, nos últimos seis anos:

Bolota produzida (ton.)	10	17	32	35	38	48
Peso Médio (Kg)	95	107	101	97	98	105

7.1 Determine a produção média de bolota nos últimos seis anos.

7.2 Calcule o desvio padrão dos dados referentes à produção de bolota.

7.3 Desenhe o diagrama de dispersão e diga se se pode concluir algo sobre a relação entre a produção anual de bolota e o peso médio dos suinos criados nessa propriedade. Justifique.

7.4 Calcule o coeficiente de correlação e interprete-o.

8. O Índice de Desenvolvimento Humano (IDH), que surge no Programa das Nações Unidas para o Desenvolvimento e no Relatório de Desenvolvimento Humano, é uma medida comparativa usada para classificar os países pelo seu grau de desenvolvimento humano. A determinação do IDH de cada país é feita com base na expectativa de vida ao nascer, taxa de fertilidade, educação e PIB per capita. A exemplo, no quadro seguinte apresentam-se os indicadores relativos a alguns países:

Grupo de desenvolvimento: Muito alto · Alto · Médio · Baixo

Ranking	País	IDH	Expectativa de vida	Média de anos de estudo	PIB per capita (US$)	Taxa de fertilidade
1	Noruega	0,943	81,1	12,6	47.557	2,0
4	EUA	0,910	78,5	12,4	43.017	2,1
45	Argentina	0,797	75,9	9,3	14.527	2,2
48	Uruguai	0,783	77,0	8,5	13.242	2,0
51	Cuba	0,776	79,1	9,9	5.416	1,5
57	México	0,770	77,0	8,5	13.245	2,2
84	Brasil	0,718	73,5	7,2	10.162	1,8
101	China	0,687	73,5	7,5	7.476	1,6
187	Rep. Dem. do Congo	0,286	48,4	3,5	280	5,5

Fonte: ONU

8.1 Determine o coeficiente de correlação linear entre os cinco indicadores constantes no quadro (IDH, expectativa de vida, média de anos de estudo, PIB per capita e taxa de fertilidade).

8.2 Interprete os resultados obtidos.

9. Determine o espaço de resultados de cada uma das seguintes experiências aleatórias:

9.1 Lançamento de uma moeda e de um dado.

9.2 Lançamento de uma moeda, seguido do lançamento de outra moeda, igual.

9.3 Lançamento de duas moedas iguais, em simultâneo.

9.4 Lançamento de dois dados, um vermelho e outro amarelo.

9.5 Lançamento de dois dados iguais.

10. Uma urna contém 50 bolas numeradas de 1 a 50. Considere a experiência que consiste em retirar, aleatoriamente e sem reposição, cinco bolas da urna.

10.1 Determine o espaço de resultados desta experiência aleatória, supondo que:

[10.1.1] as bolas são retiradas em simultâneo.

[10.1.2] as bolas são retiradas uma a uma.

10.2 Determine o número de maneiras que se podem retirar 5 bolas da urna.

11. A chave do concurso Euromilhões é **um conjunto** de 5 números extraídos, um a um, de uma tômbola contendo 50 bolas numeradas de 1 a 50.

11.1 Identifique o espaço de resultados.

11.2 Calcule quantas chaves diferentes do Euromilhões é que existem.

12. Uma escola abriu um concurso para admissão de novos professores. Entre as 80 candidaturas recebidas, 40 revelam possuir experiência anterior, 30 possuem certificado profissional e 10 revelam possuir, quer experiência profissional, quer certificado profissional.

12.1 Qual a probabilidade de que um candidato, escolhido ao acaso, tenha experiência anterior ou certificado profissional?

12.2 Qual a probabilidade de que um candidato, escolhido aleatoriamente, tenha experiência ou certificado profissional, mas não ambos?

12.3 Seleccionou-se ao acaso uma das 80 candidaturas e verificou-se que o candidato possuia experiência profissional. Qual a probabilidade deste candidato ter também certificado profissional?

13. Sejam A e B dois acontecimentos independentes tais que

$$P(A|B) = \frac{1}{3}, \quad P(A \cup B) = \frac{7}{9}$$

Determine a probabilidade de A e de B.

14. Dado um acontecimento A tal que $P(A) = \frac{4}{5}$, seja B um acontecimento tal que

$$P(B|A) = \frac{1}{5}$$

Determine a probabilidade de B que garanta que os acontecimentos A e B sejam independentes.

15. Numa região desértica em África, a seca é um dos problemas com que a população mais frequentemente se depara. Numa tentativa de minorar este problema, pretendese abrir um furo que permita o abastecimento de água à população. Depois de alguns estudos, chegou-se à conclusão que a probabilidade de existir água no subsolo de uma determinada área é de 0.3 e que, caso esta exista, a probabilidade de se encontrar água logo na primeira tentativa é de 0.5.

15.1 Qual é a probabilidade de não se encontrar água na primeira tentativa?

15.2 Sabendo que na primeira tentativa não se encontrou água, qual é a nova probabilidade atribuída à existência de água nessa área?

16 A tabela seguinte mostra as percentagens de adultos do País X distribuídas pelas diferentes categorias do Indíce de Massa Corporal (IMC).

Categoria de Peso	Sexo Masculino (%)	Sexo Feminino (%)
$IMC < 25$	15.4	23.3
$25 \leq IMC < 30$	21.9	14.9
$IMC \geq 30$	12.3	12.2

Suponha que se escolhe uma pessoa aleatoriamente do País X.

16.1 Qual a probabilidade desta pessoa ter um $IMC \geq 25$?

16.2 Qual a probabilidade desta pessoa ter $25 \leq IMC < 30$ ou ser do sexo masculino?

16.3 Considerando os problemas de saúde associados com a obesidade, a tabela anterior foi reorganizada:

Categoria de Peso	Sexo Masculino (%)	Sexo Feminino (%)
Não obeso	38	38
Obeso	12	12

Considerando esta nova tabela e utilizando a teoria das probabilidades, será o sexo um factor de risco para a obesidade?

17. Considere que lança dois dados (de 6 faces) equilibrados ao mesmo tempo.

17.1 Qual a probabilidade de obter um duplo (os dois dados com a mesma face)?

17.2 Qual a probabilidade de ao lançar os dois dados, obter 7 ou 8 (na soma das pintas dos dois dados)?

17.3 Qual a probabilidade de ao lançar os dois dados, obter 8 ou um duplo?

UAb

ATIVIDADE FORMATIVA 2

Temas:

- Variáveis Aleatórias
- Variáveis Aleatórias Discretas

Objetivos

- Definição de variável aleatória
- Função de distribuição
- Variáveis aleatórias bidimensionais discretas
- Função de probabilidade e função de distribuição
- Valor esperado. Momentos.
- Algumas distribuições paramétricas discretas: Distribuição uniforme
- Distribuição de Bernoulli. Distribuição binomial
- Distribuição geométrica. Distribuição hipergeométrica
- Distribuição de Poisson

Descrição da Atividade (Exercícios a Resolver)

1. Numa casa de jogo ilegal existe uma roleta, com casas numeradas de 1 a 20, que está viciada: a probabilidade de sair uma casa com numeração superior a 14 é o quádruplo da probabilidade de sair uma casa com numeração inferior a 15.

1.1 Indique o espaço de resultados e a variável aleatória subjacente.

1.2 Calcule a probabilidade de sair uma casa numerada com um valor superior a 17.

1.3 Determine a probabilidade de sair uma casa numerada com um número par.

2. Considere o lançamento de um par de dados usuais. Em cada jogada observa-se o número de pintas das faces que ficam viradas para cima, somam-se os pontos assim obtidos e, à soma obtida, aplica-se a regra "'noves fora"'[1]. Por exemplo, se num dado sairem 6 pintas e no outro dado 4 pintas, a soma dá 10 que, "'noves fora"', dá 1.

[1] Esta regra é equivalente à determinação do resto da divisão por 9. Dito de outro modo, o valor que se obtém por aplicação da regra "'noves fora"' a um número ´é igual ao valor do resto da divisão desse número por 9.

2.1 Identifique o espaço de resultados e a variável aleatória X subjacente.

2.2 De quantas maneiras se pode obter pelo procedimento descrito o número 2?

2.3 Calcule o valor de $P(X = 2)$ e de $P(X \leq 5|X \geq 2)$.

2.4 Determine a função de probabilidade e a função de distribuição de X.

3. Seja X uma variável aleatória discreta, com valores no conjunto $\{-1, 0, 1, 2\}$, e de função de probabilidade dada pela tabela seguinte:

x	-1	0	1	2
$f(x)$	2α	0.3	0.4	α

3.1 Determine o valor de α.

3.2 Calcule $E(X + 2)$, $E(X^2)$ e $P(X < E(X^2))$.

3.3 Determine a função de probabilidade da variável aleatória $Y = X^2$.

4. Determine a probabilidade de em cinco lançamentos de um dado usual sair duas vezes, e apenas duas vezes, o número dois.

5. Uma caixa contém 80 berlindes dos quais 15 apresentam pequenos defeitos. Assumindo que uma tiragem significa uma extração aleatória de um berlinde seguida da sua reposição na caixa, determine:

5.1 A probabilidade de em 20 tiragens ser encontrado um único berlinde defeituoso.

5.2 Quantas tiragens devem ser realizadas de modo que a probabilidade de se retirar um berlinde defeituoso seja superior a 40%.

6. No contexto do exercício anterior, suponha que são retiradas, sem reposição e aleatoriamente, 20 berlindes da caixa. Calcule a probabilidade de entre os 20 berlindes retirados, dois serem defeituosos.

7. Determine a função geradora de momentos correspondente à distribuição de Poisson.

8. Por uma questão de contenção de despesas, alguns serviços de saúde têm vindo a ser extintos e outros agrupados em novos centros de atendimento. Suponha que numa determinada freguesia existiam três centros, A, B e C, que, independentemente, prestavam serviços de saúde à população local e que o número de pessoas que se dirigiam por hora a estes centros seguia uma distribuição de Poisson a uma média, respectivamente, 5, 10 e 3 ocorrências.
Agora que os serviços de saúde prestados nestes três centros estão concentrados num único centro novo, determine a probabilidade de numa hora serem atendidas 12 pessoas neste novo centro.

9. Sejam X e Y duas variáveis aleatórias discretas com valores no conjunto, respetivamente, $\{0, 1, 2, 3\}$ e $\{0, 1, 2, 3, 4\}$, e de função de probabilidade, respetivamente,

$$f_X(x) = \begin{cases} 0.30 & se \quad x = 0 \\ 0.20 & se \quad x = 1 \\ 0.25 & se \quad x \in \{2, 3\} \end{cases} \qquad f_Y(y) = \begin{cases} 0.40 & se \quad y = 0 \\ \alpha & se \quad y = 1 \\ 0.20 & se \quad y = 2 \\ 0.10 & se \quad y = 3 \\ 0.05 & se \quad y = 4 \end{cases}$$

Relativamente à função de probabilidade conjunta das variáveis X e Y , f, conhecem-se os valores constantes da tabela seguinte:

$f(x,y)$	$y=0$	$y=1$	$y=2$	$y=3$	$y=4$
$x=0$	0.10		0.05	0	0
$x=1$	0	0.05	0.05	0.05	
$x=2$			0	0.05	
$x=3$	0.15	0		0	

9.1 Sem determinar o valor de α, complete a tabela dada.

9.2 Determine o valor de $P(Y = 1)$.

9.3 Calcule a covariância entre X e Y e utilize o valor obtido para averiguar se as variáveis X e Y são, ou não, independentes.

10. A função de probabilidade conjunta de duas variáveis aleatórias discretas X e Y com valores no conjunto dos números inteiros não negativos $\mathbb{N} \cup \{0\}$ é igual a

$$f(x, y) = c(2x + y)$$

se $x \in \{0, 1, 2\}$ e se $y \in \{0, 1, 2, 3\}$, sendo $f(x, y) = 0$ em todos os outros casos.

10.1 Determine:

[10.1.1] O valor de c.

[10.1.2] O valor de $P(X = 2, Y = 1)$.

[10.1.3] O valor de $P(X \geq 1, Y \leq 2)$.

[10.1.4] As funções de probabilidade marginal de X e de Y .

10.2 Verifique se as variáveis aleatórias X e Y são, ou não, dependentes.

11. Considere o seguinte conjunto de agregados familiares com dois titulares, cuja função de probabilidade conjunta do rendimento mensal em euros do primeiro titular X e do segundo titular Y está representada na seguinte tabela:

	$Y = 1000$	$Y = 2000$	$Y = 3000$	$Y = 4000$
$X = 1000$	0.20	0.04	0.01	0
$X = 2000$	0.10	0.36	0.09	0
$X = 3000$	0	0.05	0.10	0
$X = 4000$	0	0	0	0.05

11.1 Obtenha as funções de probabilidades marginais para as variáveis X e Y.

11.2 Considere apenas os segundos titulares com um rendimento de 2000 euros mensais. Qual a probabilidade do primeiro titular ter um rendimento igual?

11.3 Calcule o valor esperado e o desvio padrão da variável X e da variável Y.

11.4 Serão X e Y variáveis aleatórias independentes?

1.5 Calcule a covariância e o coeficiente de correlação entre X e Y.

11.6 Calcule o valor esperado e o desvio padrão do rendimento total (R), considerando $R = X + Y$.

12. De uma fornada de 300 bolinhos de feijão retiram-se 5. Admitindo que a percentagem de bolinhos de tamanho não conforme (abaixo do permitido) na fornada é de 1%, calcule a probabilidade de, entre os 5 bolinhos retirados, haver pelo menos um não conforme.

UAb

ATIVIDADE FORMATIVA 3

Temas:

- Variáveis Aleatórias Contínuas

Objetivos

- Função de distribuição e função densidade de probabilidade
- Valor esperado. Momentos
- Algumas distribuições paramétricas contínuas
- Distribuição uniforme. Distribuição exponencial
- Distribuição gama. Distribuição do qui-quadrado
- Distribuição normal
- Teorema do limite central

Descrição da Atividade (Exercícios a Resolver)

1. Considere uma variável aleatória X com a seguinte função densidade de probabilidade

$$f_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{a}{\sqrt{x}}, & 0 < x \leq 1 \\ 0, & x > 1 \end{cases}$$

1.1 Determine o valor de a.

1.2 Calcule o valor esperado de X.

1.3 Determine a função de distribuição de X.

1.4 Calcule a probabilidade de em três observações independentes da variável aleatória X se obter, em todas elas, valores superiores a 0.25.

2. A altura dos homens de qualquer região revela ser uma variável normalmente distribuída. No caso dos homens portugueses, esta variável tem uma distribuição normal com média de 167 cm de altura e um desvio padrão de 3 cm.

2.1 Qual a percentagem de homens portugueses com uma altura não superior a 167 cm? Justifique.

2.2 Qual a probabilidade de um homem escolhido ao acaso ter uma altura superior a 170 cm? Justifique.

2.3 Escolhida uma amostra aleatória de quatro homens portugueses, calcule:

[2.3.1] A probabilidade de todos eles terem uma altura superior a 170 cm.

[2.3.2] A probabilidade de apenas dois deles terem uma altura inferior à média.

3. A variável aleatória X representa o diâmetro de uma componente cilíndrica de uma peça automóvel tem uma função de densidade de probabilidade:

$$f_X(x) = \begin{cases} 0, & x \leq 1 \\ k(x-1)(x-2), & 1 < x < 2 \\ 0, & x \geq 2 \end{cases}$$

3.1 Determine k de modo a que $f_X(x)$ represente uma função de densidade de probabilidade. Justifique.

3.2 Determine o 3º Quartil da distribuição. Justifique.

3.3 O valor de venda da peça automóvel no mercado é uma variável aleatória Y com distribuição normal ($Y \sim N(\mu, \sigma^2)$, com $\mu = 100$ unidades monterárias (u.m.) e desvio padrão $\sigma = 30$ u.m., variação esta que está relacionada com o país de venda. Duas oficinas da mesma marca mas localizadas em dois países diferentes venderam 1 peça durante a manhã de sábado. Calcule a probabilidade do montante global de venda ser superior a 250 u.m.

4. A procura semanal de um certo bem produzido pela empresa Q comporta-se segundo uma distribuição normal de média 20 toneladas e desvio padrão 5 toneladas. A empresa Q incorre em certos custos que se processam do seguinte modo: serão de 2500 u.m. (unidades monetárias) se a procura for inferior a 17 toneladas; serão de 2000 u.m. para valores de procura entre 17 e 22 toneladas; serão de 3000 u.m. quando a procura é superior a 22 toneladas.

4.1 Calcule a probabilidade de numa semana qualquer a procura se situar entre 17 e 22 toneladas.

4.2 Calcule a probabilidade de em duas semanas consecutivas a procura exceder as 25 toneladas? Assuma os pressupostos que achar necessários para responder à questão.

4.3 Qual o valor esperado dos custos referidos?

ATIVIDADE FORMATIVA 1 - Proposta de Resolução

1. Para os itens em análise temos:
Dados Qualitativos: sexo, estado civil, habilitações literárias e categoria profissional.
Dados Quantitativos: idade, número de filhos, antiguidade na empresa, salário atual e avaliação obtida no ano transato.

2. O enunciado é omisso quanto ao número total N de observações. Este pode ser obtido tendo em atenção que, por definição de frequência absoluta acumulada,

$$0.8 = F_C = \frac{N_C}{N} = \frac{40}{N}$$

ou seja, $N = 50$. Deste modo, por definição de frequência relativa, tem-se

$$0.36 = f_B = \frac{n_B}{N} \Rightarrow n_B = 18$$

pelo que a frequência absoluta acumulada N_B é igual a

$$N_B = N_A + n_B = 12 + 8 = 30$$

Isto permite-nos determinar agora os valores em falta na quarta linha do quadro:

$$n_C = N_C - N_B = 40 - 30 = 10 \Rightarrow f_C = \frac{n_C}{N} = 0.20$$

Para completar a última linha, relativa a D, observe-se que

$$N = n_A + n_B + n_C + n_D = N_B + n_C + n_D = N_C + n_D = N_D$$

e

$$f_A + f_B + f_C + f_D = F_B + f_C + f_D = F_C + f_d = F_D = 1$$

Deste modo,

$$f_D = F_D - F_C = 1 - 0.80 = 0.20 \Rightarrow n_D = Nf_D = 10$$

Para o preenchimento da linha relativa a A, basta notar que $n_A = N_A = 12$ e que $f_A = F_A = \frac{n_A}{N}$. Em particular, resulta daqui que

$$F_B = F_A + n_B = 0.24 + 0.36 = 0.60$$

Fica assim completo o quadro:

x_i	n_i	N_i	f_i	F_i
A	12	12	0.24	0.24
B	18	30	0.36	0.60
C	10	40	0.20	0.80
D	10	50	0.20	1

3.

3.1 As duas primeiras questões deste exercício são semelhantes ao exercício anterior, mas ligeiramente mais complexa a determinação do tamanho N da amostra. Para tal note-se o seguinte. Como

$$f_0 + f_1 + f_2 = F_1 + f_2 = 0.74 \quad e \quad f_0 + f_1 + f_2 + f_3 + f_4 = 1$$

tem-se

$$f_3 + f_4 = 1 - 0.74 = 0.26$$

Mas, por definição de frequência relativa,

$$0.26 = f_3 + f_4 = \frac{n_3 + n_4}{N} = \frac{64 + 40}{N}$$

o que conduz a $N = 400$.

3.2 O exercício segue agora como no exercício 2. Para a segunda linha tem-se

$$N_0 = n_0 = 100, \quad F_0 = f_0 = \frac{n_0}{N} = 0.25$$

para a terceira linha,

$$f_1 = F_1 - f_0 = 0.50 - 0.25 = 0.25, \quad n_1 = Nf_1 = 100, \quad N_1 = N_0 + n_1 = 200$$

para a quarta linha,

$$n_2 = Nf_2 = 96, \quad N_2 = N_1 + n_2 = 296, \quad F_2 = F_1 + f_2 = 0.74$$

para a quinta linha,

$$N_3 = N_2 + n_3 = 360, \quad f_3 = \frac{n_3}{N} = 0.16, \quad F_3 = F_2 + f_3 = 0.90$$

e, finalmente, para a última linha, e pelas mesmas razões indicadas no exercício 2,

$$N_4 = N = 400, \quad f_4 = \frac{n_4}{N} = 0.10, \quad F_4 = 1$$

Obtém-se assim o quadro seguinte:

x_i	n_i	N_i	f_i	F_i
0	100	100	0.25	0.25
1	100	200	0.25	0.50
2	96	296	0.24	0.74
3	64	360	0.16	0.90
4	40	400	0.10	1

3.3 Ordenando por ordem crescente os dados obtidos, a mediana corresponde ao valor tal que 50% das observações são menores ou iguais a ela. Contudo, no caso presentemente em análise, a coleção tem um número par de dados: $N = 400$. Logo, neste caso não há um elemento central da coleção dos dados ordenados, mas um par de elementos centrais, correspondentes aos dados com as ordens 200 e 201.

Por observação do quadro construído na alínea anterior (que de certo modo já tem

implícita a ordenação mencionada), verifica-se que o 200º elemento da coleção ordenada foi uma vez ao seu médico de família ($N_1 = 200$ com $n_1 \neq 0$) e o 201º elemento foi duas vezes ao seu médico de família ($N_2 = 296$ e $N_1 = 200$). A mediana é então igual ao valor da semi-soma $\frac{1+2}{2} = 1.5$.
Observe-se, contudo, que não é necessário ordenar os dados para se determinar as ordens 200 e 201. Estas podem ser obtidas pelos cálculos $\frac{400}{2}$ e $\frac{400}{2} + 1$.

3.4 Pensando ainda que os dados estão ordenados por ordem crescente, o 1º quartil corresponde ao valor tal que 25% das observações são menores ou iguais a ele. Como a amostra tem um número par de dados, isto é equivalente a determinar a mediana dos primeiros 200 elementos da amostra. Como 200 é também um número par, um raciocínio semelhante ao realizado na alínea 3.3 permite então concluir que $Q_1 = 0.5$.
O mesmo valor pode ainda ser obtido a partir dos 400 dados iniciais e utilizando os cálculos $\frac{400}{4} = 100$, $\frac{400}{4} + 1 = 101$ para determinar que dados intervirão no cálculo do 1º quartil. Por observação da tabela da alínea 3.2, conclui-se que o 100º dado corresponde a nenhuma ida ao médico de família ($N_0 = 100$) e o 101º dado a uma ida ao médico de família ($N_1 = 200$), pelo que o 1º quartil é igual à semi-soma $\frac{0+1}{2} = 0.5$.
Considerações semelhantes são naturalmente válidas para o 3º quartil. Por determinaço da mediana dos 200 últimos elementos da coleção de dados ordenados, conclui-se que $Q_3 = 3$. O mesmo valor pode ainda ser obtido pelos cálculos $\frac{3\times400}{4} = 300$, $\frac{3\times400}{4} + 1 = 301$ e notando que o 300º e o 301º dados correspondem ambos a três idas ao médico de família ($N_2 = 296$, inferior a 300, mas $N_3 = 360$). Tem-se então $Q_3 = (3+3)/2 = 3$.

4.

4.1 O nível de escolaridade é um atributo qualitativo. Como existe uma relação de ordem entre os diferentes níveis de escolaridade, o item em estudo é qualitativo ordinal.

4.2 Para responder a esta questão, agrupemos primeiro os dados de acordo com as respostas dadas em cada uma das freguesias. No caso da freguesia A, 3 pessoas responderam ter o 1º Ciclo, 3 pessoas o 2º Ciclo, 3 pessoas o 3º Ciclo, duas o Secundário, duas a licenciatura e duas pessoas o mestrado. Nenhuma das pessoas inquiridas possuia o doutoramento. Feita a mesma análise para os inquéritos realizados na freguesia B obtém-se o quadro seguinte:

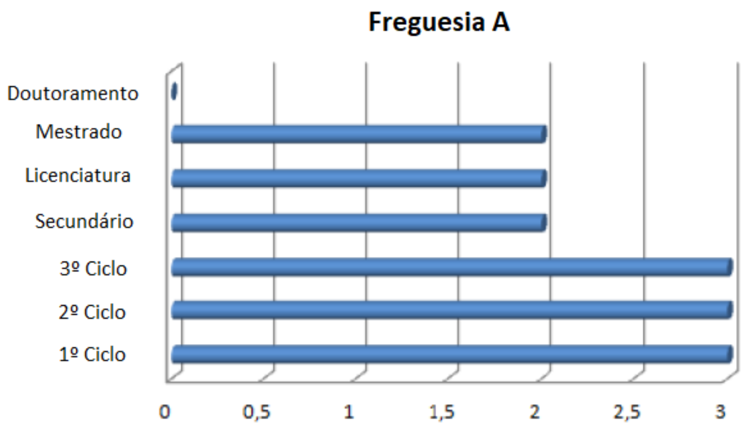
	n_i (Freguesia A)	n_i (Freguesia B)
1º Ciclo	3	1
2º Ciclo	3	2
3º Ciclo	3	1
Secundário	2	3
Licenciatura	2	5
Mestrado	2	1
Doutoramento	0	2

Na freguesia B, a licenciatura é o grau de escolaridade que acolheu mais respostas: 5 ($n_{Licenciatura} = 5$). Assim, nesta freguesia a moda é 5. Já na freguesia A, o maior número de resposta, 3, foi para o 1º Ciclo, 2º Ciclo e 3º Ciclo. Neste caso, a moda é a escolaridade entre o 1º e o 3º Ciclo.

4.3 Para cada freguesia, a amostra é de 15 elementos, um número ímpar. Ordenando os dados por ordem crescente do nível de escolaridade, o ponto central da coleção de dados corresponde ao dado com a ordem 8. Obviamente, não é necessário fazer explicitamente esta ordenação: o valor pode ser obtido pelo cálculo $(15+1)/2$.
Assim, no caso da freguesia A a oitava resposta é o 3º Ciclo. Com efeito, por análise do quadro da alínea 4.2 (que, de certo modo, tem implícita a ordenação crescente dos dados), as 3 primeiras respostas são o 1º Ciclo, as 3 respostas seguintes são o 2º Ciclo (o que já dá um total de 6 resposta), e a sétima, a oitava e a nona respostas são o 3º Ciclo. Está assim encontrado o valor da mediana para a freguesia A: 3º Ciclo.
De modo semelhante, conclui-se que a mediana de escolaridade dos habitantes da freguesia B é a licenciatura.

4.4 Sendo a variável em análise qualitativa ordinal, não faz sentido cálcular a média.

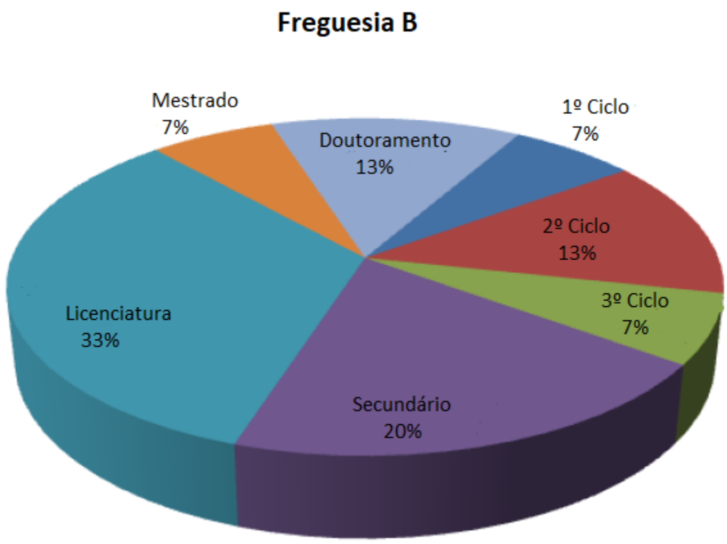
4.5 Para a freguesia A podemos, por exemplo, pensar num gráfico de barras para as frequências absolutas:



Para a freguesia B utilizaremos um gráfico circular ou de sectores representativo das frequências relativas:

	n_i	$f_i = \frac{n_i}{15}$
1º Ciclo	1	0.067
2º Ciclo	2	0.133
3º Ciclo	1	0.067
Secundário	3	0.200
Licenciatura	5	0.333
Mestrado	1	0.067
Doutoramento	2	0.133

Para se representar as frequências relativas sob a forma de um gráfico circular, comece-se por observar que o ângulo de um círculo completo é de 360º. Logo, por exemplo, $20\% = f_{secundario}$ corresponde a uma "'fatia"' do círculo com um ângulo de 72º (regra de três simples). Como a soma das frequências relativas é igual a 100%, podemos deste modo "'fatiar"' todo o círculo



4.6 Os gráficos obtidos na alínea anterior são relativos a frequências diferentes: para a freguesia A, frequências absolutas, para a freguesia B, frequências relativas. Isto significa que ter-se-á que ter alguma atenção para que, através dos dois gráficos, se possa concluir qual das freguesias apresenta maior nível de escolaridade. Mas observando ambos os gráficos, ressalta de imediato que no gráfico circular a licenciatura, mestrado e doutoramento ocupam mais de metade do ”‘bolo”’, correspondendo a mais de metade dos 15 habitantes inquiridos; enquanto que no gráfico de barras, estes três graus académicos correspondem apenas a 4 respostas, numa amostra de 15 elementos. Portanto, menos de metade das respostas. Conclusão: o grau de escolaridade entre a população com idades compreendidas entre os 30 e os 50 anos é mais elevado na freguesia B do que na freguesia A.

5.

5.1 Se o ângulo de um círculo completo é 360º, então uma ”‘fatia”’ do círculo com um ângulo de 90º corresponde a 25% do círculo (novamente, pela regra de três simples). Em termos do gráfico do enunciado, isto significa que 25% dos 400 funcionários estão a trabalhar na empresa há 7 ou mais anos, mas há menos de 15 anos. Ou seja, 100 funcionários. Tem-se, assim,

$$n_{[7,15[} = 100, \quad f_{[7,15[} = 0.25$$

O mesmo raciocínio aplica-se às restantes ”‘fatias”’ do gráfico, tendo que se ter alguma atenção com os arredondamentos. Por exemplo, o mesmo raciocínio aplicado ao sector representativo dos funcionários que trabalham há 15 ou mais anos na empresa mostra que 27.78% dos 400 empregados estão nessa situação. Se se fizer o cálculo para se determinar o número destes funcionários, obtém-se o valor 111.1, que, naturalmente, não faz qualquer sentido quando estamos a falar em pessoas. O número será então 111 empregados.
Ainda em relação ao gráfico, observe-se que no sector relativo aos funcionários mais recentes (há menos de 1 ano na empresa), não é indicado o ângulo. Tal é igual a

$$360^\circ - (100 + 90 + 80 + 45)^\circ = 45^\circ$$

Pelos argumentos indicados, constrói-se assim a tabela seguinte:

Antiguidade	n_i	N_i	f_i	F_i
< 1	50	50	0.125	0.125
$[\,1, 3\,[$	50	100	0.125	0.25
$[\,3, 7\,[$	89	189	0.2222	0.4722
$[\,7, 15\,[$	100	289	0.25	0.7222
≥ 15	111	400	0.2778	1

5.2 Como os dados estão agrupados por classes, para se calcular a média utiliza-se o ponto médio x_i' de cada classe $[\,a_i, b_i\,[$

$$x_i' = \frac{a_i + b_i}{2}$$

Para o caso em análise, note-se que o ponto médio da classe dos funcionários que estão há menos de um ano na empresa é o ponto médio do intervalo $[\,0, 1\,[$ (não há anos de antiguidade negativos), e o ponto médio da classe dos empregados que estão há 15 ou mais anos na empresa é o ponto médio do intervalo $[15, 25]$ (a empresa tem 25 anos). Tem-se assim,

	< 1	$[\,1, 3\,[$	$[\,3, 7\,[$	$[\,7, 15\,[$	≥ 15
x_i'	0.5	2	5	11	20

pelo que

$$\bar{x} = \frac{1}{n}\sum_i x_i' n_i = \sum_i x_i' f_i \quad (n = 400)$$

$$= 0.5 \times 0.125 + 2 \times 0.125 + 5 \times 0.2222 + 11 \times 0.25 + 20 \times 0.2778 = 9.7295$$

5.3 Como as classes não têm a mesma amplitude, a classe modal é aquela a que corresponde a maior frequência relativa corrigida, ou seja, é a classe $[\,a_i, b_i\,[$ que apresenta maior valor para o quociente $\frac{f_i}{h_i}$, onde $h_i = b_i - a_i$ é a amplitude da classe $[\,a_i, b_i\,[$. Por análise da tabela seguinte, conclui-se que é a classe dos funcionários que trabalham há menos de um ano na empresa:

Antiguidade	h_i	f_i	$\frac{f_i}{h_i}$
< 1	1	0.125	0.125
$[\,1, 3\,[$	2	0.125	0.0625
$[\,3, 7\,[$	4	0.2222	0.05555
$[\,7, 15\,[$	8	0.25	0.03125
≥ 15	10	0.2778	0.02778

5.4 A mediana M_e corresponde ao valor cuja frequência relativa acumulada é 50% (isto é, $F(M_e) = 0.50$). Observando a tabela da alínea 5.1, verifica-se que $F_{[7,15[} = 72.22\%$ (superior a 50%) e $F_{[3,7[} = 47.22\%$ (inferior a 50%). Logo, a mediana está localizada na classe $[\,7, 15\,[$.
Pode então acontecer que os 100 funcionários que estão na classe $[\,7, 15\,[$ estejam todos há apenas 7 anos na empressa, ou pode acontecer que 50% destes funcionários estejam há 7-12 anos na empresa e os restantes 50% há 12-14 anos, etc. Não se conhecendo como se distribuem estes funcionários pelos anos de trabalho $[\,7, 15\,[$ na empresa, assume-se, para efeitos de cálculo da mediana, que a frequência relativa acumulada se encontra uniformemente distribuída ao longo do intervalo $[\,7, 15\,[$.

Em termos gráficos, isto significa que o gráfico da função de frequências relativas acumuladas F em $[\,7, 15\,[$ é um segmento de reta, que inclui o ponto $(7, F(7))$, $F(7) := F_{[3,7[} = 0.4722$, e de declive[2]

$$\frac{f_{[7,15[}}{15-7} = \frac{0.25}{8} = 0.03125$$

Naturalmente que este declive é igual ao declive do segmento definido por dois quaisquer pontos $(x, F(x))$ e $(y, F(y))$ com $x, y \in [\,7, 15\,[$,

$$\frac{F(y)-F(x)}{y-x}, \quad \forall x, y \in [\,7, 15\,[$$

Em particular, para $x = 7$ e $y = M_e$, tem-se então

$$0.03125 = \frac{F(M_e)-F(7)}{M_e-7} = \frac{0.5-0.4722}{M_e-7}$$

pelo que $M_e = 7.8896$.

6.

6.1 Para se determinar a média somam-se todos os valores tabelados e divide-se o resultado assim obtido pela dimensão da amostra, 24. Sendo a soma de todos os valores tabelados igual a 78.67, a média é então igual a $\bar{x} = 3.28$.

6.2 Com a introdução de um novo ensaio clínico, o novo valor para a média $\bar{x}_N = 3.23$ é então o resultado da soma de todos os resultados anteriores (78.67) e do resultado do novo teste, y, a dividir por 25 (a dimensão da amostra aumentou uma unidade). Ou seja,

$$3.23 = \bar{x}_N = \frac{78.67 + y}{25}$$

pelo que o resultado do 25º teste foi $y = 2.08$.

7.

7.1 Designado por x a produção de bolota (em toneladas), a média da produção de bolota é dada por

$$\bar{x} = \frac{10+17+32+35+38+48}{6} = \frac{180}{6} = 30$$

7.2 Desvio padrão da produção de bolota:

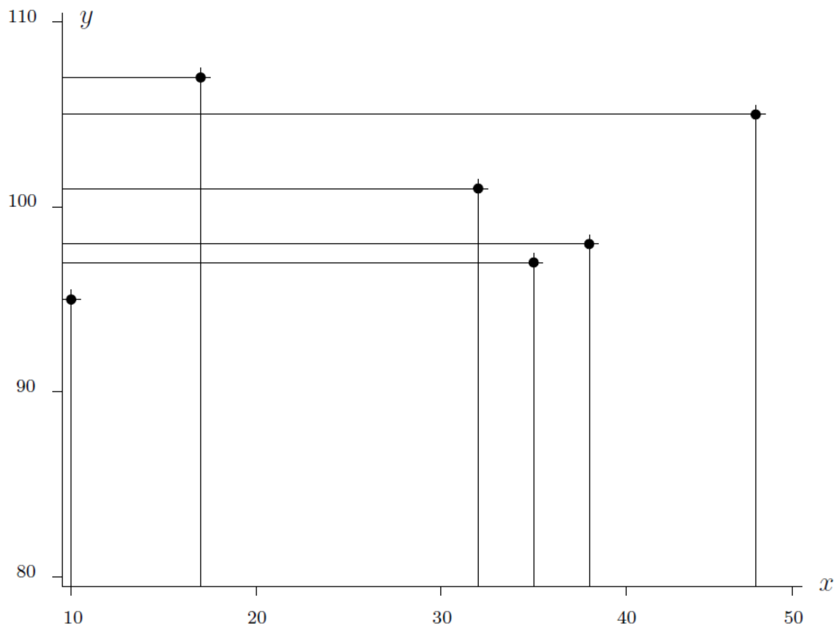
	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	10	-20	400
	17	-13	169
	32	2	4
	35	5	25
	38	8	64
	48	18	324
Totais	180		986

[2]Como a função F só está definida no intervalo $[\,7, 15\,[$, não se pode dizer que o declive da semi-recta é igual a $\frac{F(15)-F(7)}{15-7}$ Por esta razão, no numerador surge então a frequência relativa $f_{[7,15[}$ da classe $[\,7, 15\,[$.

Com estes dados calcula-se o desvio padrão:

$$s_x = \sqrt{\frac{\sum_{i=1}^{6}(x_i - \bar{x})^2}{6}} = \sqrt{\frac{986}{6}} = 12.82$$

7.3 Com os dados tabelados no enunciado obtém-se o diagrama de dispersão da figura seguinte, onde x representa a produção de bolota (em toneladas) e y o peso médio dos suinos (em quilogramas). O diagrama mostra o que parece ser uma correlação linear positiva muito fraca entre as variáveis x e y.



7.4 Assumiremos a notação da alínea anterior para as variáveis estatísticas em causa e ainda a seguinte notação para as médias e os desvios padrão:

$$\bar{x} = \frac{\sum_i x_i}{n}, \quad s_x = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}$$

$$\bar{y} = \frac{\sum_i y_i}{n}, \quad s_y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n}}, \quad n = 6$$

O coeficiente de correlação é, assim, dado por

$$r = \frac{\sum_{i=1}^{6}(x_i - \bar{x})(y_i - \bar{y})}{6 s_x s_y}$$

Note-se que o desvio padrão de x já foi calculado na alínea 7.2 e, portanto, não precisamos de voltar a repetir este cálculo. Organizando os dados e os restantes cálculos em forma de tabela tem-se a tabela seguinte.

	x_i	$x_i - \bar{x}$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
	10	-20	95	-5.5	30.25	110.0
	17	-13	107	6.5	42.25	-84.5
	32	2	101	0.5	0.25	1.0
	35	5	97	-3.5	12.25	-17.5
	38	8	98	-2.5	6.25	-20.0
	48	18	105	4.5	20.25	-81.0
Totais	180		603		111.5	70.0

Média da variável y: $\bar{y} = \frac{603}{6} = 100.5$ Kg.
O desvio padrão da variável y é $s_y = \sqrt{111.5/6} = 4.31$, concluindo-se, portanto, que

$$r = \frac{70}{6 \times 12.82 \times 4.31} = 0.21$$

Este valor traduz uma correlação linear positiva muito fraca entre as duas variáveis, confirmando o que fora concluído na alínea anterior por observação do diagrama de dispersão.

8.

8.1 Cálculos semelhantes aos realizados na alínea 7.4 conduzem aos seguintes valores para os coeficientes de correlação pedidos:

	IDH	Esperança média de vida	Média de anos de estudo	PIB per capita	Taxa Fecundidade
IDH	1	0.965302	0.939015	0.718784	-0.840703
Esperança média de vida		1	0.849489	0.525055	-0.937372
Média de anos de estudo			1	0.844305	-0.664488
PIB per capita				1	-0.290150
Taxa Fecundidade					1

8.2 Sendo o cálculo do IDH de cada país realizado com base nos indicadores expetativa de vida ao nascer, taxa de fertilidade, educação e PIB per capita, naturalmente que existe uma grande correlação entre estes indicadores e o IDH. Mas o que os valores calculados na alínea 8.1 mostram, é que existe uma forte correlação linear entre cada um destes indicadores e o IDH. À exceção da taxa de fecundidade, essa correlação é mesmo positiva. Isto significa que quanto maior (resp., menor) é a esperança de vida, a média de anos de estudo ou o PIB per capita, maior (resp., menor) é o IDH. Em sentido contrário, reflectindo a correlação linear negativa entre o IDH e a taxa de fecundidade, a maiores (resp., menores) taxas de fecundidade correspondem menores (resp., maiores) valores de IDH.
Entre os indicadores que contribuiem para o cálculo do IDH de um país, um especial destaque para os coeficientes de correlação 0.96530 entre a esperançaa de vida e o IDH, 0.939015 entre a média de anos de estudo e o IDH, ambos com um valor absoluto muito próximo de 1. Por contraste, o coeficiente de correlação entre o PIB per capita e o IDH é o que apresenta um valor absoluto mais afastado de 1.
Entre os restantes coeficientes de correlação, uma ênfase especial para a forte correlação linear negativa (-0.937372) entre a taxa de fecundidade e a esperança média de vida, o que leva a antevêr que, em países com grandes (resp., baixas) taxas de fertilidade, a esperança média de vida é menor (resp., maior).
Entre todos os coeficientes de correlação calculados, o que apresenta menor valor absoluto é o que relaciona o PIB per capita com a taxa de fecundidade. Isto significa que, embora todos os indicadores e o IDH estejam linearmente correlacionados entre si, a menor correlação linear verifica-se entre o PIB per capita e a taxa de fertilidade.

9. Recordando que o espaçoo de resultados de uma experiência aleatória é o conjunto de todos os resultados possíveis, para cada uma das experiências aleatórias do enunciado tem-se então o seguinte espaço de resultados:

9.1

$$\{(F,1),(F,2),(F,3),(F,4),(F,5),(F,6),(C,1),(C,2),(C,3),(C,4),(C,5),(C,6)\}$$

onde F e C designam, respectivamente, a face e a coroa da moeda e 1, 2, 3, 4, 5, 6 o número de pintas do dado.

9.2 Com a mesma notação da alínea anterior,

$$\{(F,F),(F,C),(C,F),(C,C)\}$$

em que em cada par ordenado a primeira coordenada é relativa ao primeiro lançamento e a segunda coordenada ao segundo lançamento.

9.3 O espaço de resultados é o mesmo que o da alínea anterior. No caso presentemente em análise, a primeira coordenada é agora relativa a uma das moedas e a segunda coordenada é relativa à outra moeda.

9.4

$$\{(i,j) : i,j = 1,2,3,4,5,6\}$$

onde em cada par ordenado a primeira coordenada diz respeito, por exemplo, ao dado vermelho e a segunda coordenada ao dado amarelo.

9.5 Tal como no caso do lançamento de duas moedas, é indiferente se o lançamento dos dados é simultâneo, ou não. Em ambos os casos o espaço de resultados é o conjunto

$$\{(i,j) : i,j = 1,2,3,4,5,6\}$$

onde, no caso dos dados serem lançaados em simultâneo, a primeira coordenada é relativa a um dos dados e a segunda coordenada ao outro dado; no caso de um dos dados ser lançaado primeiro, seguido do lançamento do outro dado, a primeira coordenada é relativa ao primeiro dado e a segunda coordenada é relativa ao segundo dado.

10.

10.1 A situação deste exercício é semelhante à do lançaamento de duas moedas (alíneas 9.2, 9.3) ou de dois dados (exercício 9.5). Independentemente das bolas serem retiradas uma a uma, ou em simultâneo, o espaço de resultados é sempre o conjunto

$$\{(a_1,a_2,a_3,a_4,a_5) : a_i \neq a_j \ \ se \ \ i \neq j, 1 \leq a_1,a_2,a_3,a_4,a_5 \leq 50\}$$

embora a interpretação de cada elemento (a_1,a_2,a_3,a_4,a_5) seja diferente. No caso das bolas serem retiradas uma após outra, cada coordenada a_i em (a_1,a_2,a_3,a_4,a_5) é relativa ao valor indicado na bola extraída na i-ésima extração ($i = 1,2,3,4,5$). Caso as bolas sejam retiradas em simultâneo, a primeira coordenada a_1 é relativa ao número inscrito numa das 5 bolas, a coordenada a_2 ao número marcado numa das restantes 4 bolas, etc.

10.2 Nesta alínea pretende-se determinar o número de elementos do espaço de resultados. Como se viu na alínea 10.1, o espaço de resultados é o mesmo, quer as cinco bolas sejam retiradas em simultâneo, quer as cinco bolas sejam retiradas uma a uma.
Para se determinar o número de elementos do espaço de resultados, observe-se que os seus elementos são da forma (a_1,a_2,a_3,a_4,a_5) com

$$a_1 \in \{1,2,\ldots,50\}, a_2 \in \{1,2,\ldots,50\} \setminus \{a_1\}, a_3 \in \{1,2,\ldots,50\} \setminus \{a_1,a_2\},$$

$$a_4 \in \{1,2,\ldots,50\} \setminus \{a_1,a_2,a_3\}\,,\, a_5 \in \{1,2,\ldots,50\} \setminus \{a_1,a_2,a_3,a_4\}$$

Isto, independentemente da interpretação dada a (a_1,a_2,a_3,a_4,a_5). É agora fácil concluir que para a_1 há 50 possibilidades, para a_2 há 49 possibilidades, para a_3 há 48 possibilidades, para a_4 há 47 possibilidades e, finalmente, para a_5 há 46 possibilidades, significando que o número total de elementos do espaço de resultados é $50 \times 49 \times 48 \times 47 \times 46$.

11.

11.1 O espaço de resultados é o indicado na alénea 10.1., agora contextualizado para o concurso do Euromilhões.

11.2 Como se verificou na alínea 10.2, existem $50\times49\times48\times47\times46$ diferentes maneiras de se retirarem 5 bolas da tômbola. Mas a pergunta é: quantas chaves diferentes é que existem?
A questão aqui é que uma chave é **um conjunto**, e não um elemento da forma $(a_1,a_2,a_3,a4,a_5)$, como os elementos do espaço de resultados. Se o concurso ditar a saída dos números pela ordem $(22,2,14,18,35)$, ou pela ordem $(2,35,22,18,14)$, a chave do Euromilhões será, em ambos os casos, a mesma: $\{2,14,18,22,35\}$.
Por outras palavras, elementos diferentes do espaço de resultados conduzem à mesma chave.
Com efeito, dados cinco números a_1,a_2,a_3,a_4,a_5, todos diferentes, todas as possíveis permutações de ordem entre os elementos de (a_1,a_2,a_3,a_4,a_5) conduzem à mesma chave: $\{a_1,a_2,a_3,a_4,a_5\}$. Para determinar quantas permutações existem, observe-se que o número, por exemplo, a_1 pode ocupar qualquer uma das cinco posições em $(.,.,.,.,.)$. Fixado este número numa posição, ficam livres quatro lugares. O número, por exemplo, a_2, pode então ocupar qualquer um destes quatro lugares livres. Completado este raciocínio, conclui-se então que há $5!$ diferentes maneiras para se localizar os námeros a_1,a_2,a_3,a_4,a_5 em $(.,.,.,.,.)$. Ou seja, há $5!$ maneiras diferentes para se obter uma mesma chave do Euromilhões.
Deste modo conclui-se que há

$$\frac{50\times49\times48\times47\times46}{5!}$$

diferentes chaves do Euromilhões.
Não é contudo surpreendente que assim seja. Com efeito,

$$\frac{50\times49\times48\times47\times46}{5!} = \frac{50!}{5!45!} = \binom{50}{5}$$

em que, por definiçãoo de coeficiente binomial, $\binom{50}{5}$ é o número de subconjuntos com 5 elementos que se podem formar a partir de um conjunto com 50 elementos. O que não é de estranhar. Como no Euromilhões tem-se um **conjunto** inicial de 50 bolas e a chave do Euromilhões é um conjunto de 5 destas 50 bolas, determinar o número de chaves diferentes no Euromilhões é equivalente a determinar o número de subconjuntos com 5 elementos de um conjunto com 50 elementos.

12. Para a resolução deste grupo de questões considerem-se os acontecimentos seguintes:

A- "'Tem experência anterior"', B - "'Tem certificado profissional"'

De acordo com o enunciado, tem-se $P(A) = \frac{40}{80}$, $P(B) = \frac{30}{80}$ e $P(A \cap B) = \frac{10}{80}$.

12.1

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{60}{80}$$

12.2 Nesta alínea pretende-se determinar

$$P\left((A \cup B) \setminus (A \cap B)\right) = P(A \cup B) - P(A \cap B)$$

que, pela introdução à resolução deste grupo e pela alínea anterior, é igual a $\frac{60}{80} - \frac{10}{80} = \frac{50}{80}$.

12.3 O enunciado diz que a candidatura seleccionada aleatoriamente tem experiência profissional. Assim, nesta alínea pretende-se determinar a probabilidade condicionada $P(B|A)$. Por definição de probabilidade condicionada, tem-se então

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{10}{40}$$

13. Como A e B são acontecimentos independentes, tem-se que

$$P(A) = P(A|B) = \frac{1}{3}$$

Por outro lado,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B)$$

em que na última igualdade se utilizou novamente a independência de A e de B. Assim,

$$\frac{7}{9} = P(A \cup B) = \frac{1}{3} + P(B) - \frac{1}{3}P(B) \Rightarrow P(B) = \frac{2}{3}$$

14. Se A e B forem independentes, então

$$P(B|A) = P(B)$$

ou seja, **se** os acontecimentos forem independentes, tem-se $P(B) = \frac{1}{5}$.
Há agora que analisar que se $P(B) = \frac{1}{5}$, **então** os acontecimentos A e B são de facto independentes. Ora,

$$P(A)P(B) = \frac{4}{5} \times \frac{1}{5} = P(A)P(B|A)$$

em que, pela definição de probabilidade condicionada, o último produto é igual a $P(A \cap B)$. Deste modo conclui-se que se $P(B) = \frac{1}{5}$, então

$$P(A \cap B) = P(A)P(B)$$

ou seja, os acontecimentos A e B são independentes.

15. Considerem-se os acontecimentos seguintes:

A - "'Existe água no subsolo"', B - "'Encontra-se água logo na 1ª tentativa"'.

De acordo com o enunciado, tem-se

$$P(A) = 0.3, \quad P(B|A) = 0.5$$

pelo que

$$P(\bar{A}) = 1 - P(A) = 0.7, \quad P(\bar{B}|A) = 1 - P(B|A) = 0.5$$

Obviamente, $P(\bar{A}|\bar{B}) = 1$.

15.1 Nesta alínea pretende-se determinar $P(\bar{B})$. Para tal, note-se que

$$\bar{B} = \bar{B} \cap (A \cup \bar{A}) = (\bar{B} \cap A) \cup (\bar{B} \cap \bar{A})$$

em que a última união é disjunta. Logo,

$$P(\bar{B}) = P\left((\bar{B} \cap A) \cup (\bar{B} \cap \bar{A})\right) = P(\bar{B} \cap A) + P(\bar{B} \cap \bar{A}) =$$

$$= P(\bar{B}|A)P(A) + P(\bar{B}|\bar{A})P(\bar{A}) = 0.85$$

15.2 Pretende-se agora calcular $P(A|\bar{B})$. Esta resulta diretamente do Teorema de Bayes:

$$P(A|\bar{B}) = \frac{P(\bar{B}|A)P(A)}{P(\bar{B})} = \frac{0.5 \times 0.3}{0.85} = 0.176$$

16.

Categoria de Peso	Sexo Masculino (%)	Sexo Feminino (%)	Total (%)
$IMC < 25$	15.4	23.3	38.7
$25 \leq IMC < 30$	21.9	14.9	36.8
$IMC \geq 30$	12.3	12.2	24.5
Total	49.6	50.4	100

16.1 Seja:
A - o acontecimento da pessoa ter um $25 \leq IMC < 30$;
B - o acontecimento da pessoa ter um $IMC \geq 30$;
Como A e B são acontecimentos mutuamente exclusivos:

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = 0.368 + 0.245 = 0.613$$

16.2 Seja:
A - o acontecimento da pessoa ter um $25 \leq IMC < 30$;
C - o acontecimento da pessoa ser do sexo masculino ;
Como A e C não são acontecimentos mutuamente exclusivos:

$$P(A \cup C) = P(A) + P(C) - P(A \cap C)$$

$$P(A \cup C) = 0.368 + 0.496 - 0.219 = 0.645$$

16.3

Categoria de Peso	Sexo Masculino (%)	Sexo Feminino (%)	Total (%)
Não obeso	38	38	76
Obeso	12	12	24
Total	50	50	100

Precisamos de verificar se existe diferença nas probabilidades condicionadas de uma pessoa ao acaso ser obesa considerando o seu sexo. Versus a probabilidade de essa pessoa ser obesa.
O - o acontecimento da pessoa ser obesa;
F - o acontecimento da pessoa ser do sexo feminino;

$$P(O|F) = \frac{P(O \cap F)}{P(F)} = \frac{12}{50} = 0.24$$
$$P(O|\bar{F}) = \frac{P(O \cap \bar{F})}{P(\bar{F})} = \frac{12}{50} = 0.24$$

A probabilidade de ser obeso é

$$P(O) = \frac{24}{100} = 0.24$$

Ou seja,

$$P(O) = P(O|F) = P(O|\bar{F})$$

O acontecimento "‘ser obeso"’ é independente dos acontecimentos "‘ser do sexo feminimo"’ (F) ou "‘ser do sexo masculino"’ ($\bar{F}$). Baseados apenas nestes dados o sexo não é um fator de risco para a obesidade.

17. Como existem 6 faces em cada dado e os dados são equilibrados, temos $6 \times 6 = 36$ possiveis resultados com igual probabilidade ($= 1/36$). A seguinte tabela mostra o espaço de resultados:

Faces	1	2	3	4	5	6
1	1-1	1-2	1-3	1-4	1-5	1-6
2	2-1	2-2	2-3	2-4	2-5	2-6
3	3-1	3-2	3-3	3-4	3-5	3-6
4	4-1	4-2	4-3	4-4	4-5	4-6
5	5-1	5-2	5-3	5-4	5-5	5-6
6	6-1	6-2	6-3	6-4	6-5	6-6

17.1 Seja:
A - o evento de obter um duplo no lançamento dos dados;
Existem 6 possiveis resultados de obter duplo:

$$P(A) = 6 \times \frac{1}{36} = 0,166(6)$$

17.2 Sejam:
B- o evento de obter 7 no lançamento dos dados;
C - o evento de obter 8 no lançamento dos dados;
No mesmo lançamento, estes dois eventos são mutamente exclusivos: $P(B \cap C) = 0$.
Então

$$P(B \cup C) = P(B) + P(C) = \frac{6}{36} + \frac{5}{36} = \frac{11}{36} = 0.305(5)$$

17.3 Sejam:

A - o evento de obter um duplo no lançamento dos dados;

C - o evento de obter 8 no lançamento dos dados;

Estes dois eventos não são mutuamente exclusivos, porque existe um resultado comum (4 em ambos os dados), que ocorre com probabilidade $1/36$.

$$P(A \cup C) = P(A) + P(C) - P(A \cap C) = \frac{6}{36} + \frac{5}{36} - \frac{1}{36} = \frac{10}{36} = 0.27(7)$$

UAb

ATIVIDADE FORMATIVA 2 - Proposta de Resolução

1.

1.1 O espaço de resultados é o conjunto de todos os valores possíveis que se podem obter quando se faz girar a roleta. Ou seja, é o conjunto

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$$

Este é o domínio da variável aleatória X, definida pelo número obtido por se fazer girar a roleta,

$$X(i) = i, \quad i \in S$$

1.2 Quando se gira a roleta só se obtém uma única casa numerada. Isto significa que

$$P(X > 17) = P(X = 18) + P(X = 19) + P(X = 20) \tag{1}$$

Por outro lado, e de acordo com o enunciado,

$$P(X = 18) = P(X = 19) = P(X = 20) \tag{2}$$

Deste modo, para se determinar a probabilidade (1), primeiro temos de calcular o valor de (2).
De acordo com o enunciado, $P(X > 14) = 4P(X < 15)$. Seja então $x = P(X < 15)$.
Como $X(S) = S$ pode ser escrito como uma união disjunta,

$$X(S) = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\} \cup \{15, 16, 17, 18, 19, 20\}$$

tem-se

$$1 = P(X \in X(S)) = P(X \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\})+$$

$$+P(X \in \{15, 16, 17, 18, 19, 20\}) = P(X < 15) + P(X > 14) = x + 4x = 5x$$

o que implica que $P(X < 15) = x = \frac{1}{5}$ e, portanto, $P(X > 14) = 4x = \frac{4}{5}$.
Obtém-se, assim,

$$\frac{4}{5} = P(X > 14) = P(X = 15) + P(X = 16) + P(X = 17) + P(X = 18)+$$

$$+P(X = 19) + P(X = 20)$$

Como, de acordo com o enunciado, as probabilidades que surgem no lado direito desta igualdade são todas iguais entre si, tem-se, ainda,

$$\frac{4}{5} = 6P(X = i)$$

para qualquer $i = 15, 16, 17, 18, 19, 20$. Deste modo

$$P(X = i) = \frac{2}{15}, \quad \imath = 15, 16, 17, 18, 19, 20$$

e, assim,

$$P(X > 17) = P(X = 18) + P(X = 19) + P(X = 20) = 3 \times \frac{2}{15} = \frac{6}{15}$$

1.3 De igual modo tem-se

$$\frac{1}{5} = P(X < 15) = \sum_{i=1}^{1} 4P(X = i) = 14P(X = i_0)$$

para qualquer $i_0 = 1, 2, \ldots, 14$, pelo que

$$P(X = i_0) = \frac{1}{70}, \quad \forall i_0 \in \{1, 2, \ldots, 14\}$$

Assim, designando por A o acontecimento "'saída de uma casa numerada com um número par"', obtém-se

$$P(A) = \sum_{k=1}^{10} P(X = 2k) = \sum_{i=1}^{7} P(X = 2k) + \sum_{i=8}^{10} P(X = 2k) = 7 \times \frac{1}{70} + 3 \times \frac{2}{15} = \frac{1}{2}$$

2.

2.1 Em relação a este exercício, observe-se que a componente aleatória está no resultado obtido pelo lançamento do par de dados.
Conhecido o número de pintas inscrito em cada um dos dados, o valor da soma do número de pintas é um resultado perfeitamente conhecido, e único. O mesmo acontece com a aplicação da regra "'noves fora"'. Ou seja, quer na soma dos valores obtidos, quer na aplicaçao da regra, não há nada de aleatório.
Isto significa que o espaço de resultados é o espaço de resultados da experiência aleatória que consiste no lançaamento de um par de dados, ou seja, o conjunto

$$\Omega := \{(i, j) : i, j = 1, 2, 3, 4, 5.6\}$$

A variável aleatória subjacente a uma determinada experiência aleatória tem por domínio o espaçoo de resultados da experiência aleatória. No caso em análise, isto significa que a variável aleatória X tem por domínio o conjunto Ω. De acordo com o enunciado, tem-se que X está definida pelo valor obtido pela aplicação da regra "'noves fora"' à soma dos pontos obtidos pelo lançamento de um par de dados. Dito de outro modo,

$$X(i, j) = \begin{cases} i + j & se \quad i + j < 9 \\ i + j - 9 & se \quad i + j \geq 9 \end{cases}$$

para qualquer par ordenado $(i, j) \in \Omega$.

2.2 Como se pode constatar por análise de cada par $(i, j) \in \Omega$, só para os pares ordenados $(1, 1), (5, 6), (6, 5)$ é que a aplicação da regra "'noves fora"' 'à soma das coordenadas dá o valor 2.

2.3 Como vimos na alínea anterior

$$X(1,1) = X(5,6) = X(6,5) = 2$$

e são só estas as 3 maneiras distintas para se obter $X = 2$. Por outro lado, num lançamento de dois dados, há 36$= \#\Omega$ resultados possíveis. Logo,

$$P(X = 2) = \frac{3}{36} = \frac{1}{12}$$

Para se calcular

$$P(X \leq 5 | X \geq 2) = \frac{P(2 \leq X \leq 5)}{P(X \geq 2)}$$

observe-se que

$$P(2 \leq X \leq 5) = \frac{n_1}{36}, \quad P(X \geq 2) = \frac{n_2}{36}$$

onde n_1 (resp., n_2) designa o número de maneiras distintas como se pode obter $2 \leq X \leq 5$ (resp., $X \geq 2$). Logo,

$$P(X \leq 5 | X \geq 2) = \frac{13}{29}$$

2.4 Em relação à variável aleatória X, observe-se que

$$X(\Omega) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$$

Isto significa que o domínio da função de probabilidade f de X é o conjunto

$$\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$$

Uma análise semelhante à realizada na alínea 2.3 para o cálculo de $P(X = 2)$ permite concluir que f, $f(x) := P(X = x)$, é definida por

x	0	1	2	3	4	5	6	7	8
$f(x) = P(X = x)$	1/9	1/12	1/12	1/12	1/12	1/9	5/36	1/6	5/36

Para responder à segunda parte da questão, comece-se por observar que o domínio de uma funçao de distribuição é sempre igual $\mathbf{R}$. Por conseguinte, o mesmo acontece com o caso particular em estudo.
Para se determinar o modo como a função de distribuição F da variável aleatória X está definida, note que, por $X(\Omega) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$, tem-se:

[•] Para valores x negativos, não há nenhum elemento $(i, j) \in \Omega$ para o qual se tenha $X(i, j) \leq x$. Ou seja, o conjunto $\{(i, j) \in \Omega : X(i, j) \leq x\}$ (abreviadamente $\{X \leq x\}$) é vazio. Logo, e pelas propriedades duma probabilidade,

$$F(x) := P(X \leq x) = P(\emptyset) = 0$$

Mas isto no caso de x ser um valor negativo.

[•] Para valores de x no intervalo $[0, 1\,[$, $\{X \leq x\} = \{X = 0\}$. Assim,

$$F(x) := P(X \leq x) = P(X = 0) = f(0) = \frac{1}{9}$$

[•] Para valores de x no intervalo $[1, 2\,[$, $\{X \leq x\} = \{X \in \{0, 1\}\}$. Logo,

$$F(x) := P(X \leq x) = P(X \in \{0, 1\}) = P(X = 0) + P(X = 1) = f(0) + f(1) = \frac{7}{36}$$

[•] Generalizando, tem-se então

$$F(x) := P(X \leq x) = \sum_{\substack{y \leq x \\ y \in \{0,1,2,3,4,5,6,7,8\}}} P(X = y) =$$

$$= \sum_{\substack{y \leq x \\ y \in \{0,1,2,3,4,5,6,7,8\}}} f(y), \forall x \in \mathbf{R}$$

Ou seja, a função de distribuição F está definida por

$$F(x) = \begin{cases} 0 & se \quad x < 0 \\ 1/9 & se \quad 0 \leq x < 1 \\ 7/36 & se \quad 1 \leq x < 2 \\ 5/18 & se \quad 2 \leq x < 3 \\ 13/36 & se \quad 3 \leq x < 4 \\ 4/9 & se \quad 4 \leq x < 5 \\ 5/9 & se \quad 5 \leq x < 6 \\ 25/36 & se \quad 6 \leq x < 7 \\ 31/36 & se \quad 7 \leq x < 8 \\ 1 & se \quad x \geq 8 \end{cases}$$

3.

3.1 Dado que a variável aleatória X assume apenas os valores -1, 0, 1 e 2, em termos da função de probabilidade f de X, $f(x) := P(X = x)$, isto significa que

$$f(-1) + f(0) + f(1) + f(2) = 2\alpha + 0.3 + 0.4 + \alpha = 1$$

ou seja, $\alpha = 0.1$.
Com efeito, por X assumir valores no conjunto $\{-1, 0, 1, 2\}$ e por este conjunto se poder escrever como uma união de conjuntos disjuntos dois a dois,

$$\{-1, 0, 1, 2\} = \bigcup_{n=-1}^{2} \{n\}$$

tem-se

$$1 = P(X \in \{-1, 0, 1, 2\}) = \sum_{n=-1}^{2} P(X = n) = \sum_{n=-1}^{2} f(n)$$

3.2 Para os vários valores pedidos:

[•] Sendo 2 uma constante, tem-se

$$E(X + 2) = E(X) + E(2) = E(X) + 2$$

com

$$E(X) = \sum_{n=-1}^{2} nP(X=n) = \sum_{n=-1}^{2} nf(n) =$$

$$= (-1)f(-1) + 0f(0) + 1f(1) + 2f(2) = 0.4$$

Assim, $E(X+2) = 0.4 + 2 = 2.4$.

[•] Tem-se

$$E(X^2) = \sum_{n=-1}^{2} n^2P(X=n) = \sum_{n=-1}^{2} n^2f(n) =$$

$$= (-1)^2f(-1) + 0^2f(0) + 1^2f(1) + 2^2f(2) = 1$$

[•] Pelo cálculo anterior tem-se $E(X^2) = 1$ e, por conseguinte,

$$P(X < E(X^2)) = P(X < 1) = P(X=-1) + P(X=0) = 0.5$$

3.3 Considere-se a variável aleatória $Y = X^2$. Atendendo aos valores assumidos pela variável aleatória X, -1, 0, 1 e 2, resulta que Y só pode assumir os valores 0, 1 e 4, tendo-se

$$P(Y=0) = P(X^2=0) = P(X=0) = 0.3$$

$$P(Y=1) = P(X^2=1) = P(X=-1) + P(X=1) = 0.6$$

$$P(Y=4) = P(X^2=4) = P(X=2) = 0.1$$

Em termos de função de probabilidade g de Y , isto significa que g está definida no conjunto $\{0,1,4\}$, tendo-se

y	0	1	4
$g(y) := P(Y=y)$	0.3	0.6	0.1

4. Considere-se a variável aleatória X que designa o número de vezes que sai o número 2 em cinco lançamentos de um dado.
Como a probabilidade de sair o número 2 mantém-se no primeiro, no segundo, no terceiro, no quarto e no quinto lançamentos, sempre igual a $p = 1/6$ e, além disto, os cinco lançamentos correspondem a cinco provas independentes, a variável aleatória X segue **uma distribuição binomial** com $n = 5$ e $p = 1/6$. Pelo conhecimento que se tem sobre a função de probabilidade correspondente a esta distribuição, conclui-se assim que a probabilidade pedida é igual a

$$P(X=2) = \binom{5}{2} p^2 (1-p)^{5-2} = \binom{5}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = \frac{10 \times 5^3}{6^5}$$

5.

5.1 Considere a variável aleatória X que designa o número de vezes que sai um berlinde defeituoso nas 20 tiragens. Como em cada tiragem existe reposição do berlinde retirado, a probabilidade de sair um berlinde defeituoso em cada tiragem é sempre igual em cada uma das 20 tiragens:

$$p = \frac{15}{80} = \frac{3}{16}$$

Acresce que, ainda devido à reposição, as 20 tiragens correspondem a 20 provas independentes. Por estas razões, a vari´avel aleatória X segue uma distribuição binomial com $n = 20$ e $p = 3/16$. Assim, a probabilidade de em 20 tiragens ser encontrado um único berlinde defeituoso é igual a

$$P(X = 1) = \begin{pmatrix} 20 \\ 1 \end{pmatrix} p^1 (1-p)^{20-1} = 20\frac{3}{16}\left(\frac{13}{16}\right)^{19} = \frac{15}{4}\left(\frac{13}{16}\right)^{19}$$

5.2 Mais geralmente, designemos por X_n a variável aleatória que designa o número de vezes que sai um berlinde defeituoso em $n \geq 1$ tiragens. Pelas mesmas razões indicadas na alínea anterior, X_n segue uma distribuição binomial com n e $p = 3/16$. Pretende-se determinar o menor valor de n tal que

$$0.4 < P(X_n = 1) = \begin{pmatrix} n \\ 1 \end{pmatrix} p (1-p)^{n-1} = n\frac{3}{16}\left(\frac{13}{16}\right)^{n-1}$$

ou seja, tal que

$$n\left(\frac{13}{16}\right)^{n} > 0.4 \times \frac{16}{3} \times \frac{13}{16} = 0.4 \times \frac{13}{3}$$

Resolvendo esta inequação por "'tentativa e erro"', tém-se

$$(n=1)\ \ \frac{13}{16} = 0.8125 < 1.7(3) = 0.4 \times \frac{13}{3}$$

$$(n=2)\ \ 2\left(\frac{13}{16}\right)^{2} = 1.3203125 < 1.7(3) = 0.4 \times \frac{13}{3}$$

$$(n=3)\ \ 3\left(\frac{13}{16}\right)^{3} = 1.609130859 < 1.7(3) = 0.4 \times \frac{13}{3}$$

$$(n=4)\ \ 4\left(\frac{13}{16}\right)^{4} = 1.743225098 > 1.7(3) = 0.4 \times \frac{13}{3}$$

pelo que serão necessárias 4 tiragens para que a probabilidade de se retirar um berlinde defeituoso seja superior a 40%.

6. Uma vez que agora os berlindes são retirados sem reposição, as várias extrações deixam de ser independentes e a probabilidade de saída de um berlinde defeituoso altera-se de extração para extração, quer por se alterar o número total de berlindes existentes na caixa, quer por eventualmente se alterar o número de berlindes defeituosos que permanecem na caixa (resultante, obviamente, de extrações anteriores de berlindes defeituosos).
Isto significa que considerando a variável aleatória Y que designa o número de berlindes defeituosos nos 20 berlindes retirados, Y segue uma **distribuição hipergeométrica** com $N = 80, M = 15$ e $n = 20$. A probabilidade pedida é então igual a

$$P(Y = 2) = \frac{\begin{pmatrix} M \\ 2 \end{pmatrix}\begin{pmatrix} N-M \\ n-2 \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} = \frac{\begin{pmatrix} 15 \\ 2 \end{pmatrix}\begin{pmatrix} 65 \\ 18 \end{pmatrix}}{\begin{pmatrix} 80 \\ 20 \end{pmatrix}}$$

7. Seja X uma variável aleatória de Poisson de parâmetro $\lambda > 0$. Neste caso, a função de probabilidade de X é dada por

$$P(X = n) = \frac{\lambda^n}{n!}e^{-\lambda}$$

sendo a função geradora de momentos Ψ_X definida por

$$\Psi_X(t) := E\left(e^{tX}\right) = \sum_{n=0}^{\infty} e^{tn}P(X = n) = e^{-\lambda}\sum_{n=0}^{\infty} e^{tn}\frac{\lambda^n}{n!} =$$

$$= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\left(\lambda e^t\right)^n}{n!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda\left(e^t-1\right)}$$

para todo o $t \geq 0$.

8. Designando por X_A, X_B e X_C as variáveis aleatórias "‘número de pessoas que se dirigiam por hora aos centros, respectivamente, A, B e C"’, decorre do enunciado que cada variável aleatória X_A, X_B, X_C tem uma distribuição de Poisson de média, respetivamente, $E(X_A) = 5$, $E(X_B) = 10$, $E(X_C) = 3$. Uma vez que o atendimento em cada um dos três centros era independente dos restantes, tem-se então que a nova variável aleatória $X = X_A + X_B + X_C$ "‘número de pessoas que se dirigem por hora ao novo centro"’ tem uma distribuição de Poisson de média $E(X) = E(X_A) + E(X_B) + E(X_C) = 18$. Assim, a probabilidade de numa hora serem atendidas 12 pessoas neste novo centro é igual a

$$P(X = 12) = \frac{18^12}{12!}e^{-18} = 0.036782$$

9. Este exercício relaciona as funções de probabilidade f_X e f_Y de duas variáveis aleatórias discretas, X e Y , com a função de probabilidade conjunta f das variáveis X e Y (ou do par aleatório (X, Y)). Para se perceber a relação entre estas noções, note-se que por Y ter valores no conjunto $\{0, 1, 2, 3, 4\}$ tem-se

$$P(X = x) = P(X = x, Y \in \{0, 1, 2, 3, 4\})$$

Por outro lado, observe-se que o conjunto $\{0, 1, 2, 3, 4\}$ pode claramente escrever-se como uma união de conjuntos disjuntos dois a dois,

$$\{0, 1, 2, 3, 4\} = \bigcup_{y=0}^{4}\{y\}$$

Isto permite concluir, por definição de probabilidade, que

$$P(X = x) = P(X = x, Y \in \{0, 1, 2, 3, 4\}) = P\left(X = x, Y \in \bigcup_{y=0}^{4}\{y\}\right) = \sum_{y=0}^{4} P\left(X = x, Y = y\right)$$

Atendendo a que a função de probabilidade conjunta f é, por definição, dada por

$$f(x, y) = P\left(X = x, Y = y\right)$$

e que a função de probabilidade f_X da variável aleatória X é, também por definição, definida por

$$f_X(x) = P(X = x)$$

obtemos assim a igualdade bem conhecida (f_X coincide com a função de probabilidade marginal de X)

$$f_X(x) = P(X = x) = P(X = x, Y \in \{0, 1, 2, 3, 4\}) = \sum_{y=0}^{4} P(X = x, Y = y) = \sum_{y=0}^{4} f(x, y) \tag{3}$$

Um raciocínio análogo permite ainda derivar a igualdade

$$f_Y(y) = \sum_{x=0}^{3} f(x, y) \tag{4}$$

9.1 Por recurso à igualdade (3), tem-se, assim,

$$0.30 = f_X(0) = \sum_{y=0}^{4} f(0, y) = 0.10 + f(0, 1) + 0.05 + 0 + 0$$

pelo que fica determinado o valor em falta na primeira linha da tabela: $f(0, 1) = 0.15$. O mesmo raciocínio permite também determinar o valor em falta na segunda linha da tabela, $f(1, 4) = 0.05$.
Utilizando agora a igualdade (4), obtém-se

$$0.40 = f_Y(0) = \sum_{x=0}^{3} f(x, 0) = 0.10 + 0 + f(2, 0) + 0.15 \Rightarrow f(2, 0) = 0.15$$

$$0.20 = f_Y(2) = \sum_{x=0}^{3} f(x, 2) = 0.05 + 0.05 + 0 + f(3, 2) \Rightarrow f(3, 2) = 0.10$$

Ficam assim a faltar três valores: $f(3, 4)$, $f(2, 1)$ e $f(2, 4)$. Para o primeiro, note-se que, por (3),

$$0.25 = f_X(3) = 0.15+0+f(3, 2)+0+f(3, 4) = 0.15+0.10+f(3, 4) \Rightarrow f(3, 4) = 0$$

e por (4)

$$0.05 = f_Y(4) = 0 + f(1, 4) + f(2, 4) + f(3, 4) = 0.05 + f(2, 4) + 0 \Rightarrow f(2, 4) = 0$$

Por fim, tem-se por (3)

$$0.25 = f_X(2) = f(2, 0)+f(2, 1)+0+0.05+f(2, 4) = 0.15+f(2, 1)+0.05 \Rightarrow f(2, 4) = 0.05$$

com o que fica completa a tabela:

$f(x, y)$	$y = 0$	$y = 1$	$y = 2$	$y = 3$	$y = 4$
$x = 0$	0.10	0.15	0.05	0	0
$x = 1$	0	0.05	0.05	0.05	0.05
$x = 2$	0.15	0.05	0	0.05	0
$x = 3$	0.15	0	0.10	0	0

9.2 Esta alínea resolve-se de forma semelhante à pergunta 3.1. Uma vez que a variável aleatória Y assume apenas os valores 0, 1, 2, 3 e 4, em termos da função de probabilidade f_Y de Y , $f_Y(y) = P(Y = y)$, isto significa que

$$f_Y(0) + f_Y(1) + f_Y(2) + f_Y(3) + f_Y(4) = 0.40 + \alpha + 0.20 + 0.10 + 0.05 = 1$$

ou seja, $\alpha = f_Y(1) = P(Y = 1) = 0.25$.

9.3 Para se determinar o valor de

$$Cov(X,Y) := E(XY) - E(X)E(Y)$$

comece-se por notar que, cálculos semelhantes aos realizados na alínea 3.2 para se determinar o valor de $E(X)$, conduzem a

$$E(X) = \sum_{x=0}^{3} x f_X(x) = 0.20 + 2 \times 0.25 + 3 \times 0.25 = 1.45$$

$$E(Y) = \sum_{y=0}^{4} y f_Y(y) = 0.25 + 2 \times 0.20 + 3 \times 0.10 + 4 \times 0.05 = 1.15$$

Por outro lado

$$E(XY) = \sum_{x=0}^{3}\sum_{y=0}^{4} xyP(X=x, Y=y) = \sum_{x=0}^{3}\sum_{y=0}^{4} xyf(x,y) =$$

$$= \sum_{x=0}^{3} (x \times 0 \times f(x,0) + x \times 1 \times f(x,1) + x \times 2 \times f(x,2) +$$

$$+ x \times 3 \times f(x,3) + x \times 4 \times f(x,4)) =$$

$$= 1 \times 1 \times f(1,1) + 2 \times 1 \times f(2,1) + 3 \times 1 \times f(3,1) +$$

$$+ 1 \times 2 \times f(1,2) + 2 \times 2 \times f(2,2) + 3 \times 2 \times f(3,2) +$$

$$+ 1 \times 3 \times f(1,3) + 2 \times 3 \times f(2,3) + 3 \times 3 \times f(3,3) +$$

$$+ 1 \times 4 \times f(1,4) + 2 \times 4 \times f(2,4) + 3 \times 4 \times f(3,4) = 1.5$$

Assim,

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 1.5 - 1.45 \times 1.15 = -0.1675$$

Sendo este valor diferente de zero, conclui-se, assim, que as variáveis X e Y não são independentes.

10.

10.1

[10.1.1] A tabela seguinte exibe apenas os pontos amostrais (x, y) para os quais a probabilidade conjunta $f(x,y) = P(X = x, Y = y)$ é igual a $c(2x + y)$. Para todos os outros casos, $f(x,y) = 0$.

	$y=0$	$y=1$	$y=2$	$y=3$	Totais
$x=0$	0	c	$2c$	$3c$	$4c$
$x=1$	$2c$	$3c$	$4c$	$5c$	$14c$
$x=2$	$4c$	$5c$	$6c$	$7c$	$22c$
Totais	$6c$	$9c$	$12c$	$15c$	$42c$

Sendo f uma função de probabilidade conjunta, o total

$$\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} f(x,y) = \sum_{x=0}^{2}\sum_{y=0}^{3} f(x,y) = 42c$$

Deverá ser igual a 1, logo, $c = 1/42$.

[10.1.2] Pela tabela conclui-se que

$$P(X = 2, Y = 1) = f(2,1) = 5c = \frac{5}{42}$$

[10.1.3] Para se determinar o valor pretendido, dos valores da tabela há que somar aqueles destacados na tabela:

	$y=0$	$y=1$	$y=2$	$y=3$	Totais
$x=0$	0	c	$2c$	$3c$	$4c$
$x=1$	**2c**	**3c**	**4c**	$5c$	$14c$
$x=2$	**4c**	**5c**	**6c**	$7c$	$22c$
Totais	$6c$	$9c$	$12c$	$15c$	$42c$

Para este processo obtém-se

$$P(X \geq 1, Y \leq 2) = \sum_{x=1}^{2}\sum_{y=0}^{2} f(x,y) = 24c = \frac{4}{7}$$

[10.1.4] A função de probabilidade marginal de X é dada por

$$f_X(x) = P(X = x) = \sum_{x=0}^{3} f(x,y), \quad x \in \mathbf{N} \cup \{0\}$$

e pode ser obtida a partir dos totais marginais da coluna da direita das tabelas anteriores. Deste modo, obtém-se

$$f_X(x) = \begin{cases} 6c = 1/7 & se \quad x = 0 \\ 14c = 1/3 & se \quad x = 1 \\ 22c = 11/21 & se \quad x = 2 \\ 0 & se \quad x \in \mathbf{N}, x \geq 3 \end{cases}$$

De modo semelhante, a função de probabilidade marginal de Y

$$f_Y(y) = P(Y = y) = \sum_{y=0}^{2} f(x,y), \quad y \in \mathbf{N} \cup \{0\}$$

pode ser obtida a partir dos totais marginais da última linha das tabelas anteriores, o que conduz a

$$f_Y(y) = \begin{cases} 6c = 1/7 & se \quad y = 0 \\ 9c = 3/14 & se \quad y = 1 \\ 12c = 2/7 & se \quad y = 2 \\ 15c = 5/14 & se \quad y = 3 \\ 0 & se \quad x \in \mathbf{N}, y \geq 4 \end{cases}$$

10.2 As variáveis X e Y são dependentes. Com efeito, pela alínea 10.1.2 tem-se

$$P(X=2, Y=1) = f(2,1) = \frac{5}{42}$$

diferente do valor do produto

$$P(X=2)P(Y=1) = f_X(2)f_Y(1) = \frac{11}{21} \times \frac{3}{14}$$

11.

11.1 Seja $p_X(x)$: função de probabilidade marginal da v.a. X.

$$\forall x \;:\; p_X(x) = P(X=x) = \sum_y p_{X,Y}(x,y)$$

e $p_Y(x)$: função de probabilidade marginal da v.a. Y.

$$\forall y \;:\; p_Y(y) = P(Y=y) = \sum_x p_{X,Y}(x,y)$$

A tabela seguinte apresenta as funções de probabilidades marginais:

$p_{X,Y}(x,y)$	$Y=1000$	$Y=2000$	$Y=3000$	$Y=4000$	$p_X(x)$
$X=1000$	0.20	0.04	0.01	0	0.25
$X=2000$	0.10	0.36	0.09	0	0.55
$X=3000$	0	0.05	0.10	0	0.15
$X=4000$	0	0	0	0.05	0.05
$p_Y(y)$	0.3	0.45	0.2	0.05	

11.2 Seja $p_{X|Y}(X=2000|Y=2000)$ a probabilidade do primeiro titular ter um rendimento mensal de 2000 euros, admitindo que o segundo titular tem um rendimento idêntico.

$$p_{X|Y}(X=2000|Y=2000) = \frac{p_{X,Y}(X=2000, Y=2000)}{p_Y(Y=2000)} =$$

$$= \frac{0.36}{0.45} = 0.89$$

11.3

$$E(X) = \sum_x x p_X(x) = 1000\times 0.25 + 2000\times 0.55 + 3000\times 0.15 + 4000\times 0.05 = 2000$$

$$E(Y) = \sum_y y p_Y(y) = 1000\times 0.30 + 2000\times 0.45 + 3000\times 0.20 + 4000\times 0.05 = 2000$$

$$Var(X) = \sum_x (X-E(X))^2 p_X(x) = (1000-2000)^2\times 0.25 + (2000-2000)^2\times 0.55 +$$

$$+(3000-2000)^2 \times 0.15 + (4000-2000)^2 \times 0.05 = 600000$$

$$\sigma_X = \sqrt{600000} = 774.6$$

$$Var(Y) = \sum_y (Y-E(Y))^2 p_Y(y) = (1000-2000)^2\times 0.30 + (2000-2000)^2\times 0.45 +$$

$$+(3000-2000)^2 \times 0.20 + (4000-2000)^2 \times 0.05 = 700000$$

$$\sigma_Y = \sqrt{700000} = 836.7$$

11.4 As variáveis X e Y são independentes se se verificarem as seguintes condições, para todos os valores de x e y:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = p_X(x)$$

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_X(x)} = p_Y(y)$$

donde,

$$\forall x, y : p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

Mas como pela alínea anterior $p_{X|Y}(X = 2000|Y = 2000) = 0.8$ e $p_X(X = 2000) = 0.55$, conclui-se que as variáveis não são independentes.

11.5

$$Cov(X,Y) = E\left[(x - E(X))\,(y - E(Y))\right] = \sum_x \sum_y (x - E(X))\,(y - E(Y))\,p_{X,Y}(x,y)$$

também se poderia utilizar

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

substituindo

$$Cov(X,Y) = (1000-2000)(1000-2000)0.20 + (1000-2000)(2000-2000)0.04 + \cdots +$$

$$+(3000-2000)(3000-2000)0.10 + (4000-2000)(4000-2000)0.05 = 490000$$

O coeficiente de correlação linear entre X e Y

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} =$$

$$= \frac{490000}{774.6 \times 836.7} = 0.756$$

11.6

$$E(R) = E\left(X + Y\right) == E(X) + E(Y) =$$

$$= 2000 + 2000 = 4000$$

$$Var(R) = Var\left(X + Y\right) = Var(X) + Var(Y) + 2Cov(X,Y) =$$

$$= 600000 + 700000 + 2 \times 490000 = 2280000$$

$$\sigma_R = \sqrt{2280000} = 1509.97$$

12. Seja Y - v.a. número de bolinhos de feijão de tamanho não conforme (abaixo do permitido) entre os 5 retirados da fornada.
Então, Y segue uma distribuição Hipergeométrica $H(Mp, M(1-p), N)$ em que p representa a proporção de bolinhos não conformes na formada ($p = 0.01$) e M o número de bolinhos que compõe a fornada ($M = 300$).

$$Y \sim H(3, 297, 5)$$

A função de probabilidade de Y é:

$$p(y) = P(Y = y) = \frac{\begin{pmatrix} 3 \\ y \end{pmatrix} \begin{pmatrix} 297 \\ 5-y \end{pmatrix}}{\begin{pmatrix} 300 \\ 5 \end{pmatrix}}$$

Pretende-se

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - p(0)$$

e

$$p(0) = P(Y = 0) = \frac{\begin{pmatrix} 3 \\ 0 \end{pmatrix} \begin{pmatrix} 297 \\ 5 \end{pmatrix}}{\begin{pmatrix} 300 \\ 5 \end{pmatrix}} = \frac{297!\ 5!\ 295!}{5!\ 300!\ 292!} = 0.951$$

$$P(Y \geq 1) = 1 - 0.951 = 0.049$$

UAb

ATIVIDADE FORMATIVA 3 - Proposta de Resolução

1.

1.1 Sendo f_X uma função densidade de probabilidade,

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

ou seja, atendendo ao modo como a função f_X está definida,

$$\int_0^1 \frac{a}{\sqrt{x}}dx = 1$$

Sendo a uma constante, tem-se

$$\int_0^1 \frac{a}{\sqrt{x}}dx = a\int_0^1 \frac{1}{\sqrt{x}}dx = \frac{a}{1/2}\left[x^{-\frac{1}{2}+1}\right]_{x=0}^{x=1} = 2a$$

pelo que $a = \frac{1}{2}$.

1.2 Tem-se

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx = \frac{1}{2}\int_0^1 \sqrt{x}dx = \frac{1}{3}\left[x^{\frac{1}{2}+1}\right]_{x=0}^{x=1} = \frac{1}{3}$$

1.3 Por definição a função de distribuição de uma variável aleatória,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$

Assim:
• Se $x \leq 0$,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \int_{-\infty}^{x} 0dy = 0$$

• Se $0 < x \leq 1$,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \int_{-\infty}^{0} f_X(y)dy + \int_{0}^{x} f_X(y)dy =$$

$$= F_X(0) + \int_0^x f_X(y)dy = \frac{1}{2}\int_0^x \frac{1}{\sqrt{x}}dy = \sqrt{x}$$

• Se $x > 1$,

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \int_{-\infty}^{1} f_X(y)dy + \int_{1}^{x} f_X(y)dy = F_X(1) + \int_1^x 0dy = F_X(1)$$

com $F_X(1) = \frac{1}{\sqrt{1}} = 1$.
Conclusão:

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \sqrt{x}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

1.4 Nesta questão pretende-se determinar

$$P\left(X_1 > 0.25, X_2 > 0.25, X_3 > 0.25\right)$$

Onde $X_i, i = 1, 2, 3$ designa-se a *i-ésima* observação da variável aleatória X. Como as observações são, por hipótese, independentes,

$$P\left(X_1 > 0.25, X_2 > 0.25, X_3 > 0.25\right) = P(X_1 > 0.25)P(X_2 > 0.25)P(X_3 > 0.25)$$

com $P(X_i > 0.25) = P(X > 0.25), i = 1, 2, 3$. Assim

$$P\left(X_1 > 0.25, X_2 > 0.25, X_3 > 0.25\right) = \left(1 - F_X(0.25)\right)^3 = \frac{1}{8}$$

2.

2.1 Seja X a variável aleatória definida pela altura dos homens portugueses em centímetros. De acordo com ao enunciado, tem-se $X \sim N(\mu, \sigma^2)$, $\mu = 176$, $\sigma = 3$.
Sendo a lei $X \sim N(\mu, \sigma^2)$ simétrica em relação a μ, isto é:

$$P(X \leq \mu - x) = P(X \geq \mu + x), \quad \forall x \in \mathbb{R}$$

resulta, para $x = 0$, que $P(X \leq 167) = 1/2$. Ou seja, a percentagem pedida é igual a 50%.

2.2 Considere-se a variável aleatória

$$Z = \frac{X - 167}{3} \sim N(0, 1)$$

Tem-se

$$P(X > 170) = P\left(\frac{X - 167}{3} > \frac{170 - 167}{3}\right) = P(Z > 1) = 1 - P(Z \leq 1)$$

Para $P(Z \leq 1) = \Phi(1)$ (Φ função de distribuição normal reduzida). De acordo com a tabela da distribuição normal, $\Phi(1) = 0.841$, pelo que $P(X > 170) = 1 - 0.841 = 0.159$.

2.3

[2.3.1] Como os homens são **independentes entre si** e, em particular, a altura entre eles, a variável aleatória Y_1 que designa o número de homens com uma altura superior a 170 cm segue uma distribuição binomial com $n = 4$ e $p = P(X > 170) = 0.159$ (cf. alínea 2.2). Logo,

$$P(Y_1 = 4) = p^4 = 0.000639129$$

[2.3.2] Considere-se a variável aleatória Y_2 que desigma o número de homens com uma altura inferior à média (167 cm). Devido à **independência** entre as alturas dos diferentes homens, Y_2 tem uma distribuição binomial com $n = 2$ e $p = P(X < 167) = P(X \leq 167) = 1/2$ (cf. alínea 2.1). Deste modo

$$P(Y_2 = 2) = \binom{4}{2} p^2(1-p)^2 = \frac{6}{2^4} = 0.375$$

3.

3.1 A função de densidade de probabilidade é uma função não negativa $f_X(x) \geq 0 \quad \forall x$, e o integral da função ao longo de todo o domínio de X é igual a 1.

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1 \Leftrightarrow \int_1^2 k(x-1)(x-2)dx = 1$$

$$\Leftrightarrow k\left[\frac{x^3}{3} - \frac{3x^2}{2} + 2x\right]_{x=1}^{x=2} = 1 \Leftrightarrow k = -6$$

Sendo $k < 0$ e $1 < x < 2$ temos $(x-1) > 0$, $(x-2) < 0$ e portanto $k(x-1)(x-2) > 0$.

3.2 O 3º Quartil, é o valor de x que acumula a probablida de 0.75.

$$\int_1^{Q_3} -6(x-1)(x-2)dx = 0.75 \Leftrightarrow -6\left[\frac{x^3}{3} - \frac{3x^2}{2} + 2x\right]_{x=1}^{x=Q_3} = 0.75$$

$$\Leftrightarrow \left(\frac{Q_3^3}{3} - \frac{3Q_3^2}{2} + 2Q_3 - \frac{5}{6}\right) = -\frac{0-75}{6} \Leftrightarrow 2Q_3^3 - 9Q_3^2 + 12Q_3 = 4.25$$

Recorrendo a um de vários softwares disponíveis online para podemos obter mais facilmente as raízes (que são 3) e escolher a que é aplicável para $1 < x < 2$.
Como $f_X(x)$ é não nula entre 1 e 2, o valor de Q_3 é em $x = 1.6726\ldots$ (as restantes raízes estão fora do intervalo).
(**Nota:** Numa prova presencial não apareceria uma equação com o presente grau de dificuldade de cálculos - 3 raízes - sem algum dado adicional).

3.3 Interessa identificar que o problema diz respeito ao cálculo de ocorrência de uma variável, chamemos S_2, que é uma soma de variáveis aleatórias independentes com distribuição Normal, correspondente à soma de 2 vendas, nos 2 países. Deve apresentar-se os respetivos parâmetros, ou seja, a média e a variância da soma. Deve apresentar-se os cálculos explicitando a passagem para a variável Normal padrão (normal reduzida) e a consulta da respetiva tabela.
Seja S_2- Montante global de venda das duas peças vendidas (em 2 países).
Assumindo que as vendas nos 2 países são independentes, pretende-se calcular $P(S_2 > 250)$

$$S_2 = Y_1 + Y_2 = \sum_{i=1}^{2} Y_i \sim N\left(\mu_2; \sigma_2^2\right)$$

com $\mu_2 = 2 \times 100 = 200$ e $\sigma_2 = \sqrt{2 \times (30)^2} = 42.43$

$$P(S_2 > 250) = 1 - P\left(Z \leq \frac{250 - \mu_2}{\sigma_2}\right) =$$

$$= 1 - P\left(Z \leq \frac{250 - 200}{42.43}\right)$$

$$= 1 - \Phi(1.178) \cong 1 - \Phi(1.18) = 1 - 0.881 = 0.119$$

4. X - representa a procura semanal do bem em toneladas e tem distribuição $N(\mu, \sigma^2)$ com $\mu = 20$ toneladas e $\sigma = 5$ toneladas. Se $X < 17 \Rightarrow$ custos de 2500; se $17 \leq X \leq 22 \Rightarrow$ custos de 2000; se $X < 22 \Rightarrow$ custos de 3000.

4.1

$$P(17 \leq X \leq 22) = P\left(\frac{17-20}{5} < \frac{X-20}{5} < \frac{22-20}{5}\right) = P(-0.6 < Z < 0.4) =$$

$$= \Phi(0.4) - \Phi(-0.6) = \Phi(0.4) - (1 - \Phi(0.6)) = 0.6554 - (1 - 0.7257) = 0.3811$$

4.2 Assumindo que a procura semanal é independente entre semanas consecutivas, a probabilidade pedida é a probabilidade de uma intersecção de acontecimentos independentes

$$P(X_1 > 25 \;\; e \;\; X_2 > 25) = P(X_1 > 25 \cap X_2 > 25) = P(X_1 > 25)P(X_2 > 25) =$$

$$= [1 - P(X \leq 25)]^2$$

onde

$$1 - P\left(Z \leq \frac{25-20}{5}\right) = 1 - \Phi(1) = 1 - 0.84134 = 0.15866$$

4.3

$$P(C = 2000) = P(17 \leq X \leq 22) = 0.3811$$

$$P(C = 2500) = P(X < 17) = P\left(\frac{X-20}{5} < \frac{17-20}{5}\right) = P(Z < -0.6) = 1 - \Phi(0.6) = 0.2743$$

$$P(C = 3000) = P(X > 22) = 1 - P(X < 22) = 1 - P(Z < 0.4) = 1 - 0.6554 = 0.3445$$

O custo C tem a seguinte distribuição de probabilidades:

Custo	2000	2500	3000
$P(C = c_i)$	0.3811	0.2743	0.3446

O valor esperado de C é

$$E[C] = 2000 \times 0.3811 + 2500 \times 0.2743 + 3000 \times 0.3446 = 2481.75$$

FIM