



# ELEMENTOS DE PROBABILIDADES E ESTATÍSTICA

## 21037

### ATIVIDADE FORMATIVA 1

#### Período de realização

Decorreu de 21 a 31 de Março de 2023

#### Proposta de Resolução da Atividade

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 omissivo 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,

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

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,

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

para a quarta linha,

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

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 determinação 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  $\frac{3 \times 400}{4} = 300$ ,  $\frac{3 \times 400}{4} + 1 = 301$  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 possuía o doutoramento. Feita a mesma análise para os inquéritos

realizados na freguesia *B* obtém-se o quadro seguinte:

	$n_i$ (Freguesia <i>A</i> )	$n_i$ (Freguesia <i>B</i> )
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.

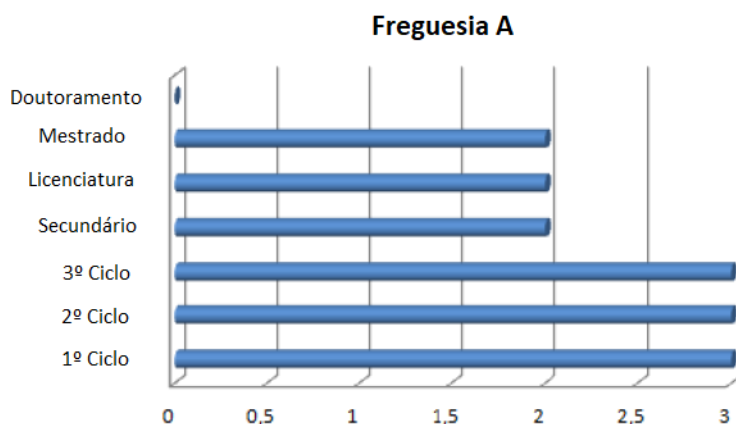
**4.3** Para cada freguesia, a amostra é de 15 elementos, um número ímpar. Ordenando os dados por ordem crescente do nível de escolaridade, o ponto central da coleção de dados corresponde ao dado com a ordem 8. Obviamente, não é necessário fazer explicitamente esta ordenação: o valor pode ser obtido 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 calcular a média.

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

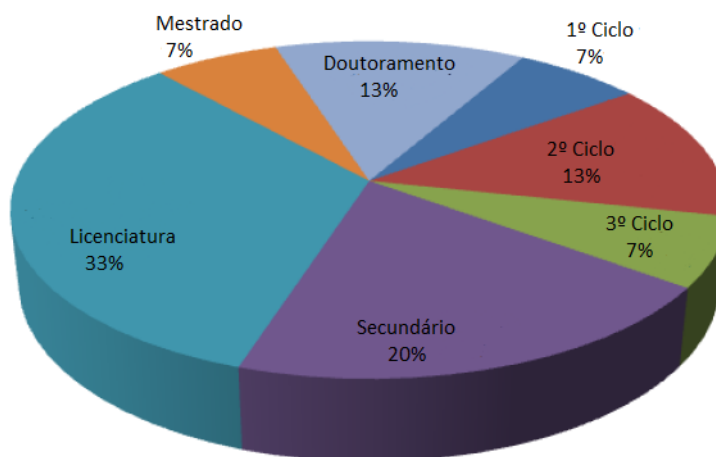


Para a freguesia *B* utilizaremos um gráfico circular ou de sectores representativo das frequências relativas:

	$n_i$	$f_i = \frac{n_i}{15}$
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  $360^\circ$ . Logo, por exemplo,  $20\% = f_{\text{secundario}}$  corresponde a uma "fatia" do círculo com um ângulo de  $72^\circ$  (regra de três simples). Como a soma das frequências relativas é igual a 100%, podemos deste modo "fatiar" todo o círculo

**Freguesia B**



**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  $360^\circ$ , então uma "fatia" do círculo com um ângulo de  $90^\circ$  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
$\geq 15$	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,

	$< 1$	$[ 1, 3 [$	$[ 3, 7 [$	$[ 7, 15 [$	$\geq 15$
$x'_i$	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
$\geq 15$	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 empresa, 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<sup>1</sup>

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

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

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

<sup>1</sup>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  $[7, 15[$ .

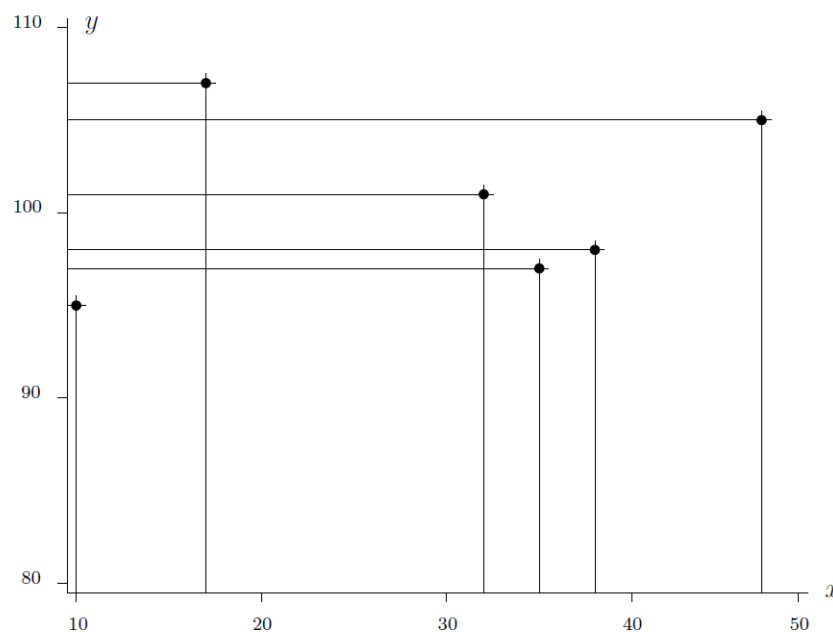
## 7.2 Desvio padrão da produção de bolota:

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

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 suínos (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

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



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 expectativa 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 contribuem para o cálculo do IDH de um país, um especial destaque para os coeficientes de correlação 0.96530 entre a esperança 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 espaço 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 presente-mente 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 lançados 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 lançado 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 lançamento 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

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

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

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

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

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 alínea 10.1., agora contextualizado para o concurso do Euromilhões.

**11.2** Como se verificou na alínea 10.2, existem  $50 \times 49 \times 48 \times 47 \times 46$  diferentes maneiras de se retirarem 5 bolas da tómbola. 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  $(a_1, a_2, a_3, a_4, a_5)$ , 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 números  $a_1, a_2, a_3, a_4, a_5$  em  $(., ., ., ., .)$ . Ou seja, há  $5!$  maneiras diferentes para se obter uma mesma chave do Euromilhões.

Deste modo conclui-se que há

$$\frac{50 \times 49 \times 48 \times 47 \times 46}{5!}$$

diferentes chaves do Euromilhões.

Não é contudo surpreendente que assim seja. Com efeito,

$$\frac{50 \times 49 \times 48 \times 47 \times 46}{5!} = \frac{50!}{5!45!} = \binom{50}{5}$$

em que, por definição 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 experiência 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

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

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,

$$\begin{aligned} P(\bar{B}) &= P((\bar{B} \cap A) \cup (\bar{B} \cap \bar{A})) = P(\bar{B} \cap A) + P(\bar{B} \cap \bar{A}) = \\ &= P(\bar{B}|A)P(A) + P(\bar{B}|\bar{A})P(\bar{A}) = 0.85 \end{aligned}$$

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 feminino" ( $F$ ) 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  $6 \times 6 = 36$  possíveis 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 possíveis 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 mutuamente 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  $1/36$ .

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