

N.º 23 – FORMA DA DISTRIBUIÇÃO DE ALGUMAS VARIÁVEIS

Por: Maria Eugénia Graça Martins
Departamento de Estatística e Investigação Operacional da FCUL
memartins@fc.ul.pt

TAREFA — Forma da distribuição de algumas variáveis tendo em conta a natureza dos dados.¹

(Adaptada de *Sheaffer, R. L. et al*, 2004, p.15)

TÓPICO - Representações gráficas.

OBJECTIVO - Desenvolver a capacidade de leitura de representações gráficas e a capacidade de esquematizar estas representações atendendo à natureza das variáveis em estudo.

¹ As situações apresentadas nesta tarefa têm um grau de dificuldade médio/elevado.

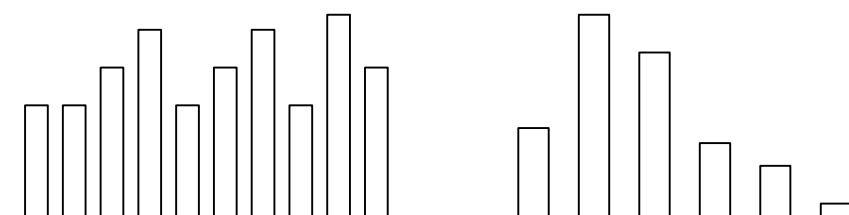


Forma da distribuição de algumas variáveis tendo em conta a natureza dos dados.

Dados – São dadas algumas representações gráficas e uma lista de variáveis e pretende-se estabelecer uma correspondência entre cada uma das representações e a variável que essa representação pretende modelar. Por exemplo, considerem-se as duas variáveis seguintes:

- Número de carros de cada um de 70 agregados familiares.
- O último dígito do número do BI de 40 alunos.

e as seguintes representações gráficas:



Pretende-se associar cada uma das representações anteriores a cada uma das variáveis indicadas.

Sabemos que o número de carros por agregado familiar deve andar à volta de 0, 1 ou 2 carros, de um modo geral. Assim, o gráfico de barras do lado direito pode representar aquela variável com as barras nos pontos 0, 1, 2, 3, 4 e 5. Quanto ao último dígito do número do BI, esperamos que a distribuição seja aproximadamente uniforme, isto é, esperamos um gráfico de barras com as barras nos pontos 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9 e com as alturas das barras aproximadamente iguais. Chamamos a atenção para que a dimensão da amostra é pequena, e nesta situação não podemos esperar um comportamento “mais uniforme”. Se recolhêssemos uma amostra de maior dimensão, esperaríamos uma maior regularidade, a chamada regularidade “a longo termo”. Pensa, por analogia, o que acontece se lançares uma moeda de um euro, equilibrada. É perfeitamente razoável em 10 lançamentos obter a seguinte sequência, onde representamos por N a face nacional e por E a face euro:

N N N E N E E N N N

Nestes 10 lançamentos, a frequência da saída da face nacional foi de 70%! Experimenta continuar a lançar a moeda muitas vezes. Vais ver que aquela frequência começa a aproximar-se de 50%.²

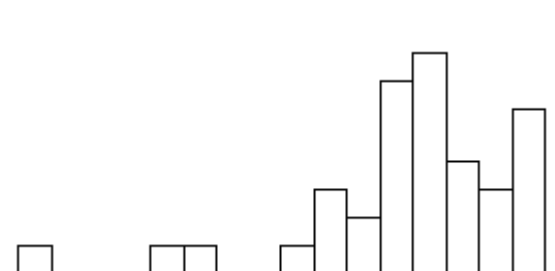
² Sugerimos uma consulta ao capítulo III do Curso de Noções de Probabilidades do ALEA, em particular às páginas 5 e 6 http://www.alea.pt/html/probabil/html/cap_03/html/cap3_1_05.html.



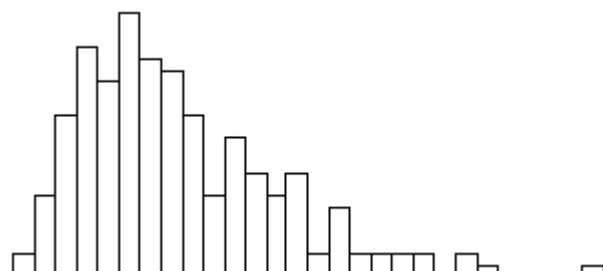
Tarefa – Considera a lista de variáveis abaixo e associa, a cada uma, uma das representações gráficas que se lhe segue:

- a) Número de medalhas ganhas pelos países medalhados nos Jogos Olímpicos de Atenas em 2004;
- b) Idade da morte de um grupo de pessoas;
- c) Notas obtidas por um conjunto de alunos num teste de Estatística, muito simples;
- d) Notas obtidas por um conjunto de alunos num teste de Estatística, equilibrado;
- e) Alturas de um grupo de alunos da faculdade;
- f) Idades em que um grupo de mães tiveram o seu primeiro filho.

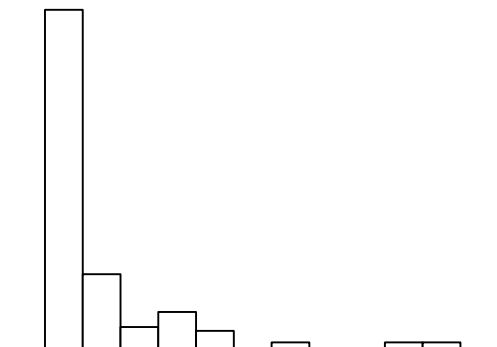
1



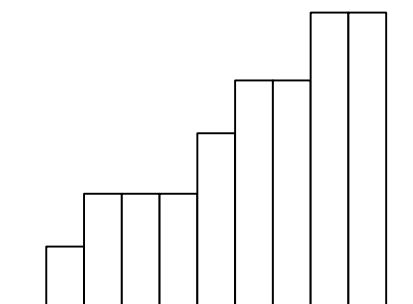
2



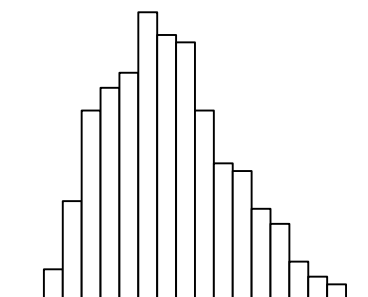
3



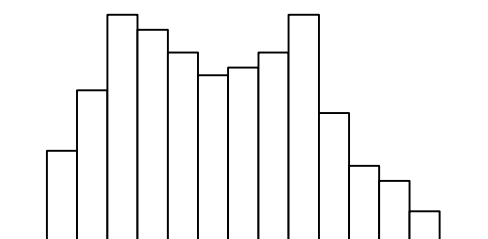
4



5



6



Variável	a	b	c	d	e	f
Gráfico						

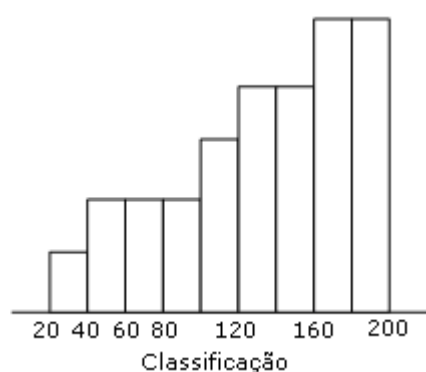
Justifica as tuas opções.



Tarefa - Forma da distribuição de algumas variáveis tendo em conta a natureza dos dados.

Resolução:

Para associar a cada conjunto de dados, de cada uma das variáveis, a representação gráfica respectiva, vamos começar por considerar as situações que nos parecem mais evidentes. Assim, começamos por associar o histograma 4. ao conjunto de notas de um teste muito simples, ou seja, da variável c), pois neste caso esperamos que as notas boas predominem, relativamente às notas más. Podemos até imaginar que a escala poderia ser a seguinte, numa classificação de 0 a 200:



Uma grande percentagem de alunos (superior a 50%) teve nota positiva e, tendo em conta o enviesamento para a esquerda, é de esperar que a média seja inferior à mediana.

Já se o teste for equilibrado, esperamos que haja uma grande percentagem de alunos numa zona central, com nota nem muito baixa, nem muito alta, e poucos alunos com nota baixa ou nota alta. Assim, sugerimos o gráfico 5. para representar os dados da alínea d).

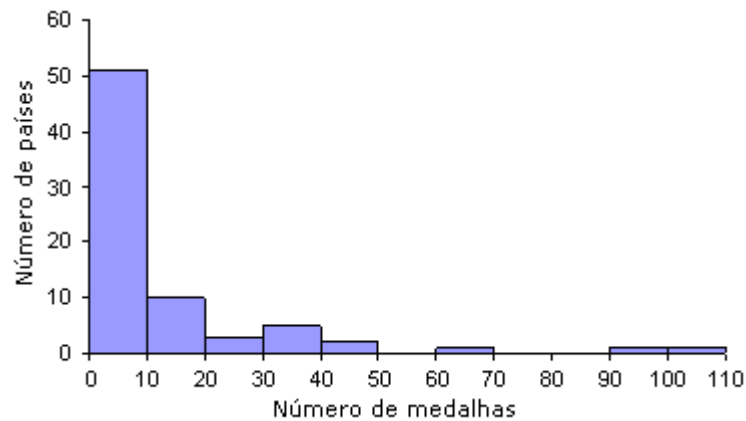
De seguida, o gráfico 6. sugere-nos uma mistura de duas populações, situação típica de dados de alturas de adultos, em que se podem considerar 2 subpopulações, a das mulheres e a dos homens, em que, de um modo geral, as mulheres têm alturas um pouco mais baixas que os homens. Numa situação de mistura de duas populações, como a que acabámos de descrever, a bimodalidade será tanto mais evidente quanto mais afastadas estiverem as zonas centrais de cada das distribuições e se a percentagem de indivíduos de cada uma das subpopulações no conjunto de dados considerados for aproximadamente igual. Assim associamos o gráfico 6 aos dados da alínea e).

Quanto ao número de medalhas por cada país nos Jogos Olímpicos, é sabido que a maior parte dos países não ganha qualquer medalha ou ganha muito poucas e só 2 ou 3 países ganham um número razoável de medalhas. Assim, o gráfico que melhor descreve esta situação é o gráfico 3. Para atribuir valores às classes consideradas, fomos pesquisar o que se passou efectivamente nos Jogos Olímpicos de Atenas³ e concluímos o seguinte: foram ganhas 929 medalhas, por 74 países. Embora estejamos perante um conjunto de dados discretos, o número de valores distintos é grande, pelo que é aconselhável organizar os

³(http://pt.wikipedia.org/wiki/Anexo:Quadro_de_medalhas_dos_Jogos_Ol%C3%ADmpicos_de_Ver%C3%A3o_de_2004)



dados em classes. Considerámos classes de amplitude 10, tendo-se obtido o seguinte histograma (idêntico ao gráfico 3).



A título de curiosidade, verificámos que o país que ganhou mais medalhas foi os EUA, com 103 medalhas, seguindo-se a Rússia com 92 e a China com 63.

Considerando os dados relativos à variável "Idade da morte de um grupo de pessoas", não temos dúvidas em associar o gráfico 1, pois este gráfico reflecte a situação em que poucos indivíduos morrem cedo, verificando-se uma maior frequência de mortes nas idades mais avançadas, como seria de esperar.

Ao contrário da variável analisada no parágrafo anterior, esperamos que o gráfico associado aos dados da variável *Idade da mãe quando teve o primeiro filho* tenha um enviesamento para a direita, característico de uma grande frequência à volta de uma idade referente a um adulto jovem, e uma pequena frequência para as idades mais avançadas. Assim, o gráfico 2 poderá ser uma boa representação dos dados desta variável.

Em conclusão, temos a seguinte tabela:

Variável	a	b	c	d	e	f
Gráfico	3	1	4	5	6	2