

AULA AbERTA

Introdução à Estatística

Estatística Descritiva com R

Teresa A. Oliveira¹

Amílcar Oliveira¹

Nuno Sousa²

Secção de Matemática¹

Secção de Informática, Física e Tecnologia²

Departamento de Ciências e Tecnologia



UNIVERSIDADE
AbERTA
www.uab.pt

2017

UNIVERSIDADE
AbERTA¹
www.uab.pt



Conteúdo

1 Revisão de Conceitos Básicos

1.1	Introdução ao curso	7
1.2	População e amostra	8
1.3	Classificação de variáveis.....	9

2 Abordagem ao software R

2.1	Breve abordagem ao software R	12
2.2	Instalação do R	12
2.3	Mais sobre o R	13

3 Estatística descritiva com R

3.1	Organização de dados.....	17
3.2	Representação gráfica dos dados.....	20
3.3	Redução dos dados: Média e desvio padrão.....	25

4 Exercícios

Exercícios Propostos.....	30
---------------------------	----

5 Resumindo...

5.1	O que vos foi dito	32
5.2	O que não vos foi dito	33



Enquadramento

Esta Aula Aberta enquadra-se numa iniciativa da Universidade Aberta de oferecer formação aberta à comunidade em geral. Nessa perspectiva a Secção de Matemática oferece três cursos distintos neste regime aberto, nas áreas de Análise Matemática e Análise Numérica, Geometria e Estatística. Estes cursos, por serem de oferta aberta, são apenas pequenas introduções a cada uma das áreas, devendo para uma abordagem mais profunda ser considerada a oferta online da secção disponível em

<http://www.uab.pt/web/guest/organizacao/departamentos/dcet/sm/cursos>

Nesse sentido, para este curso de Introdução à Estatística - Estatística descritiva com R, é conveniente ter conhecimento prévio de algumas noções básicas de estatística e de algumas metodologias, ao nível do ensino secundário, nomeadamente a noção de média e de desvio padrão, a construção de tabelas de frequências e a elaboração dos respetivos diagramas de representação dos dados.

Além deste texto, que servirá de guião ao curso, serão disponibilizados recursos multimédia para auxílio à aprendizagem, nomeadamente um vídeo sobre a apresentação e contexto desta aula aberta, um vídeo de Noções Básicas de Estatística e outro de Estatística descritiva com R.

Capítulo 1

Revisão de Conceitos Básicos

1.1 Introdução ao curso

Sejam bem vindos à Aula Aberta de Introdução à Estatística, onde apresentaremos uma breve abordagem à Estatística Descritiva com R. Esta aula aberta conta com a participação dos docentes Teresa Oliveira, Amílcar Oliveira e Nuno Sousa.

Vamos em primeiro lugar procurar uma definição para o conceito de Estatística, que surge muitas vezes no dia-a-dia com conotações diferentes de acordo com o contexto, o que pode conduzir o cidadão comum a absorver uma ideia demasiado simplista daquilo que é efetivamente a Estatística.

Por exemplo na comunicação social é comum a tendência para referirem a Estatística como uma simples informação numérica, por exemplo os resultados das sondagens em percentagens.

Numa visão mais alargada, a Estatística é vista como um conjunto de técnicas para tratamento e análise de dados, mas ainda assim estamos longe do seu significado mais amplo.

Atualmente é consensual falarmos em Estatística como sendo simultaneamente uma ciência e uma arte que permite obter conclusões sobre um conjunto de dados com base na informação de uma ou mais amostras retiradas da população em estudo. Na primeira semana estudaremos estes e outros conceitos básicos e na segunda semana estes conceitos serão abordados com recurso ao software R.

1.2 População e amostra

Os conceitos de amostra e de população surgem em inúmeras situações onde se utilizam técnicas estatísticas envolvendo análises sobre conjuntos de indivíduos. É então conveniente que comecemos por rever a definição destes conceitos.

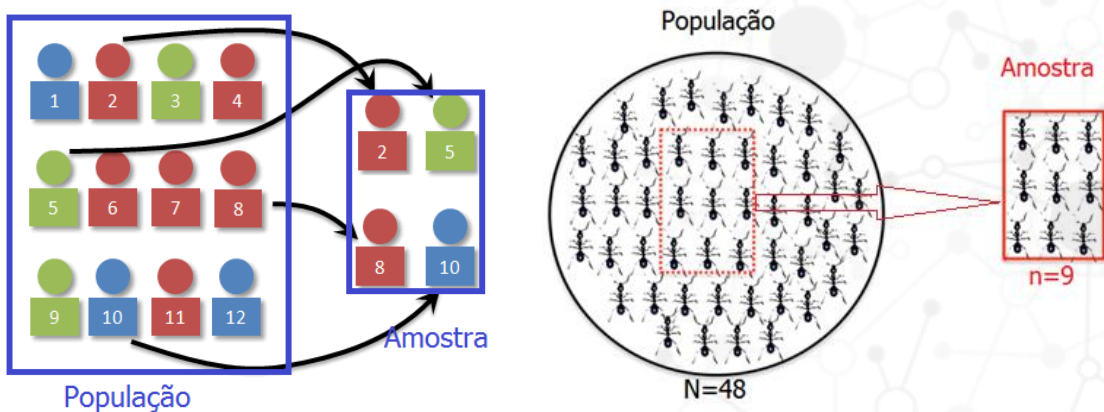
O que é uma População?

População é uma coleção de elementos individuais, por exemplo, animais, pessoas, objectos, ou resultados experimentais, que tenham uma ou mais características em comum.

E o que é uma amostra? Uma amostra é um subconjunto de elementos individuais da população. Exemplos de População:

- Estudantes da Universidade Aberta
- Pessoas com mais de 18 anos residentes em Portugal Continental
- Conjunto de formigas na região de Lisboa
- Conjunto de PCs numa determinada empresa

Ilustração de População/Amostra:



Num contexto mais real e adequado a esta época, podemos ilustrar estes conceitos com a figura que se segue, onde se observa uma amostra de amêndoas com tamanho 4, retirada a partir de uma população de tamanho N desconhecido:



1.3 Classificação de variáveis

Sobre as populações poderemos analisar as mais variadas características, consoante a natureza das mesmas e o objetivo do estudo. São exemplos dessas características:

- idade (em anos) dos estudantes da Universidade Aberta;
- situação profissional das pessoas com mais de 18 anos residentes em Portugal Continental;
- capacidade de memória dos PCs numa determinada empresa.

Assim, de acordo com o âmbito de cada estudo, para além da identificação da população e da amostra, é crucial identificar e classificar as características com interesse, designadas por variáveis uma vez que variam:

- de indivíduo para indivíduo, por exemplo a altura e o número de acessos diários à Web;
- de objeto para objeto, por exemplo a resistência e a cor.

As variáveis podem ser classificadas:

- quanto ao caráter, conforme exprimam qualidades ou quantidades:
 - qualitativo: referente a qualidades (por exemplo a cor dos olhos)
 - quantitativo: referente a quantidades mensuráveis (por exemplo a altura)
- quanto ao tipo, podem ser do tipo discreto ou do tipo contínuo, conforme se exprimam numa escala discreta (assumindo valores concretos) ou contínua (assumindo valores em intervalos), respetivamente.

Passamos agora a alguns exercícios para os quais gostaríamos que verificasse se faz a escolha correta.

1. Classifique as seguintes séries estatísticas de acordo com o caráter:

Série A: “Classificação de 8 parafusos de acordo com a calibragem”:
perfeito, defeituoso, defeituoso, perfeito, perfeito, perfeito, perfeito, perfeito

Série B: “ Número de gravações efetuadas na *pen-drive* por mês, por um determinado aluno do Curso de Matemática e Aplicações, ao longo de um ano letivo”:

7, 8, 12, 5, 8, 10, 3, 5, 8, 8, 8,10

2. Classifique as seguintes séries estatísticas de acordo com o tipo:

Série C: “Número registado de faltas ao emprego por motivo de doença na empresa XPTO em março de 2017”:

Nºde faltas	0	1	2	3	4	5
Nºde funcionários	25	8	10	6	5	2

Série D: “Classificações obtidas por 230 candidatos a uma prova de seleção de um jogo de cálculo mental”:

Classificação	Nº de candidatos
$[0,10[$	7
$[10,20[$	11
$[20,30[$	18
$[30,40[$	9
$[40,50[$	41
$[50,60[$	37
$[60,70[$	52
$[70,80[$	25
$[80,90[$	22
$[90,100[$	8

Respostas

Série A – Carater Qualitativo; Série B – Carater Quantitativo; Série C – Tipo Discreto; Série D – Tipo Contínuo

Capítulo 2

Abordagem ao software R

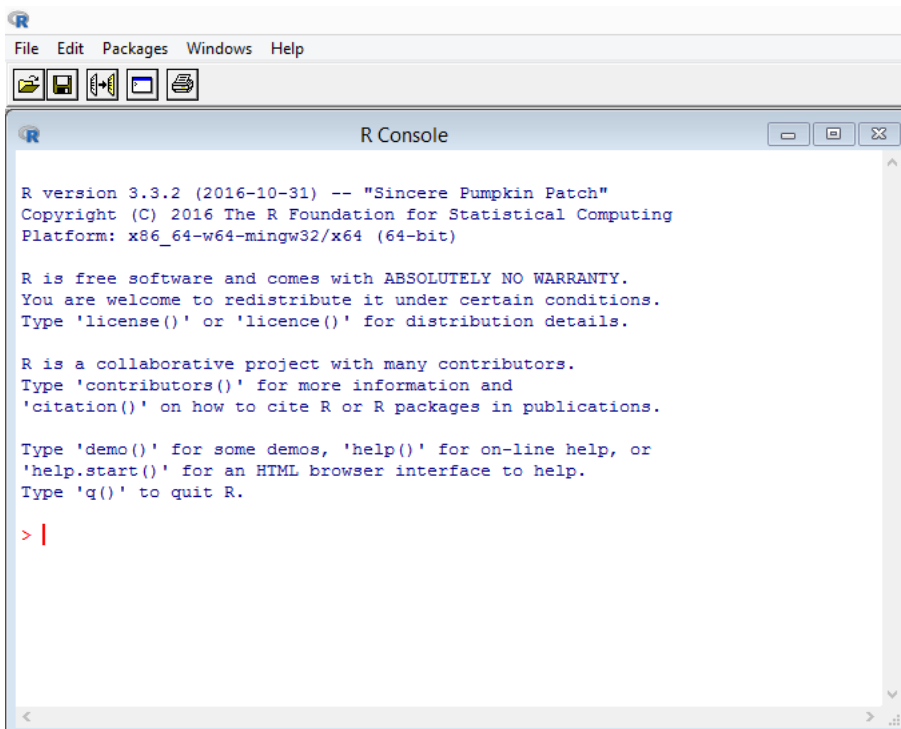
2.1 Introdução ao ambiente e software R

Sejam bem vindos ao mundo do R. O R não é simplesmente um software. O R é um ambiente com uma linguagem própria para computação estatística e visualização gráfica, que permite desenvolver um método de pesquisa específico e realizar estudo ímpares, tendo em conta a sua permanente contribuição, atualização, validação e envolvimento de especialistas a nível internacional. Os códigos e os programas desenvolvidos em R permitem não só conduzir e partilhar experiências, mas também trabalhar com dados que podem ser reais ou simulados. O R destaca-se pelo seu potencial e pela possibilidade de controlar diretamente os cálculos, pela sua flexibilidade e simplicidade, e pelo acesso livre à comunidade científica. A sua versão base faz praticamente tudo o que é elementar e muito do que é avançado, e tem suplementos disponíveis para todo o tipo de temáticas mais específicas.

2.2. Instalação do R

O R pode ser descarregado de <http://cran.r-project.org/> para os três sistemas operativos mais usados: Windows, Mac OS, Linux (várias distribuições). Para fazer a descarga e instalação do R bastará seguir as instruções na página-mãe do CRAN. Em caso de dificuldades, o participante pode e deve contactar os professores pela plataforma.

A instalação base do CRAN é do tipo linha-de-comandos, i.e. os comandos são dados por escrita e não recorrendo a menus. Eis o aspeto da linha-de-comandos quando trabalhamos no sistema operativo Windows:



```
R version 3.3.2 (2016-10-31) -- "Sincere Pumpkin Patch"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

No que se segue, assumimos que o participante já está familiarizado com a terminologia estatística entretanto abordada nesta Aula Aberta.

Recomenda-se também que o participante tenha a sua instalação do R efectuada e aberta, que vá introduzindo os dados e reproduzindo os comandos indicados.

2.3 Entrada de dados no R

Existem essencialmente duas formas de entrar dados no R: manualmente, através da linha-de-comandos, ou carregando um ficheiro externo. Neste documento vamos abordar apenas a entrada manual de dados.

No R, os dados são guardados sob a forma de *vetores*, ou seja, listas ordenadas de valores. Para exemplificar a entrada de dados, vamos recorrer a um exemplo. Suponhamos que queremos estudar o número de vezes que cada uma de um conjunto de 12 pessoas foi ao supermercado no último mês.

Os dados recolhidos foram os seguintes:

Nr.	Pessoa	Idas ao supermercado
1	Abel	6
2	Rafael	3
3	Miguel	4
4	Gabriel	5
5	Ismael	4
6	Isabel	1
7	Samuel	1
8	Daniel	2
9	Raquel	4
10	Manuel	3
11	Ezequiel	7
12	Joel	2

Para entrar estes dados, basta escrever na linha-de-comandos:

```
> idas <- c(6, 3, 4, 5, 4, 1, 1, 2, 4, 3, 7, 2)
```

Fica assim criado o objeto "idas", que é neste caso um vetor com os 12 valores indicados. O comando "c" significa: "combinar a sequência de valores que se segue num vetor". Para vermos o conteúdo que `idas` tem, basta escrever `idas` na linha-de-comandos seguido de *enter*:

```
> idas
[1] 6 3 4 5 4 1 1 2 4 3 7 2
```

O número 1 entre parêntesis, [1], significa apenas que essa é a 1ª linha de output de dados. Pode-se também juntar o nome da pessoa a `idas`, mas isso não será relevante para o que se segue.

Se tiver havido um erro na entrada de dados, podemos substituir individualmente o elemento errado. Por exemplo, se o Ezequiel tivesse ido 6 vezes ao supermercado (em vez de 7), essa alteração seria feita com:

```
> idas[11] = 6
> idas
[1] 6 3 4 5 4 1 1 2 4 3 6 2
```

O comando `idas[11] = 6` significa, em linguagem corrente, “mude o 11º elemento de `idas` para 6”. Também se pode sempre voltar a entrar todo o vetor `idas`, o que é mais fácil usando as setas “up/down” para navegar na lista de comandos entrados anteriormente e, encontrado o comando `idas = c(...,` substituir o valor 7 por 6. No restante deste documento, vamos assumir que os dados são introduzidos de forma manual e que `idas[11]` têm valor original 7. O número de idas ao supermercado é uma variável estatística do tipo discreto, i.e. uma variável que só pode tomar uma gama fixa de valores, usualmente inteiros. Para o caso de variáveis contínuas, a entrada de dados é feita exatamente da mesma forma, mas os comandos R para visualização de dados são diferentes. Neste texto vamos abordar apenas o caso das variáveis discretas, que é o mais simples de tratar.

Exemplo: Introduza, recorrendo à linha-de-comandos do R, os seguintes dados, relativos ao número de irmãos que 40 pessoas têm:

Número de Irmãos

0, 1, 2, 0, 0, 2, 4, 1, 2, 3, 2, 1, 1, 1, 1, 0, 0, 0, 1, 2,
3, 4, 2, 2, 1, 1, 0, 1, 0, 1, 0, 1, 0, 2, 1, 1, 2, 0, 1, 1

(Atenção que o R não permite acentos nos nomes das variáveis, pelo que deverá dar à variável um nome sem acentos, como por exemplo:)

```
> irmaos <- c(0,1,2,0,0,...
```

As variáveis `idas` e `irmaos` vão servir-nos de ferramenta de trabalho para o resto destas breves notas.

O próximo passo será representar graficamente estas variáveis e obter medidas de redução, como a média e o desvio-padrão.

Capítulo 3

Estatística Descritiva com R

3.1 Organização de dados no R

A tabela de frequências é um instrumento usual em estatística e é a forma mais simples de organizar e de fazer a “contabilidade” dos dados recolhidos. Como dito acima, apenas abordaremos o caso mais simples, as variáveis discretas.

O comando do R `table()`, devolve um primeiro esboço dos valores da tabela, nesta caso as frequências absolutas simples resultantes das contagens efectuadas.

```
> table(idas) idas
 1 2 3 4 5 6 7
 2 2 2 3 1 1 1
```

Ou seja: “2 pessoas foram 1 vez ao supermercado.” “2 pessoas foram 2 vezes ao supermercado.” “2 pessoas foram 3 vezes ao supermercado.” “3 pessoas foram 4 vezes ao supermercado.” “1 pessoa foi 5 vezes ao supermercado.” “1 pessoas foi 6 vezes ao supermercado.” “1 pessoa foi 7 vezes ao supermercado.” Para organizar esta informação de uma forma de mais fácil leitura, podemos utilizar o comando `transform()`:

```
> transform(table(idas))
  idas Freq
 1  1     2
 2  2     2
 3  3     2
 4  4     3
 5  5     1
 6  6     1
 7  7     1
```

A coluna `Freq` pode ser passada a outros argumentos de `transform`. Por exemplo se quisermos escrever uma terceira coluna, com as frequências relativas simples, bastará fazer

```
> transform(table(idas), FreqRel=Freq/12)
```

idas	Freq	FreqRel	
1	1	2	0.16666667
2	2	2	0.16666667
3	3	2	0.16666667
4	4	3	0.25000000
5	5	1	0.08333333
6	6	2	0.08333333
7	7	2	0.08333333

Ou, em alternativa podemos obter as frequências relativas simples usando o comando `prop.table()`

```
> transform(prop.table(table(idas)))
```

idas	Freq	
1	1	0.16666667
2	2	0.16666667
3	3	0.16666667
4	4	0.25000000
5	5	0.08333333
6	6	0.08333333
8	7	0.08333333

Recorde que foram 12 pessoas a indicar quantas vezes foram ao supermercado. É necessário dar um nome, por exemplo `FreqRel`, ao cabeçalho da coluna, caso contrário o comando `transform()` simplesmente não imprimirá essa coluna. Experimente!

De forma semelhante, usando o comando `cumsum()` podemos obter as frequências absolutas acumuladas:

```
> cumsum(table(idas))
```

1	2	3	4	5	6	7
2	4	6	9	10	11	12

e também as frequências relativas acumuladas:

```
> cumsum(prop.table(table(idas)))  
  1      2      3      4      5      6      7  
0.1666667 0.3333333 0.5000000 0.7500000 0.8333333 0.9166667 1.0000000
```

O R possui uma linguagem muito flexível e existem muitas outras formas de obter tabelas e de as formatar a gosto. A que apresentámos aqui é a mais simples e permite *copy-paste* para outras ferramentas como por exemplo uma tabela de processador de texto ou folha de cálculo (Word, Excel, Pages, Numbers, etc.).

Exercício proposto: obtenha a tabela de frequências para a variável `irmaos`. Recorde que neste caso a amostra é de 40 pessoas, e não 12.

3.2 Visualização de dados usando o R

Tratada a questão da construção das tabelas de frequências, vejamos como construir alguns gráficos que representem os dados. Dois dos tipos de gráfico mais usados para variáveis discretas são os diagramas de barras e os diagramas circulares. Para variáveis contínuas o diagrama de barras é substituído pelo histograma.

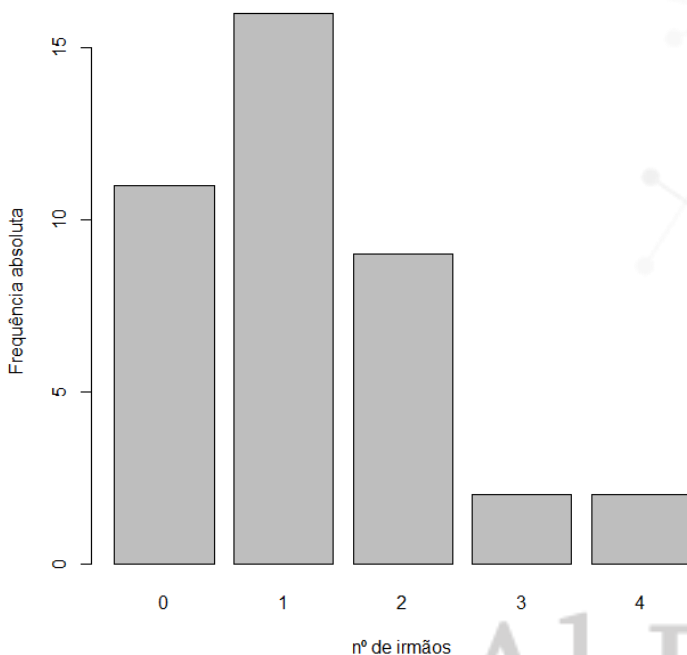
Diagrama de barras

Começemos pelo diagrama de barras. O comando R para fazer um tal diagrama é `barplot()`. No caso de pretendermos a representação dos dados referentes ao exemplo do número de irmãos, usamos:

```
> table(irmaos)
irmaos
 0  1  2  3  4
11 16  9  2  2

> barplot(table(irmaos))

> barplot(table(irmaos), xlab="nº de irmãos", ylab="Frequência absoluta")
```



O diagrama, conforme descrito, é algo escasso em informação. Para o enriquecer recorre-se aos argumentos que o comando `barplot()` permite utilizar. Por exemplo:

<code>space</code>	Controla o espaçamento entre barras.
<code>names.arg</code>	Permite substituir o nome das barras (0,1,2,3,4) por outra designação que se pretenda.

Por exemplo, para colocar zero, um, dois, etc., escreve-se:

```
names.arg=c("zero", "um", "dois", "tres", "quatro").
```

As aspas significam que o que está entre elas é texto, e são essenciais. Caso contrário o R irá interpretar os nomes como variáveis, cujo valor não estará definido (e dará erro). Evitar também colocar acentos no texto, dado que o tratamento correto destes depende um pouco da versão utilizada do R e pode nalguns casos gerar problemas.

<code>main</code>	Dá um título ao diagrama.
<code>xlab, ylab</code>	Referencia os eixos horizontal e vertical.
<code>cex.axis</code>	Controla o tamanho dos números no eixo vertical.
<code>cex.names</code>	Controla o tamanho dos números/letras no eixo horizontal.

Estes são apenas alguns dos argumentos que podem ser utilizados, mas o R dispõe de uma grande versatilidade de opções, como por exemplo para as cores, para os sombreados das barras, etc.. Podemos sempre utilizar a ajuda do R utilizando `help()` no R.

Em <http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf> existe também uma vasta lista de nomes de cores.

Vejamos um exemplo atribuindo valores a todos os parâmetros descritos:

```
> barplot(table(idas), space = 0, names.arg =  
c("um", "dois", "tres", "quatro", "cinco", "seis", "sete"), main = "Idas  
ao supermercado no ultimo mês \n do grupo de pessoas em estudo",  
xlab = "Numero de idas", ylab = "Frequencia absoluta", cex.axis =  
1.5, cex.names = 1.2)
```

Notar que o símbolo $\backslash n$ no título do gráfico, é usado para obrigar a uma mudança de linha.

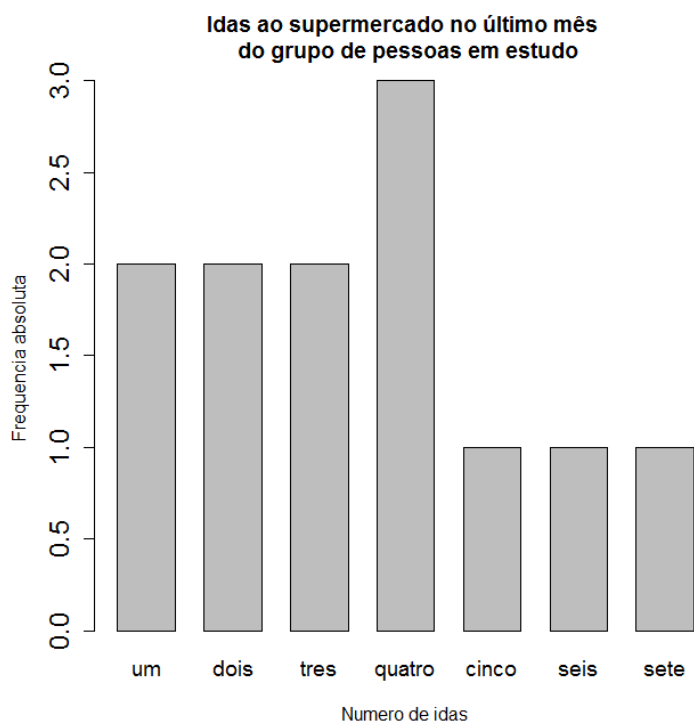


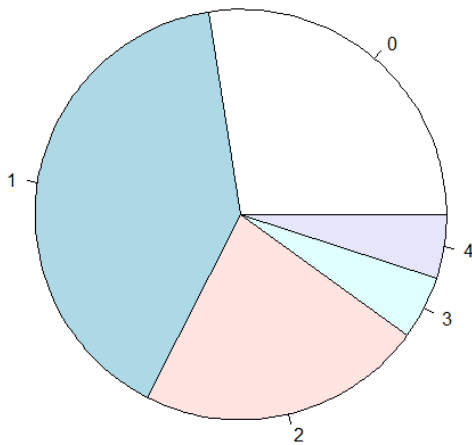
Diagrama circular

O comando utilizado na construção de diagramas circulares é `pie()` ("piechart").

Utilizando os dados referidos anteriormente, teremos

```
> pie(table(irmaos))
```

cujo resultado é:

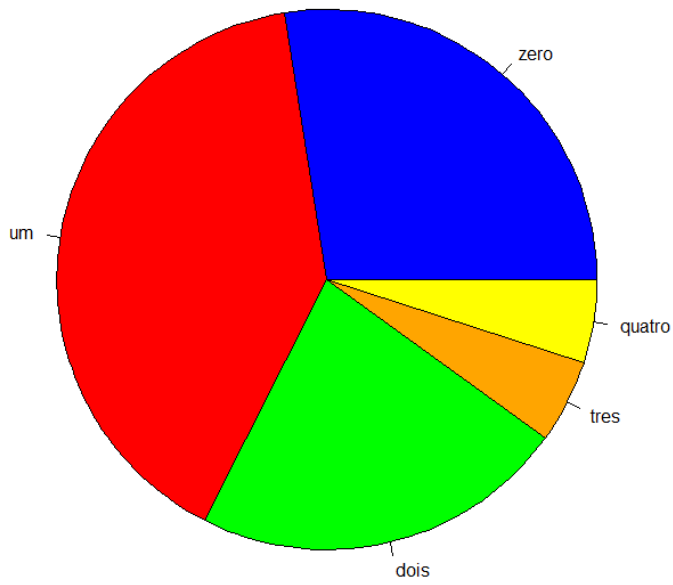


Também aqui se pode "enriquecer" o gráfico utilizando por exemplo:

```
> pie(table(irmaos), labels =  
c("zero", "um", "dois", "tres", "quatro"), radius = 1.05,  
col=c("blue1", "red", "green", "orange", "yellow"), main =  
"Distribuição do número de irmãos de 40 pessoas")
```

obtendo-se neste caso o seguinte resultado:

Distribuição do número de irmãos de 40 pessoas



Exercício: obtenha o diagrama circular para a variável idas.

3.3 Redução de dados: média aritmética e desvio-padrão

A média aritmética e o desvio-padrão são duas grandezas, que resumem num número apenas, informação sobre os dados recolhidos. A média aritmética é uma medida de localização, no sentido em que nos diz onde está localizada uma determinada característica dos dados (já vamos ver qual), e o desvio-padrão é uma medida de dispersão, uma vez que nos indica o quão dispersos os dados estão.

Média aritmética

Todos nós temos uma noção intuitiva do que é uma média. Matematicamente, a média aritmética é o valor ao qual os dados mais se aproximam e é representada quase universalmente pelo símbolo \bar{x} . A média aritmética de uma amostra é uma medida de localização obtida através do quociente entre a soma de todos os valores observados e o número total de observações. Trata-se da medida de tendência central mais utilizada e de mais fácil interpretação.

Se representarmos as N observações por x_1, x_2, \dots, x_N será:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

onde,

x_i são os valores individuais observados

N é a dimensão da população (ou amostra, utilizando-se nesse caso n em vez de N).

Caso os dados sejam do tipo discreto e estejam agrupadas (há valores que se repetem na amostra ou população), podemos obter a média com base na expressão:

$$\bar{x} = \frac{F_1x_1 + F_2x_2 + \dots + F_kx_k}{N} = \frac{1}{N} \sum_{i=1}^k F_i x_i$$

onde,

F_i é a frequência absoluta da característica x_i

k é o número de classes do agrupamento

No R, a média aritmética é obtida através do comando `mean()`:

```
> mean(idas)
```

```
[1] 3.5
```

Por vezes na literatura usam-se outras expressões para o cálculo da média (média ponderada, média para dados agrupados em classes, etc.). Essas expressões referem-se a situações em que os dados estão agregados em categorias ou classes. Quando os dados estão desagregados, ou seja, quando são considerados um-a-um, que é como os entramos no R, a fórmula acima é suficiente.

Desvio-padrão estimado

O desvio-padrão estimado a partir de uma amostra representa-se pelo símbolo "s" e é dado pela expressão:

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

A fórmula é um pouco mais complicada de entender que a da média, mas no R basta usar o comando `sd()` para o obter o seu valor sem esforço:

```
> sd(idas)
```

```
[1] 1.882938
```

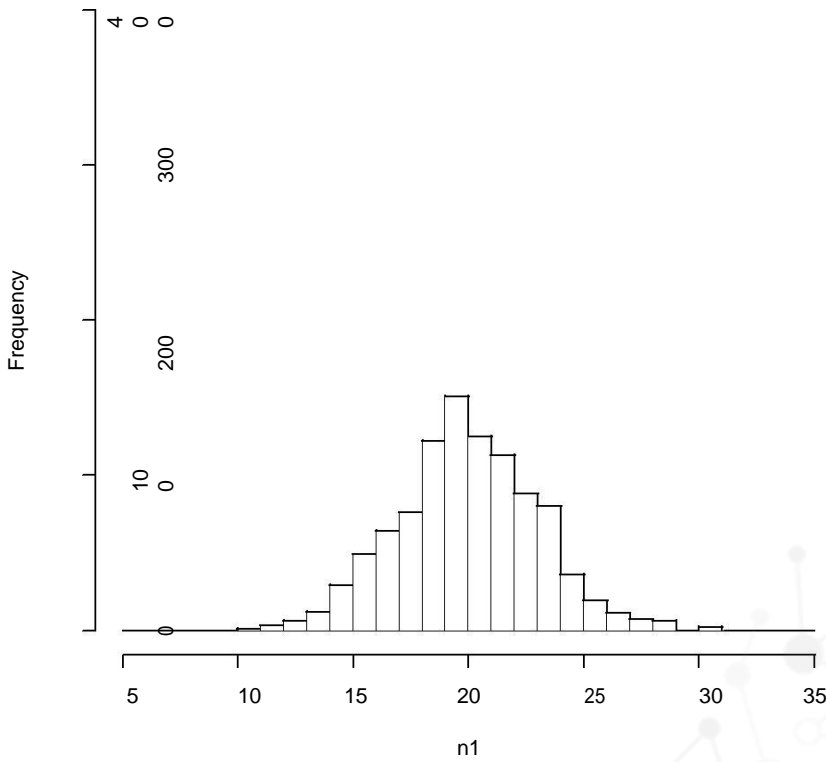
Note-se que o valor da dispersão só ganha significado quando comparado com o resto dos dados. Para perceber melhor este aspeto, vamos recorrer a uma interpretação gráfica.

Os dois diagramas abaixo foram gerados a partir de 1000 observações de variáveis estatísticas, ambas com a média igual a 20, mas com diferentes valores do desvio-padrão.

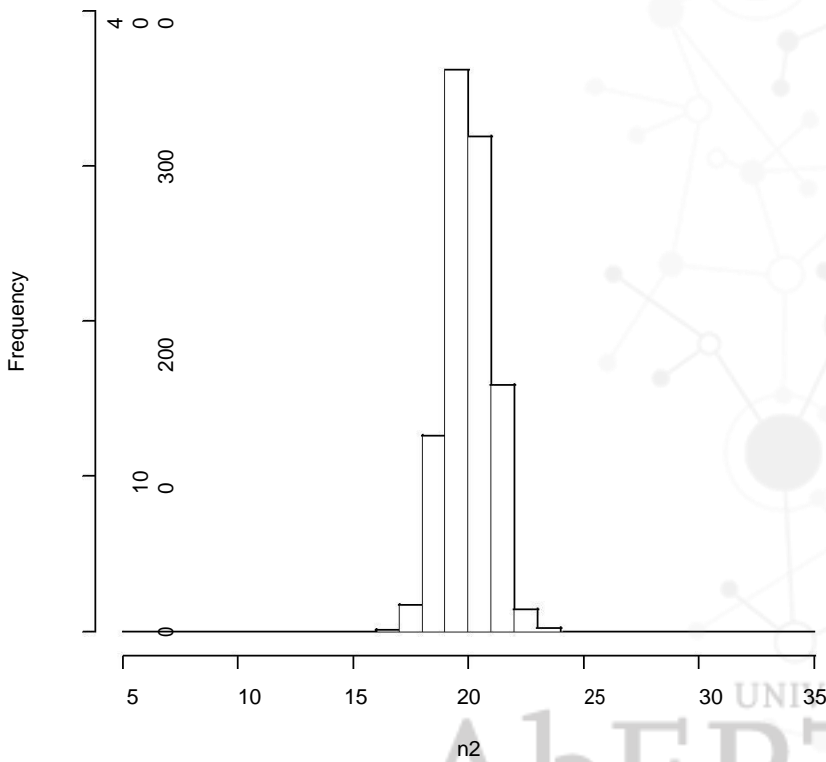
No caso dos gráficos abaixo, a distribuição referente ao primeiro tem desvio-padrão 3, enquanto que no seguinte o desvio-padrão é 1, logo com menor dispersão.



Histogram of n1



Histogram of n2



O gráfico de cima representa a variável associada a um maior desvio-padrão. O seu pico é bem mais esbatido e disperso do que o do gráfico de baixo, que por seu turno evidencia um pico muito mais pronunciado e bem definido, associado a um menor valor do desvio-padrão, logo com menor dispersão. É esse o significado de desvio-padrão: quanto maior for, por comparação com o resto dos dados, mais dispersos vão estar esses dados. Quanto menor for, menos dispersos estarão, i.e. estarão mais concentrados em torno da média.

Exercício: Calcule no R a média e desvio-padrão da variável `irmaos`. Compare as duas grandezas e avalie o grau de dispersão dos dados.

Capítulo 4

Exercícios

Exercícios Propostos

4.1 Numa amostra de 12 smartphones de uma determinada marca e modelo, verificaram-se os seguintes tempos de duração, em horas, até descarregamento total da bateria:

18, 20, 19, 18, 21, 20, 19, 19, 20, 22, 17, 20

- Calcule a média aritmética pela definição e usando o comando `mean()`.
- Calcule o desvio-padrão pela definição e usando o comando `sd()`.

4.2 De acordo com os dados da tabela seguinte, a distribuição dos salários (em Euros) dos colaboradores de uma empresa é:

x_i	F_i
1000	10
1100	8
1200	12
1300	8
1400	10
1500	12

- Indique o número de trabalhadores da empresa.
- Obtenha as frequências absolutas acumuladas e as frequências relativas simples.
- Determine a média e o desvio padrão desta distribuição.
- Represente as frequências absolutas simples num diagrama de barras.

4.3 Considere a seguinte amostra, de dados referente aos pesos (em Kg), de 15 alunos de uma turma de Educação Física:

70, 68, 83, 56, 49, 68, 73, 68, 55, 59, 48, 62, 63, 67, 75

Usando o R, determine:

- a)** A média aritmética.
- b)** O desvio padrão.

4.4 Considere os seguintes valores referentes às taxas de inflação de 6 países da união europeia, num dado ano.

Portugal	França	Espanha	Itália	Alemanha	U.K.
2.2	2.0	0.4	1.3	1.4	1.9

- a) Represente os dados num diagrama de barras.
- b) Represente os dados num diagrama circular.
- c) Obtenha a média das taxas de inflação dos 6 países.

Capítulo 5

Resumindo...

5.1 O que vos foi dito

Nesta Aula Aberta considerámos:

- Rever conceitos básicos de Estatística e algumas metodologias, tal como os conceitos de População e Amostra, a construção de tabelas de frequências e diagramas de barras.
- Abordar o conceito de variável e a classificação de variáveis quanto ao carater e quanto ao tipo. Debruçamo-nos apenas sobre o caso de variáveis discretas.
- Ilustrar o cálculo de medidas de redução de dados e escolhemos os exemplos de uma medida de localização - a média e de uma medida de dispersão - o desvio padrão.
- Ajudar a instalação do software R e fomentar a sua utilização, mostrando a sua utilização na resolução de problemas.

5.2 O que não vos foi dito

Nesta Aula Aberta foram deixadas algumas pontas soltas, que enumeramos de seguida. Estas implicam o conhecimento mais profundo de noções e metodologias estatísticas, que saem fora do âmbito desta Aula Aberta. A saber:

- O estudo e análise de variáveis contínuas.
- O estudo e exploração de mais medidas de redução de dados fundamentais para a caracterização da população e da amostra.
- Mais formas de ilustração gráfica que facilitam a interpretação visual dos dados.
- Aprofundar os conhecimentos de comandos do R para manuseamento de dados formatados em bases de dados diversas.
- Aprofundar os conhecimentos de comandos do R para resolução de problemáticas mais avançadas.

Caso tenha interesse em aprofundar os conhecimentos, poderá frequentar a Licenciatura em Matemática e Aplicações, ou inscrever-se nas unidades curriculares da área da Estatística, em oferta como unidades curriculares isoladas.

Os cursos da secção estão listados em:

<http://www.uab.pt/web/guest/organizacao/departamentos/dcet/sm/cursos>

Bibliografia

[1] Amílcar Oliveira e Teresa Oliveira (2011). Elementos de Estatística Descritiva. Repositório da Universidade Aberta.

<https://repositorioaberto.uab.pt/bitstream/10400.2/1986/1/Estat%20Descritiva.pdf>

[2] TEAM (2014) R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available at URL www.R-project.org.

[3] Robert J. Knell (2014). Introductory R. Published by Robert Knell, Hersham, Walton on Thames United Kingdom KT12, 5RH.

<http://www.introductoryr.co.uk/Introductory%20example%20chapters.pdf>