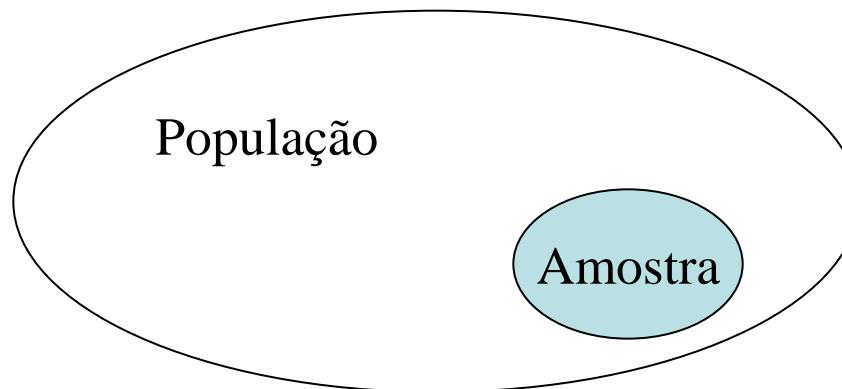


Estatística descritiva

Introdução

- **População:**
 - **Def:** Conjunto de todos os objectos que se pretende estudar.
 - **Ex:** Conjunto de todos os alunos do ISEP.
- **Amostra:**
 - **Def:** Subconjunto finito de uma população.



Introdução

- **Estatística descritiva:**
 - **Objectivo:** Representar a informação contida nos *dados* de forma *sintética* e *organizada*. Por exemplo, usando distrib. de freq. ou medidas descritivas.
- **Dados:**
 - Qualitativos
 - **Nominais.** Ex: Cor do cabelo (<{preto, castanho, louro, ...})
 - **Ordinais.** Ex: Classificação num teste (<{mau, insuficiente, suficiente, bom, ...})
 - Quantitativos:
 - **Discretos.** Ex: Nº de falhas diárias de uma máquina (<{0, 1, 2, 3, ...})
 - **Contínuos.** Ex: Tempo de atendimento de um cliente ([0, +∞[)
- **Variável:**
 - **Def:** Uma *variável* representa uma característica de uma população. Pode ser *qualitativa* (não numérica) ou *quantitativa* (numérica).
 - **Ex:** X: “Nº de filhos de um casal de uma cidade”

Introdução

- *Classes:*
 - **Def:** Uma *classe* (ou *categoria*) de uma variável qualitativa é qualquer um dos valores que a variável pode tomar.
 - **Def:** Uma *classe* de uma variável quantitativa discreta é qualquer um dos valores numéricos que a variável pode tomar.
 - **Def:** Uma *classe* de uma variável quantitativa contínua é um intervalo, na forma $[a, b[$ ou $]a, b]$, de valores numéricos que a variável pode tomar.
- *Marca de classe (variáveis contínuas):*
 - **Def:** A *marca da classe* i , x_i , de uma variável contínua é o ponto médio dessa classe, isto é,

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

Introdução

- *Frequência absoluta:*

- **Def:** A *frequência absoluta* da classe i , n_i , é o números de observações dessa classe.
- **Nota:** Se n é o tamanho de uma amostra com c classes, então

$$n = \sum_{i=1}^c n_i$$

- *Frequência relativa:*

- **Def:** A *frequência relativa* da classe i , f_i , é

$$f_i = \frac{n_i}{n}$$

- **Nota:** $\sum_{i=1}^c f_i = 1$

Introdução

- *Frequência relativa acumulada (dados quantitativos):*
 - *Def:* A frequência relativa acumulada da classe i , F_i , é

$$F_i = \sum_{k=1}^i f_k$$

- *Distribuição de frequência:*
 - *Def:* É uma *representação tabular ou gráfica dos dados* que mostra a frequência de cada classe observada.

Dados qualitativos

- **Tabela de frequências de dados qualitativos**
 - **Ex:** A seguinte *tabela* representa a distribuição das *frequências absoluta* e *relativa* das preferências de ocupação dos tempos livres de 50 habitantes de uma cidade.

X:”Tipo de preferência”

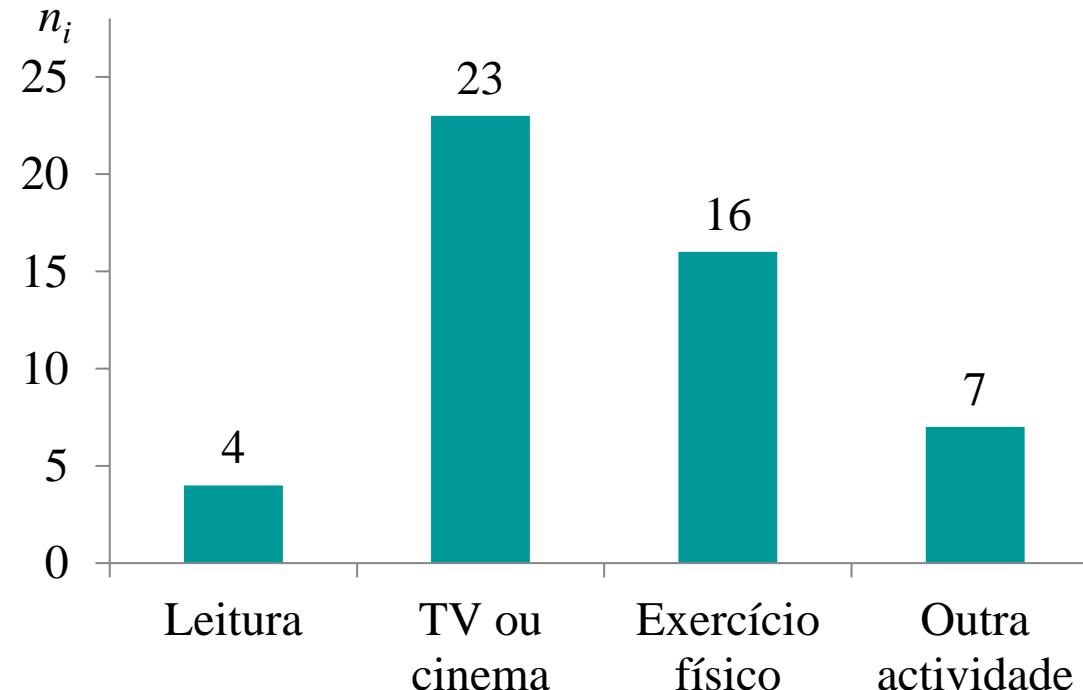
Tipo de preferência (x_i)	Frequência absoluta (n_i)	Frequência relativa (f_i)
Leitura	4	0,08
TV ou cinema	23	0,46
Exercício físico	16	0,32
Outra actividade	7	0,14
Total	50	1

Dados qualitativos

- *Gráfico de barras:*

- *Def:* É uma *representação gráfica de dados qualitativos* onde se usam barras separadas (de igual largura) cuja altura é proporcional à frequência (absoluta ou relativa) da classe correspondente.

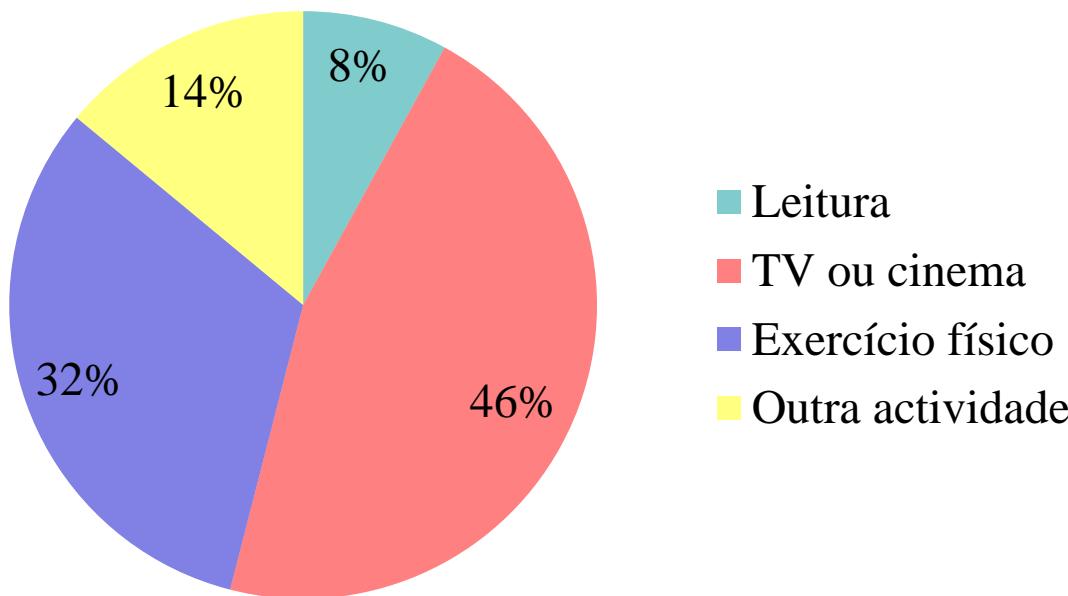
- *Ex:*



R: barplot()

Dados qualitativos

- *Gráfico circular:*
 - *Def:* É uma *representação gráfica de dados qualitativos num círculo dividido em sectores circulares de área proporcional à frequência (relativa) da classe correspondente.*
 - *Ex:*



R: pie()

Dados quantitativos discretos

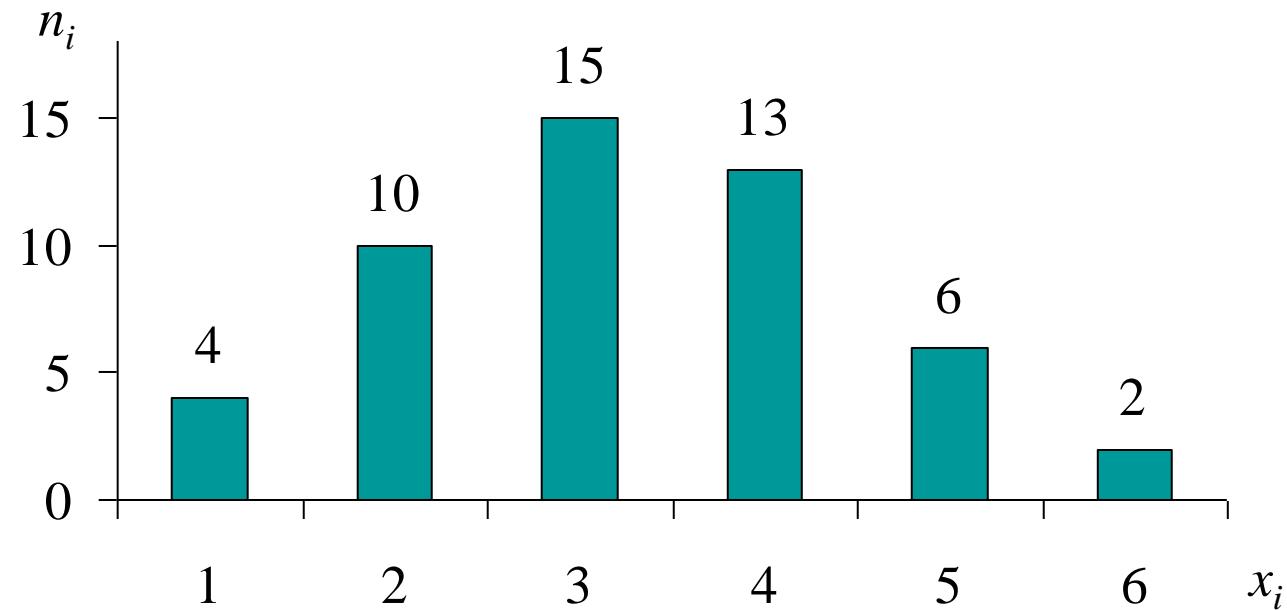
- *Tabela de frequências de dados quantitativos discretos*
 - **Ex:** A seguinte *tabela* representa a distribuição das *frequências absoluta, relativa e relativa acumulada* dos tamanhos dos agregados familiares registados num inquérito a 50 famílias duma dada cidade.

X: "Tamanho do agregado familiar"

Tamanho do agregado (x_i)	Frequência absoluta (n_i)	Frequência relativa (f_i)	Freq. Relat. acumulada (F_i)
1	4	0,08	0,08
2	10	0,2	0,28
3	15	0,3	0,58
4	13	0,26	0,84
5	6	0,12	0,96
6	2	0,04	1
Total	50	1	

Dados quantitativos discretos

- *Gráfico de barras:*
 - *Def:* É uma *representação gráfica de dados discretos* onde se usam barras separadas (de igual largura) cuja altura é proporcional à frequência (absoluta ou relativa) da classe correspondente.
 - *Ex:*



Dados quantitativos contínuos

- *Tabela de frequências de dados quantitativos contínuos:*
 - **Ex:** A seguinte *tabela* representa a distribuição das *frequências absoluta, relativa e relativa acumulada* do tempo de execução, em segundos, de 34 processos duma empresa.

X : "Tempo de execução dum processo, em segundos"

Tempo (Classe)	Marca de classe (x_i)	Freq. abs. (n_i)	Freq. relat. (f_i)	Freq. relat. acumulada (F_i)
[5, 7[6	5	0.1471	0.1471
[7, 9[8	13	0.3824	0.5295
[9, 11[10	8	0.2353	0.7648
[11, 15[13	8	0.2353	1.0001
	Total	34	1.0001	

Dados quantitativos contínuos

- *Como determinar o número de intervalos de classe e a sua amplitude?*
 - Não há um número ideal de intervalos de classe.
 - A *regra de Sturges* costuma dar bons resultados para grandes amostras ($n \geq 30$).
 - **Def:** De acordo com a *regra de Sturges*, o nº de intervalos de classe, k , é o menor inteiro tal que

$$k > 1 + \log_2(n) \approx 1 + 3.322 \log_{10}(n)$$

- A amplitude, h , dos intervalos pode ser obtida fazendo

$$h \geq \frac{x_{\max} - x_{\min}}{k}$$

onde x_{\min} e x_{\max} são os valores mínimo e máximo observados na amostra, respectivamente.

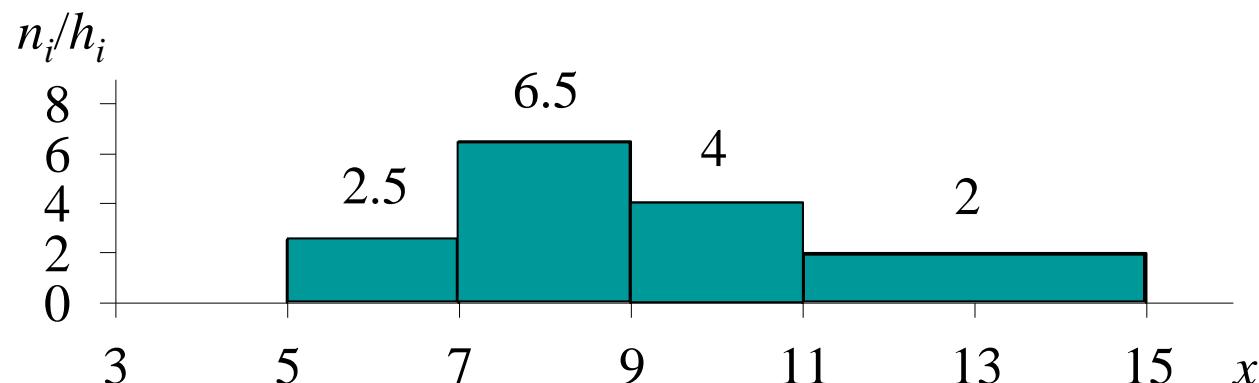
Dados quantitativos contínuos

- **Histograma:**

- **Def:** É uma *representação gráfica de dados contínuos* onde se usam barras contíguas de área proporcional à frequência (absoluta ou relativa) da classe correspondente.

- **Ex:**

Classe	Freq. abs. (n_i)	Amplitude da classe (h_i)	Altura da barra (n_i/h_i)
[5, 7[5	2	2.5
[7, 9[13	2	6.5
[9, 11[8	2	4
[11, 15[8	4	2



R: `hist()`