

# Analisi dei dati

kanopo

2022

## Indice

<b>1</b>	<b>Statistica descrittiva</b>	<b>1</b>
1.1	Le grandezze che sintetizzano i dati . . . . .	1
1.1.1	Media . . . . .	1
1.1.2	Mediana . . . . .	1
1.1.3	Moda . . . . .	1
1.1.4	Varianza e deviazione standard . . . . .	1
1.1.5	Percentili campionari e box plot . . . . .	2

## Elenco delle figure

## Elenco delle tabelle

## 1 Statistica descrittiva

### 1.1 Le grandezze che sintetizzano i dati

#### 1.1.1 Media

Dato un insieme  $x_1, x_2, \dots, x_n$  di dati, la media è la media aritmetica:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

#### 1.1.2 Mediana

Dato un'insieme di dati di ampiezza  $n$ , lo si ordina dal minore al maggiore. La mediana è il valore che occupa la posizione  $\frac{n+1}{2}$  in caso di un'insieme dispari, oppure la media fra  $\frac{n}{2}$  e  $\frac{n+1}{2}$  se pari.

#### 1.1.3 Moda

La moda campionaria di un'insieme di dati, se esiste, è l'unico valore che ha la frequenza massima.

#### 1.1.4 Varianza e deviazione standard

Dato un'insieme di dati  $x_1, x_2, \dots, x_n$ , si dice varianza campionaria ( $s^2$ ), la quantità

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

Una comodità per il calcolo è

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

Si dice **deviazione standard campionaria** e si denota con  $s$ , la quantità

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

### 1.1.5 Percentili campionari e box plot

Sia  $k$  un numero intero  $0 \leq k \leq 100$ .

Dato un campione di dati, esiste sempre un dato che è contemporaneamente maggiore del  $k$  per cento dei dati e minore del  $100 - k$  per cento.