

# ELEMENTI DI STATISTICA DESCRITTIVA

Vlacci Fabio

Dipartimento di Matematica “U. Dini”, Università di Firenze  
Viale Morgagni 67/A, 50134 - Firenze, Italy, vlacci@math.unifi.it

A.A. 2016-17

# Terminologia

- ▶ In un *esperimento* ogni risultato delle caratteristiche *osservabili* dell'esperimento si dice *esito*.

# Terminologia

- ▶ In un *esperimento* ogni risultato delle caratteristiche *osservabili* dell'esperimento si dice *esito*.
- ▶ Se un esperimento può essere ripetuto si dice *prova* ogni singola esecuzione dell'esperimento.

# Terminologia

- ▶ In un *esperimento* ogni risultato delle caratteristiche *osservabili* dell'esperimento si dice *esito*.
- ▶ Se un esperimento può essere ripetuto si dice *prova* ogni singola esecuzione dell'esperimento.
- ▶ Un *evento* è un insieme di esiti.

# Terminologia

- ▶ In un *esperimento* ogni risultato delle caratteristiche *osservabili* dell'esperimento si dice *esito*.
- ▶ Se un esperimento può essere ripetuto si dice *prova* ogni singola esecuzione dell'esperimento.
- ▶ Un *evento* è un insieme di esiti. I singoli esiti sono anche detti *eventi elementari* e, in questo senso, un evento è anche detto *evento composto* (da eventi elementari)).

# Terminologia

- ▶ In un *esperimento* ogni risultato delle caratteristiche *osservabili* dell'esperimento si dice *esito*.
- ▶ Se un esperimento può essere ripetuto si dice *prova* ogni singola esecuzione dell'esperimento.
- ▶ Un *evento* è un insieme di esiti. I singoli esiti sono anche detti *eventi elementari* e, in questo senso, un evento è anche detto *evento composto* (da eventi elementari)).
- ▶ Un esito o un evento di un esperimento si dicono *casuali* o *aleatori* se non è possibile prevederne il verificarsi a priori in modo certo.
- ▶ La totalità degli eventi elementari associati ad un esperimento è lo *spazio campionario* dell'esperimento.

# Terminologia

- ▶ In un *esperimento* ogni risultato delle caratteristiche *osservabili* dell'esperimento si dice *esito*.
- ▶ Se un esperimento può essere ripetuto si dice *prova* ogni singola esecuzione dell'esperimento.
- ▶ Un *evento* è un insieme di esiti. I singoli esiti sono anche detti *eventi elementari* e, in questo senso, un evento è anche detto *evento composto* (da eventi elementari)).
- ▶ Un esito o un evento di un esperimento si dicono *casuali* o *aleatori* se non è possibile prevederne il verificarsi a priori in modo certo.
- ▶ La totalità degli eventi elementari associati ad un esperimento è lo *spazio campionario* dell'esperimento.

evento  $\mathcal{A} \rightarrow A \subset S$  spazio campionario

## Dati campionari

In genere in una raccolta di dati campionari di una popolazione o *rilevazione campionaria*, oltre agli esiti ovvero alle caratteristiche della popolazione che si intendono registrare, vengono indicati anche le frequenze (relative e/o assolute) per tali dati.



# Dati campionari

In genere in una raccolta di dati campionari di una popolazione o *rilevazione campionaria*, oltre agli esiti ovvero alle caratteristiche della popolazione che si intendono registrare, vengono indicati anche le frequenze (relative e/o assolute) per tali dati.

Inoltre *prima* di procedere alla rilevazione campionaria andrebbe appositamente definito una procedura per il *piano di campionamento*

## Dati campionari

In genere in una raccolta di dati campionari di una popolazione o *rilevazione campionaria*, oltre agli esiti ovvero alle caratteristiche della popolazione che si intendono registrare, vengono indicati anche le frequenze (relative e/o assolute) per tali dati.

Inoltre *prima* di procedere alla rilevazione campionaria andrebbe appositamente definito una procedura per il *piano di campionamento*. Ad esempio, se bisogna campionare un'area territoriale (abbastanza estesa), sarà cura del ricercatore stabilire se (per ragioni pratiche o teoriche) sia meglio seguire uno schema di rilevamento casuale o con geometria regolare (ad esempio uniforme o a grappoli).

## Rappresentazione dei dati statistici

Le frequenze ma anche le serie storiche dei dati possono essere visualizzati usando dei diagrammi; in particolare potranno essere usati dei *diagrammi a punti*, dei *diagrammi a linea* o *istogrammi* e degli *areogrammi* come i diagrammi a torta o pie charts.

## Rappresentazione dei dati statistici

Le frequenze ma anche le serie storiche dei dati possono essere visualizzati usando dei diagrammi; in particolare potranno essere usati dei *diagrammi a punti*, dei *diagrammi a linea* o *istogrammi* e degli *areogrammi* come i diagrammi a torta o pie charts. Quando il numero di osservazioni è molto grande, è bene considerare una rappresentazione a istogrammi per intervalli o *classi* di valori.

# Rappresentazione dei dati statistici

Le frequenze ma anche le serie storiche dei dati possono essere visualizzati usando dei diagrammi; in particolare potranno essere usati dei *diagrammi a punti*, dei *diagrammi a linea* o *istogrammi* e degli *areogrammi* come i diagrammi a torta o pie charts. Quando il numero di osservazioni è molto grande, è bene considerare una rappresentazione a istogrammi per intervalli o *classi* di valori. La scelta del numero di intervalli è una questione delicata ed *influenza* l'aspetto dell'istogramma.

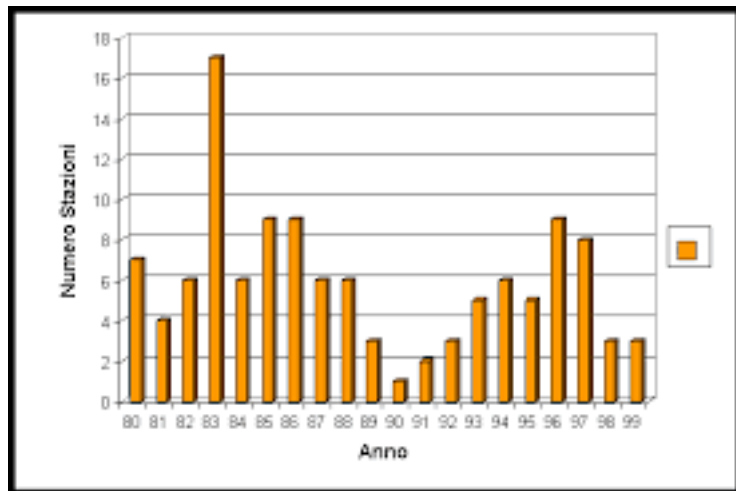
# Rappresentazione dei dati statistici

Le frequenze ma anche le serie storiche dei dati possono essere visualizzati usando dei diagrammi; in particolare potranno essere usati dei *diagrammi a punti*, dei *diagrammi a linea* o *istogrammi* e degli *areogrammi* come i diagrammi a torta o pie charts. Quando il numero di osservazioni è molto grande, è bene considerare una rappresentazione a istogrammi per intervalli o *classi* di valori. La scelta del numero di intervalli è una questione delicata ed *influenza* l'aspetto dell'istogramma.

regola (empirica) di Sturges

*numero classi* =  $1 + 3,322 \log_{10}(n)$  con  $n$  numero di osservazioni.

## Esempio di Istogramma

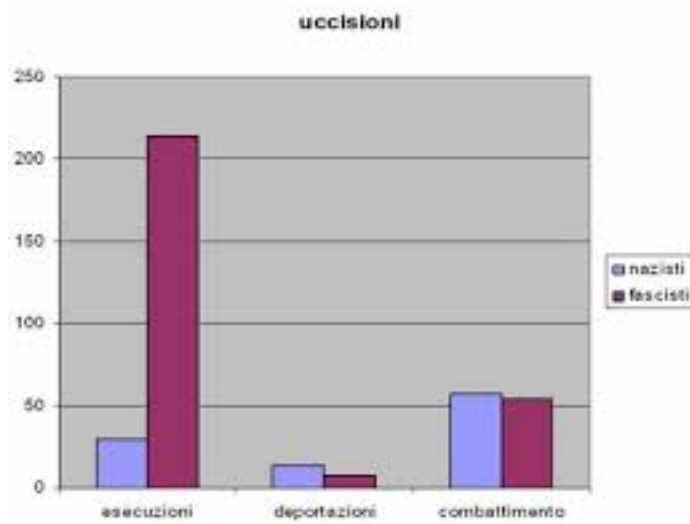


## Esempio di Istogramma



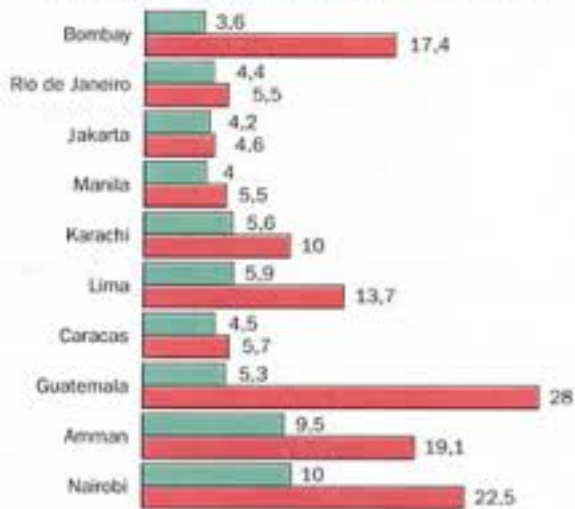


## Esempio di Istogramma



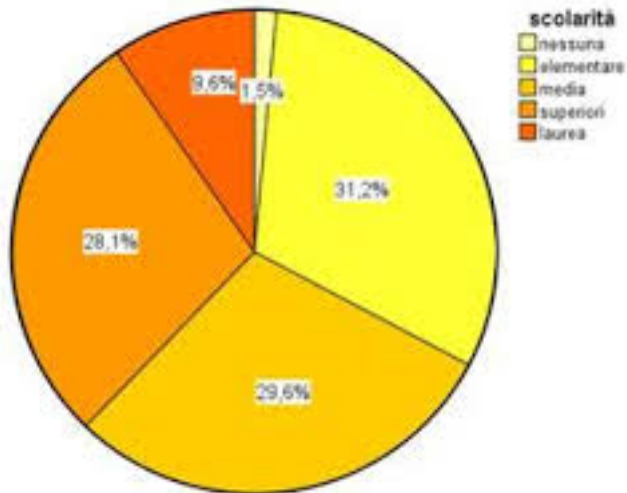
## Esempio di Istogramma

Crescita comparata delle città e delle baraccopoli



crescita annuale in %  delle città  delle baraccopoli

## Pie chart o Diagramma a torta



# Indicatori statistici di dati numerici

Per ottenere una descrizione sommaria di un insieme di dati numerici e per confrontare fra loro insiemi diversi di dati numerici campionari è opportuno ricavare dei termini riassuntivi detti *indicatori statistici*.

# Indicatori statistici di dati numerici

Per ottenere una descrizione sommaria di un insieme di dati numerici e per confrontare fra loro insiemi diversi di dati numerici campionari è opportuno ricavare dei termini riassuntivi detti *indicatori statistici*.

In particolare tra gli indicatori di *tendenza centrale* si evidenziano

- ▶ media campionaria

# Indicatori statistici di dati numerici

Per ottenere una descrizione sommaria di un insieme di dati numerici e per confrontare fra loro insiemi diversi di dati numerici campionari è opportuno ricavare dei termini riassuntivi detti *indicatori statistici*.

In particolare tra gli indicatori di *tendenza centrale* si evidenziano

- ▶ media campionaria
- ▶ mediana

## Media campionaria o Media aritmetica dei dati

Se  $x_1, x_2, \dots, x_n$  sono i dati numerici rilevati, la media campionaria è il numero

$$\bar{x} := \frac{x_1 + x_2 + \dots + x_n}{n}$$

# Media campionaria o Media aritmetica dei dati

Se  $x_1, x_2, \dots, x_n$  sono i dati numerici rilevati, la media campionaria è il numero

$$\bar{x} := \frac{x_1 + x_2 + \dots + x_n}{n}$$

- ▶ La media campionaria  $\bar{x}$  NON è necessariamente uno dei dati;



# Media campionaria o Media aritmetica dei dati

Se  $x_1, x_2, \dots, x_n$  sono i dati numerici rilevati, la media campionaria è il numero

$$\bar{x} := \frac{x_1 + x_2 + \dots + x_n}{n}$$

- ▶ La media campionaria  $\bar{x}$  NON è necessariamente uno dei dati;
- ▶ tuttavia  $\min\{x_i\}_{i=1,\dots,n} \leq \bar{x} \leq \max\{x_i\}_{i=1,\dots,n}$ .
- ▶ Esistono anche ALTRE medie (geometrica, armonica, ecc.), ma quella campionaria o aritmetica può essere presa come *baricentro* dei dati.

## Osservazione importante

Poichè la somma di numeri reali è commutativa e associativa, allora

$$\bar{x} := \frac{\overbrace{x_1 + \dots + x_1}^{n_1}}{n} + \frac{\overbrace{x_2 + \dots + x_2}^{n_2}}{n} + \dots + \frac{\overbrace{x_n + \dots + x_n}^{n_n}}{n}$$

con  $n_1 + n_2 + \dots + n_n = n$ .

## Osservazione importante

Poichè la somma di numeri reali è commutativa e associativa, allora

$$\bar{x} := \frac{\overbrace{x_1 + \dots + x_1}^{n_1}}{n} + \frac{\overbrace{x_2 + \dots + x_2}^{n_2}}{n} + \dots + \frac{\overbrace{x_n + \dots + x_n}^{n_n}}{n}$$

con  $n_1 + n_2 + \dots + n_n = n$ .

Si noti che

$$\bar{x} = x_1 p_1 + x_2 p_2 + \dots + x_n p_n$$

ove  $p_j = \frac{n_j}{n}$  è la frequenza relativa di  $x_j$ .

# Media geometrica

Supponiamo che i dati campionari  $x_1, x_2, \dots, x_n$  siano tutti numeri positivi.

## Media geometrica

Supponiamo che i dati campionari  $x_1, x_2, \dots, x_n$  siano tutti numeri positivi. Allora la loro *media geometrica* è data da

$$M_g = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \sqrt[n]{\prod_{j=1}^n x_j}$$

## Media geometrica

Supponiamo che i dati campionari  $x_1, x_2, \dots, x_n$  siano tutti numeri positivi. Allora la loro *media geometrica* è data da

$$M_g = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \sqrt[n]{\prod_{j=1}^n x_j}$$

Si noti questo **importante** legame fra media geometrica e media aritmetica.

$$\log(M_g) = \log \sqrt[n]{\prod_{j=1}^n x_j} = \frac{\sum_{j=1}^n \log(x_j)}{n}$$

# Mediana

Un altro indicatore statistico di tendenza centrale è la *mediana campionaria*.

# Mediana

Un altro indicatore statistico di tendenza centrale è la *mediana campionaria*.

Esso è il valore che divide in due parti uguali i dati, quando questi sono ordinati in senso crescente o decrescente.



# Mediana

Un altro indicatore statistico di tendenza centrale è la *mediana campionaria*.

Esso è il valore che divide in due parti uguali i dati, quando questi sono ordinati in senso crescente o decrescente. Più precisamente, se il numero  $n$  dei dati è dispari, allora il valore della mediana è dato dal dato (detto *mediano*) di posizione  $(n + 1)/2$  nell'elenco ordinato di dati. Se invece  $n$  è pari vi saranno due dati mediani (quello di posizione  $n/2$  e il successivo) e allora andrà considerata come mediana la media campionaria dei due dati mediani.

## Quartili e percentili

Se il numero di osservazioni/dati è elevato, può essere utile suddividere ulteriormente l'insieme dei dati procedendo al calcolo della mediana della prima e della seconda metà dei dati una volta ordinati in senso crescente o decrescente.

## Quartili e percentili

Se il numero di osservazioni/dati è elevato, può essere utile suddividere ulteriormente l'insieme dei dati procedendo al calcolo della mediana della prima e della seconda metà dei dati una volta ordinati in senso crescente o decrescente.

In questo modo i dati sono suddivisi in 4 classi dette *quartili*.

## Quartili e percentili

Se il numero di osservazioni/dati è elevato, può essere utile suddividere ulteriormente l'insieme dei dati procedendo al calcolo della mediana della prima e della seconda metà dei dati una volta ordinati in senso crescente o decrescente.

In questo modo i dati sono suddivisi in 4 classi dette *quartili*. In generale, se i dati possono essere divisi in 100 classi, dette *percentili*, il primo quartile corrisponde al 25-esimo percentile mentre il 50-esimo percentile coincide con la mediana o secondo quartile.

## Quartili e percentili

Se il numero di osservazioni/dati è elevato, può essere utile suddividere ulteriormente l'insieme dei dati procedendo al calcolo della mediana della prima e della seconda metà dei dati una volta ordinati in senso crescente o decrescente.

In questo modo i dati sono suddivisi in 4 classi dette *quartili*.

In generale, se i dati possono essere divisi in 100 classi, dette *percentili*, il primo quartile corrisponde al 25-esimo percentile mentre il 50-esimo percentile coincide con la mediana o secondo quartile.

Per un valore  $p$  ( $0 \leq p \leq 100$ ) si parla di  $p$ -esimo percentile.

# Moda

La *moda campionaria* (detta anche *valore normale* o *norma*) è il dato (o i dati) che si presenta(no) con frequenza massima.

# Moda

La *moda campionaria* (detta anche *valore normale* o *norma*) è il dato (o i dati) che si presenta(no) con frequenza massima. Di conseguenza si parla di *distribuzione dei dati* di tipo *unimodale* o *bimodale* e in generale *plurimodale*.

# Moda

La *moda campionaria* (detta anche *valore normale* o *norma*) è il dato (o i dati) che si presenta(no) con frequenza massima. Di conseguenza si parla di *distribuzione dei dati* di tipo *unimodale* o *bimodale* e in generale *plurimodale*.

La moda è di facile identificazione in un istogramma;



# Moda

La *moda campionaria* (detta anche *valore normale* o *norma*) è il dato (o i dati) che si presenta(no) con frequenza massima. Di conseguenza si parla di *distribuzione dei dati* di tipo *unimodale* o *bimodale* e in generale *plurimodale*.

La moda è di facile identificazione in un istogramma; se i dati sono raggruppati in classi, allora la classe con massima frequenza è detta *classe modale*.

## Indicatore di dispersione dei dati

Se la media campionaria  $\bar{x}$  ha il ruolo di *baricentro* dei dati  $\{x_1, \dots, x_n\}$ , allora la dispersione dei dati  $x_j$  da  $\bar{x}$  può essere calcolata considerando gli *scarti*

$$d_j := x_j - \bar{x}$$

oppure

$$|d_j| = |x_j - \bar{x}| \geq 0.$$

## Indicatore di dispersione dei dati

Se la media campionaria  $\bar{x}$  ha il ruolo di *baricentro* dei dati  $\{x_1, \dots, x_n\}$ , allora la dispersione dei dati  $x_j$  da  $\bar{x}$  può essere calcolata considerando gli *scarti*

$$d_i := x_i - \bar{x}$$

oppure

$$|d_i| = |x_i - \bar{x}| \geq 0.$$

Tuttavia, risulta

$$\sum_{i=1}^n d_i = 0$$

## Indicatore di dispersione dei dati

Se la media campionaria  $\bar{x}$  ha il ruolo di *baricentro* dei dati  $\{x_1, \dots, x_n\}$ , allora la dispersione dei dati  $x_j$  da  $\bar{x}$  può essere calcolata considerando gli *scarti*

$$d_j := x_j - \bar{x}$$

oppure

$$|d_j| = |x_j - \bar{x}| \geq 0.$$

Tuttavia, risulta

$$\sum_{i=1}^n d_i = 0$$

in quanto

$$\sum_{i=1}^n d_i = \sum_{i=1}^n (x_i - \bar{x}) = n\bar{x} - n\bar{x} = 0.$$

## Varianza campionaria o stimata dei dati

Si dice *varianza campionaria* o *varianza stimata* dei dati  $\{x_1, \dots, x_n\}$  di media campionaria  $\bar{x}$  il numero (non negativo)

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

## Varianza campionaria o stimata dei dati

Si dice *varianza campionaria* o *varianza stimata* dei dati  $\{x_1, \dots, x_n\}$  di media campionaria  $\bar{x}$  il numero (non negativo)

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

Si dice infine che

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

è la *deviazione standard stimata* o *scarto quadratico medio stimato* dei dati  $\{x_1, \dots, x_n\}$ .

## Alcune osservazioni

Si parla di *Varianza* (Var) e *Scarto quadratico* o *Deviazione standard* ( $\sigma$ ) se nelle formule precedenti si considera  $n$  invece di  $n - 1$ , vale a dire

$$\text{Var} := \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \sum_{i=1}^n \frac{d_i^2}{n}$$

$$\sigma = \sqrt{\text{Var}}$$

## Alcune osservazioni

Si parla di *Varianza* (Var) e *Scarto quadratico* o *Deviazione standard* ( $\sigma$ ) se nelle formule precedenti si considera  $n$  invece di  $n - 1$ , vale a dire

$$\text{Var} := \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \sum_{i=1}^n \frac{d_i^2}{n}$$
$$\sigma = \sqrt{\text{Var}}$$

La differenza fra la varianza e la varianza campionaria o stimata (e quindi fra la deviazione standard e quella stimata) risulta rilevante solo per  $n$  piccoli; per  $n$  grandi tali differenze sono trascurabili.



## Alcune osservazioni

Si parla di *Varianza* (Var) e *Scarto quadratico* o *Deviazione standard* ( $\sigma$ ) se nelle formule precedenti si considera  $n$  invece di  $n - 1$ , vale a dire

$$\text{Var} := \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \sum_{i=1}^n \frac{d_i^2}{n}$$

$$\sigma = \sqrt{\text{Var}}$$

La differenza fra la varianza e la varianza campionaria o stimata (e quindi fra la deviazione standard e quella stimata) risulta rilevante solo per  $n$  piccoli; per  $n$  grandi tali differenze sono trascurabili.

Si osservi inoltre che la deviazione standard e la deviazione standard stimata hanno la stessa unità di misura dei dati.

## Una formula alternativa....

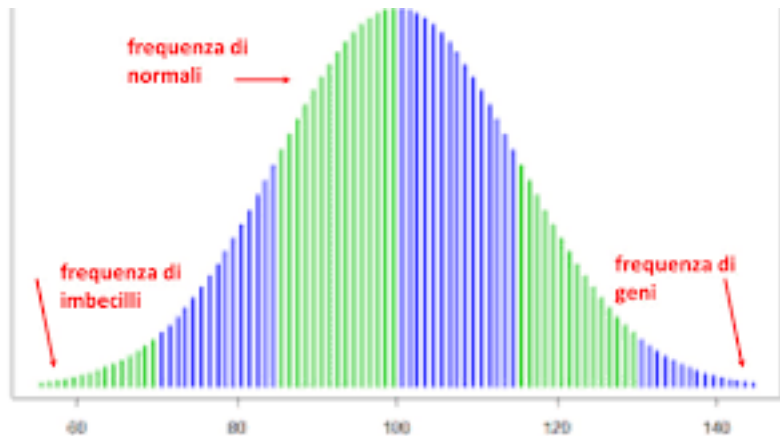
Da  $(x_i - \bar{x})^2 = x_i^2 - 2x_i\bar{x} + \bar{x}^2$  e dal fatto che  $n\bar{x} = x_1 + \dots + x_n$  si ricava anche

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

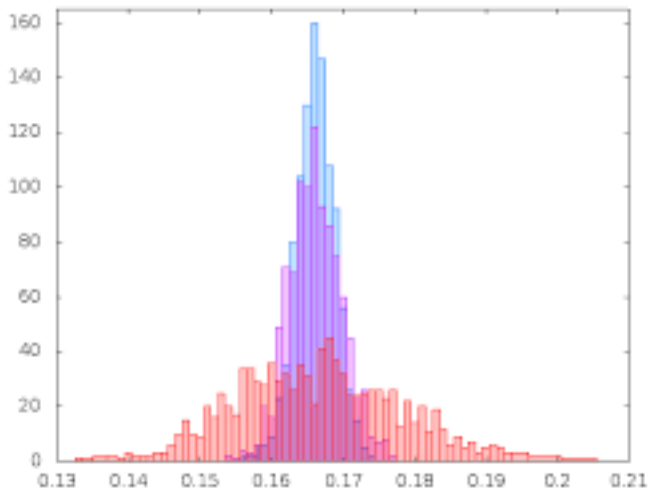
## Dati relativi a grandezze continue

Se volessimo rappresentare in forma di istogramma di frequenze un insieme di misure di una grandezza che può variare con continuità, allora, procedendo con la suddivisione dell'intero intervallo di misure in un numero finito  $n$  di intervallini della medesima ampiezza, verremmo a ottenere un istogramma in cui l'area di ciascun rettangolino verticale risulti proporzionale al numero di misure che cadono nella base dell'intervallo considerato.

## Esempio di istogramma di frequenza di dati relativi a grandezza che può variare con continuità



## Esempio di istogramma di frequenza di dati relativi a grandezza che può variare con continuità



## Dati relativi a grandezze continue

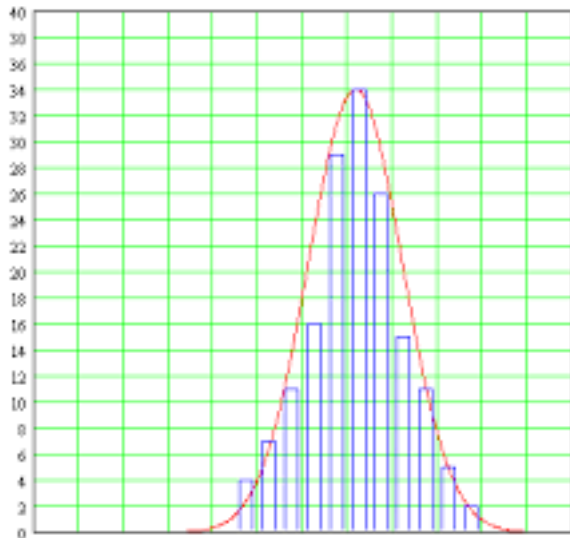
Se inoltre si fa l'ipotesi che la numerosità del campione sia molto grande, allora i rettangolini dell'istogramma di frequenza diventeranno sempre più sottili e l'istogramma stesso sarà assimilabile ad una curva continua detta *curva di distribuzione delle frequenze*.

## Dati relativi a grandezze continue

Se inoltre si fa l'ipotesi che la numerosità del campione sia molto grande, allora i rettangolini dell'istogramma di frequenza diventeranno sempre più sottili e l'istogramma stesso sarà assimilabile ad una curva continua detta *curva di distribuzione delle frequenze*.

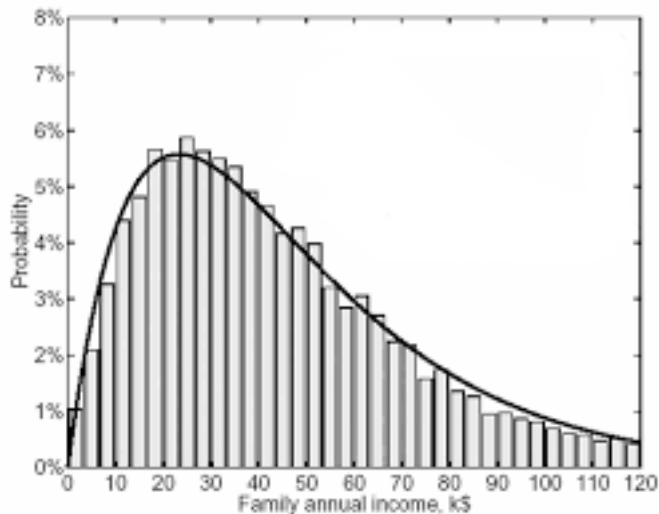
In genere, come per gli istogrammi "discreti", si suppone (ovvero si fa in modo) che l'area complessiva del sottografico di una curva di distribuzione di frequenze sia uguale a 1.

## Esempio di istogramma con curva di distribuzione di frequenze





## Esempio di istogramma con curva di distribuzione di frequenze



## Distribuzione Normale o Gaussiana

Spesso la curva di distribuzione di frequenza è il grafico di una funzione (dipendente dai parametri reali  $A$ ,  $B$  e  $C$ )

$$f(x) = Ae^{-B(x-C)^2}$$

e in questi casi si dice che la distribuzione è *Normale* o *Gaussiana*.

# Distribuzione Normale o Gaussiana

Spesso la curva di distribuzione di frequenza è il grafico di una funzione (dipendente dai parametri reali  $A$ ,  $B$  e  $C$ )

$$f(x) = Ae^{-B(x-C)^2}$$

e in questi casi si dice che la distribuzione è *Normale* o *Gaussiana*.

Vi sono ragioni empiriche ma anche teoriche per stabilire o prevedere che certi dati si distribuiscano seguendo una curva di tipo gaussiano.

# Distribuzione Normale o Gaussiana

Spesso la curva di distribuzione di frequenza è il grafico di una funzione (dipendente dai parametri reali  $A$ ,  $B$  e  $C$ )

$$f(x) = Ae^{-B(x-C)^2}$$

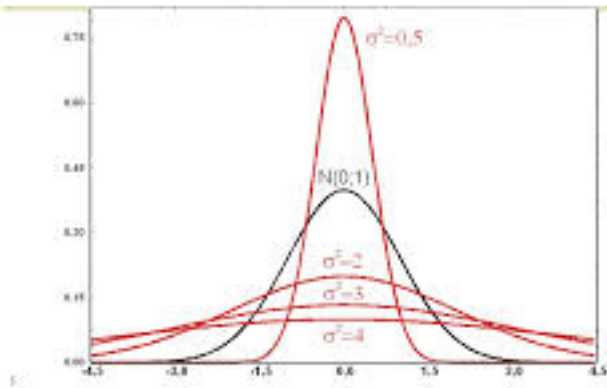
e in questi casi si dice che la distribuzione è *Normale* o *Gaussiana*.

Vi sono ragioni empiriche ma anche teoriche per stabilire o prevedere che certi dati si distribuiscano seguendo una curva di tipo gaussiano.

In linea di principio questo accertamento richiederebbe degli specifici *test statistici*; a livello pratico, sfruttando il fondamentale **Teorema Limite Centrale**, se si considerano medie campionarie di dati (numerosi) relativi ad una grandezza continua, allora queste tendono a disporsi seguendo una distribuzione normale o gaussiana.

# Esempio di gaussiane

## Distribuzione Normale



# Distribuzione Normale o Gaussiana

Se quindi si sa che una distribuzione di dati di media aritmetica  $\mu$  e scarto quadratico medio  $\sigma$  è di tipo gaussiano allora nell'espressione

$$f(x) = Ae^{-B(x-C)^2}$$

si pone

$$A := \frac{1}{\sigma\sqrt{2\pi}} \quad B := \frac{1}{2\sigma^2} \quad C := \mu$$

# Distribuzione Normale o Gaussiana

In questo modo la funzione distribuzione di frequenza diventa

$$x \mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

# Distribuzione Normale o Gaussiana

In questo modo la funzione distribuzione di frequenza diventa

$$x \mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

La curva associata a tale distribuzione di frequenze è una curva a campana più o meno ripida a seconda del valore di  $\sigma$ , simmetrica rispetto a  $x = \mu$  (che è anche un punto di massimo assoluto) e ha l'ulteriore proprietà che l'area del sottografico vale 1, ossia

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$



# Distribuzione Normale o Gaussiana Standardizzata

Si dice distribuzione Normale o Gaussiana *Standardizzata* quella distribuzione per cui  $\mu = 0$  e  $\sigma = 1$ ; con tale scelta l'espressione della relativa distribuzione di frequenze diventa (più semplicemente)

$$x \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

## Proprietà della Distribuzione Normale o Gaussiana

Quale che siano i valori di  $\mu$  e  $\sigma$  della gaussiana considerata, nell'intervallo

- ▶  $[\mu - \sigma, \mu + \sigma]$  si trova circa il 68% delle misure, ossia

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \simeq 0.68$$

## Proprietà della Distribuzione Normale o Gaussiana

Quale che siano i valori di  $\mu$  e  $\sigma$  della gaussiana considerata, nell'intervallo

- ▶  $[\mu - \sigma, \mu + \sigma]$  si trova circa il 68% delle misure, ossia

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \simeq 0.68$$

- ▶  $[\mu - 2\sigma, \mu + 2\sigma]$  si trova circa il 95% delle misure, ossia

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \simeq 0.95$$

## Proprietà della Distribuzione Normale o Gaussiana

Quale che siano i valori di  $\mu$  e  $\sigma$  della gaussiana considerata, nell'intervallo

- ▶  $[\mu - \sigma, \mu + \sigma]$  si trova circa il 68% delle misure, ossia

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \simeq 0.68$$

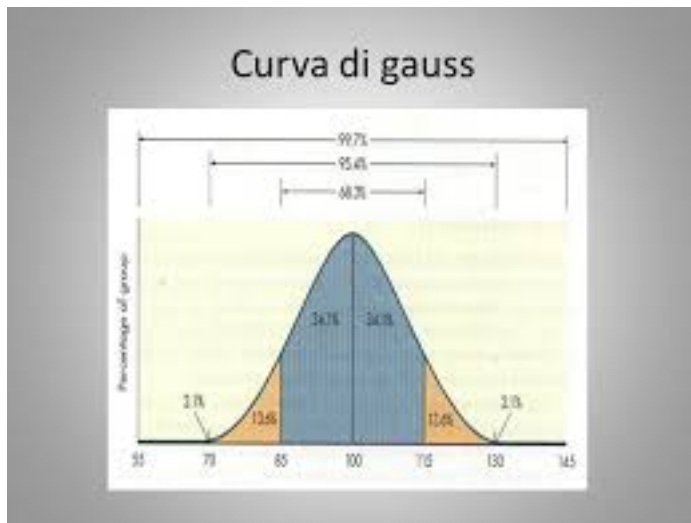
- ▶  $[\mu - 2\sigma, \mu + 2\sigma]$  si trova circa il 95% delle misure, ossia

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \simeq 0.95$$

- ▶  $[\mu - 3\sigma, \mu + 3\sigma]$  si trova circa il 99.7% delle misure, ossia

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \simeq 0.997$$

# Esempio di gaussiane



# Proprietà di distribuzioni Gaussiane

Indicata con  $\mathcal{N}(\mu, \sigma)$  una distribuzione Normale di media  $\mu$  e scarto quadratico medio  $\sigma$ , risulta che

- ▶ posto  $z =: \frac{x - \mu}{\sigma}$ , la distribuzione associata è Normale standard (ossia di media 0 e varianza/scarto quadratico medio 1)

# Proprietà di distribuzioni Gaussiane

Indicata con  $\mathcal{N}(\mu, \sigma)$  una distribuzione Normale di media  $\mu$  e scarto quadratico medio  $\sigma$ , risulta che

- ▶ posto  $z =: \frac{x - \mu}{\sigma}$ , la distribuzione associata è Normale standard (ossia di media 0 e varianza/scarto quadratico medio 1)
- ▶ La somma di distribuzioni Normali è Normale di media la somma delle medie e di varianza

# Proprietà di distribuzioni Gaussiane

Indicata con  $\mathcal{N}(\mu, \sigma)$  una distribuzione Normale di media  $\mu$  e scarto quadratico medio  $\sigma$ , risulta che

- ▶ posto  $z =: \frac{x - \mu}{\sigma}$ , la distribuzione associata è Normale standard (ossia di media 0 e varianza/scarto quadratico medio 1)
- ▶ La somma di distribuzioni Normali è Normale di media la somma delle medie e di varianza



## Referenze bibliografiche

Capitolo 1 del testo

### **Metodi Matematici e Statistici nelle Scienze della Terra**

Volume terzo: Tecniche statistiche

A. Buccianti, F. Rosso, F. Vlacci

Liguori Editore 2003

Capitolo 10 del testo

### **Matematica. Comprendere e interpretare fenomeni delle scienze della vita**

V Edizione

G. Gentili, V. Villani

McGraw Hill 2012

### **Metodi Matematici e Statistici per le Scienze Applicate**

G. Prodi

McGraw Hill 1992