

# Statistica della Formazione

## Slides 3

A.A. 2025-2026

Docente: ANNA LINA SARRA

# OBIETTIVI DI APPRENDIMENTO

**Definire** le misure di tendenza centrale e spiegare a cosa servono in ambito statistico

**Distinguere** moda, mediana e media aritmetica e indicare per quali tipi di carattere ciascuna può essere calcolata

**Calcolare** la moda per distribuzioni qualitative sconnesse, qualitative ordinabili e quantitative (discrete e in classi)

**Calcolare** la mediana per N dispari e N pari, sia per caratteri quantitativi che qualitativi ordinabili

**Calcolare** quartili e decili a partire da frequenze percentuali cumulate e interpretarne il significato

**Calcolare** la media aritmetica per distribuzioni disaggregate, a modalità singole e raggruppate in classi

**Applicare** la media aritmetica ponderata quando le unità statistiche hanno peso diverso

**Riconoscere** la presenza di valori anomali (outlier) e spiegare perché la mediana è più robusta della media

# Modulo 1: elementi di statistica descrittiva

- **Le misure di tendenza centrale**

“A statistician can have his head in an oven and his feet in ice, and he will say that on the average he feels fine.”

# Misure di tendenza centrale:

## Medie

- Le medie sono lo strumento con cui si **sintetizzano** i dati statistici.
- L'uso della media consente all'individuo di rappresentarsi mentalmente l'“**ordine di grandezza**” di un fenomeno, di effettuare **comparazioni** tra le manifestazioni di uno stesso fenomeno in tempi, luoghi o situazioni diverse, di comunicare ad altri tale informazione.

**PROPRIETA' DI INTERNALITA'**: se  $a$  e  $b$  sono il minimo e il massimo dell'insieme dei numeri  $x_1, x_2, \dots, x_N$ , la media è compresa tra queste due quantità:  $a \leq m \leq b$

Medie che è possibile calcolare in relazione ai diversi tipi di carattere

|              |            | Indici di sintesi ed Operazioni |                                                                            |                                                                                            |
|--------------|------------|---------------------------------|----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
|              |            | Moda                            | Statistiche d'ordine<br>(Mediana, Quartili, Decili, Percentili, Quantili ) | Medie algebriche<br>(media aritmetica, media armonica, media geometrica, media quadratica) |
| Caratteri    |            | =, ≠                            | > , <                                                                      | +, -, *, /                                                                                 |
| Qualitativi  | sconnessi  | sì                              | no                                                                         | no                                                                                         |
|              | ordinabili | sì                              | sì                                                                         | no                                                                                         |
| Quantitativi |            | sì                              | sì                                                                         | sì                                                                                         |

# Moda

La **moda** di un collettivo, distribuito secondo un carattere di qualsiasi natura, è la modalità prevalente del carattere ossia quella **modalità a cui è associata la massima frequenza**.

Quando il carattere è quantitativo e le modalità sono **raggruppate in classi**, si parla di **classe modale** con riferimento alla classe avente la **densità di frequenza più elevata**.

Si possono avere distribuzioni *unimodali, bimodali, multimodali e zeromodali*

# Moda esempio (1)

## **Stato estero di cittadinanza alunni stranieri**

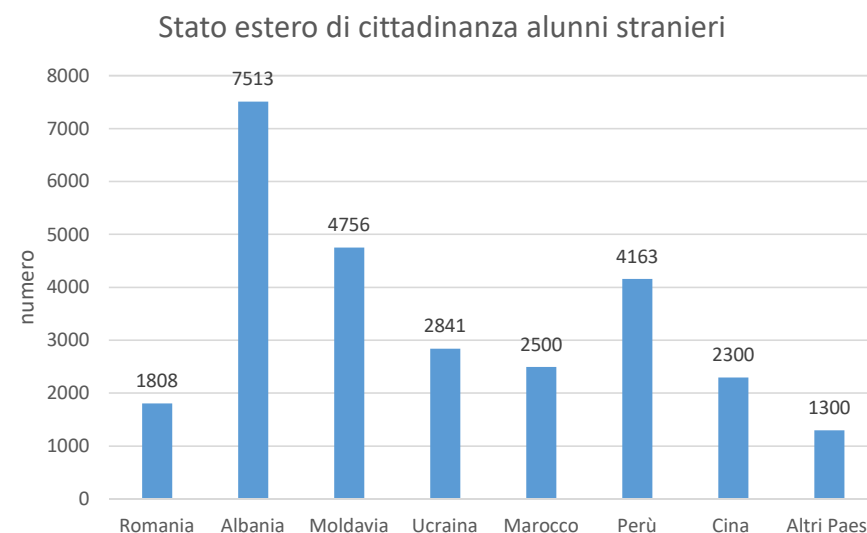
| <b>Cittadinanza</b> | <b>numero</b> |
|---------------------|---------------|
| Romania             | 1808          |
| Albania             | 7513          |
| Moldavia            | 4756          |
| Ucraina             | 2841          |
| Marocco             | 2500          |
| Perù                | 4163          |
| Cina                | 2300          |
| Altri Paesi         | 1300          |
| <b>Totale</b>       | <b>27181</b>  |

### Distribuzione unimodale

Moda: **Albania**

Frequenza massima: **7513**

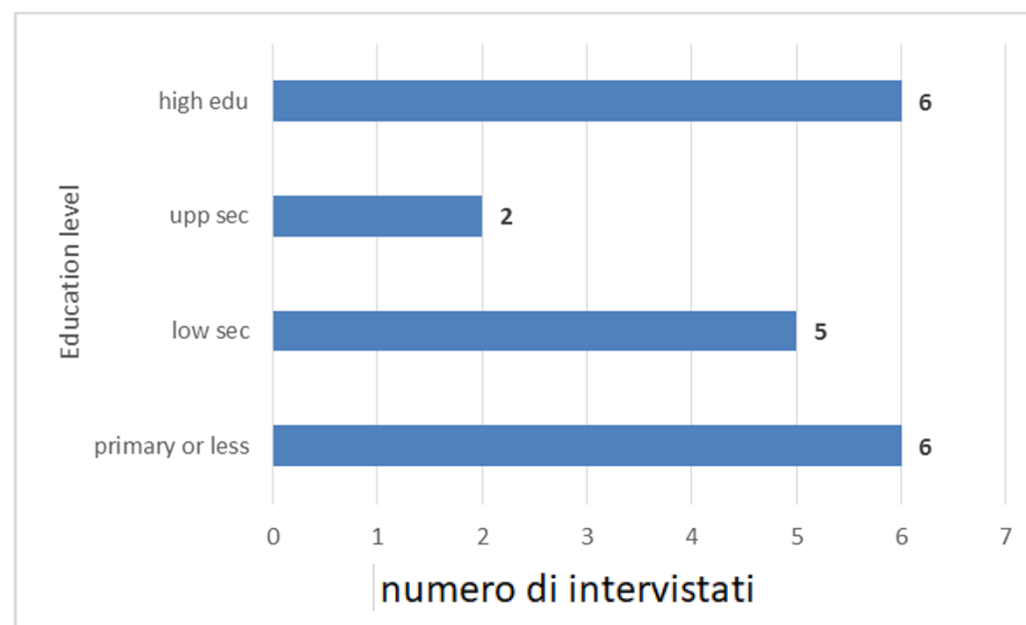
## Carattere qualitativo sconnesso



## Moda esempio (2)

*Carattere qualitativo ordinabile*

| Education level | frequenza assoluta |
|-----------------|--------------------|
| primary or less | 6                  |
| low sec         | 5                  |
| upp sec         | 2                  |
| high edu        | 6                  |
| totale          | 19                 |



### Distribuzione bimodale

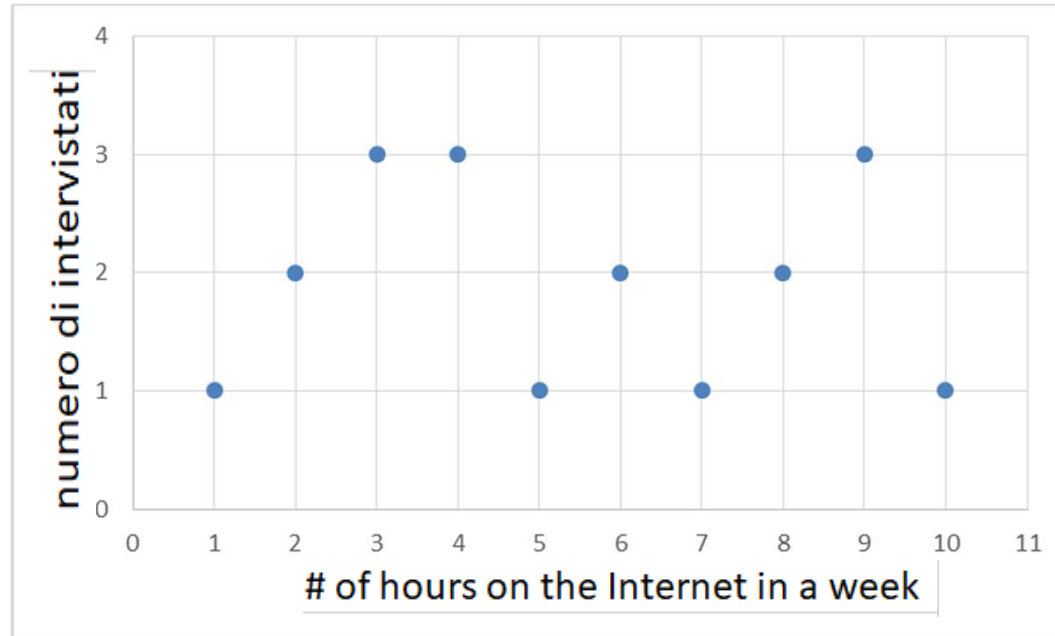
Mode: **primary or less; high education**

Frequenza massima: **6**

# Moda esempio (3)

*Carattere quantitativo discreto*

| # OF HOURS ON THE INTERNET IN A WEEK | frequenza assoluta |
|--------------------------------------|--------------------|
| 1                                    | 1                  |
| 2                                    | 2                  |
| 3                                    | 3                  |
| 4                                    | 3                  |
| 5                                    | 1                  |
| 6                                    | 2                  |
| 7                                    | 1                  |
| 8                                    | 2                  |
| 9                                    | 3                  |
| 10                                   | 1                  |
| totale                               | 19                 |



Distribuzione trimodale

Moda: 3; 4; 9

Frequenza massima: 3

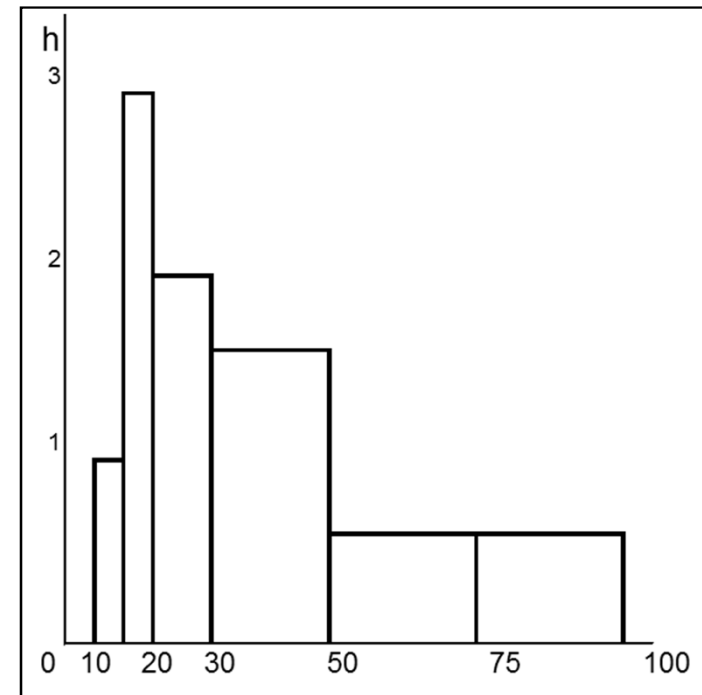
# Carattere quantitativo continuo in classi: Classe modale

| Classi di peso<br>(in Kg) | Frequenza<br>assoluta | Ampiezza<br>di classe | Densità di<br>frequenza |
|---------------------------|-----------------------|-----------------------|-------------------------|
| $c_{i-1}   - c_i$         | $n_i$                 | $d_i$                 | $h_i$                   |
| 10   -- 15                | 5                     | 5                     | 1                       |
| <b>15   -- 20</b>         | 15                    | 5                     | <b>3</b>                |
| 20   -- 30                | 20                    | 10                    | 2                       |
| 30   -- 50                | 30                    | 20                    | 1,5                     |
| 50   -- 75                | 15                    | 25                    | 0,6                     |
| 75   -- 100               | 15                    | 25                    | 0,6                     |
| <b>Totale</b>             | <b>100</b>            |                       |                         |

Densità  
Massima

$$h_i = \frac{n_i}{d_i}$$

$$d_i = c_i - c_{i-1}$$



Per determinare la moda è necessario calcolare la densità di frequenza  $h_i$ , data dal rapporto fra frequenza assoluta  $n_i$  e ampiezza di classe  $d_i$ .

La classe modale è la classe con maggiore densità di frequenza

## Distribuzione unimodale

Classe modale: 15 | -- 20

Densità di frequenza massima:

**3**

# Mediana

La mediana è quella modalità che suddivide ogni **distribuzione ordinata** in due distribuzioni aventi ciascuna una numerosità (o una quantità) che è il 50% della numerosità (o della quantità) della distribuzione totale.

Si può calcolare per i caratteri qualitativi ordinabili e quantitativi.

Sia  $x_1, x_2, \dots, x_N$ , una distribuzione statistica disaggregata.

Sia  $y_1, y_2, \dots, y_N$ , con  $y_1 \leq y_2 \leq \dots \leq y_N$ , la corrispondente distribuzione dei termini ordinati.

## Mediana per N dispari

➤ N dispari : la mediana è la modalità che nella distribuzione ordinata occupa il posto

$$\frac{N+1}{2}$$

*Distribuzione disaggregata dell'età di 9 intervistati*

$x_1=21$   $x_2=18$   $x_3=28$   $x_4=27$   $x_5=30$   $x_6=28$   $x_7=30$   $x_8=25$   $x_9=22$

Distribuzione ordinata

$y_1=18$   $y_2=21$   $y_3=22$   $y_4=25$   $y_5=27$   $y_6=28$   $y_7=28$   $y_8=30$   $y_9=30$



Posizione occupata dalla mediana:

$$\frac{N+1}{2} = \frac{9+1}{2} = 5$$

Mediana:

$$y_{\frac{N+1}{2}} = y_5 = 27$$

## Mediana per N pari

N pari : si hanno due modalità mediane che nella distribuzione ordinata occupano rispettivamente i posti

$$\frac{N}{2} \text{ e } \frac{N}{2} + 1$$

*Distribuzione disaggregata dell'età di 10 intervistati*

$$x_1=21 \ x_2=18 \ x_3=28 \ x_4=27 \ x_5=30 \ x_6=28 \ x_7=30 \ x_8=25 \ x_9=22 \ x_{10}=28$$

Distribuzione ordinata

$$y_1=18 \ y_2=21 \ y_3=22 \ y_4=25 \ y_5=27 \ y_6=28 \ y_7=28 \ y_8=28 \ y_9=30 \ y_{10}=30$$

Posizioni occupate dalle mediane

$$\frac{N}{2} = \frac{10}{2} = 5; \frac{N}{2} + 1 = 5 + 1 = 6$$

Mediane

$$y_{\frac{N}{2}} = y_5 = 27; y_{\frac{N}{2}+1} = y_6 = 28$$

Mediana

$$\frac{y_{\frac{N}{2}} + y_{\frac{N}{2}+1}}{2} = \frac{27 + 28}{2} = 27.5$$

MODULO 1. Elementi di statistica descrittiva

Se **N è pari**, e il carattere è quantitativo si può assumere come mediana la media aritmetica dei termini che occupano le due posizioni centrali della graduatoria dei termini ordinati, ossia le posizioni  $N/2$  e  $N/2 + 1$ .

## Esempio: Mediana per caratteri qualitativi ordinabili - N dispari

Distribuzione disaggregata:

Sufficiente, Pessimo, Insufficiente, Ottimo, Sufficiente, Insufficiente, Ottimo

Distribuzione ordinata:

Pessimo, Insufficiente, Insufficiente, Sufficiente, Sufficiente, Ottimo, Ottimo

Posizione occupata dalla mediana

$$\frac{N+1}{2} = \frac{7+1}{2} = 4$$

Mediana:

$$y_{\frac{N+1}{2}} = y_4 = \text{Sufficiente}$$

## Esempio: Mediana per caratteri qualitativi ordinabili - N pari

Distribuzione disaggregata:

Licenza media, Diploma, Diploma, Laurea, Licenza media, Licenza elementare

Distribuzione ordinata:

Licenza elementare, Licenza media, Licenza media, Diploma, Diploma, Laurea,

Posizioni occupate dalle mediana

$$\frac{N}{2} = \frac{6}{2} = 3; \quad \frac{N}{2} + 1 = \frac{6}{2} + 1 = 4;$$

Mediane

$$y_{\frac{N+1}{2}} = y_3 = \text{Licenza media};$$

$$y_{\frac{N}{2}+1} = y_4 = \text{Diploma}$$

Carattere quantitativo: La somma degli scarti in valore assoluto dei valori  $x_1, x_2, \dots, x_N$  da una costante  $c$  è minima quando  $c$  è uguale alla mediana

|   |   |   |   |    |
|---|---|---|---|----|
| 1 | 3 | 5 | 6 | 11 |
|---|---|---|---|----|

Me=5

$$\sum_{i=1}^5 |x_i - Me| = |1 - 5| + |3 - 5| + |5 - 5| + |6 - 5| + |11 - 5| = \\ = 4 + 2 + 0 + 1 + 6 = 13$$

Se al posto di 5 (che corrisponde alla mediana dei valori) metto un qualsiasi altro valore, questa somma sarà sempre  $>13$

# Oltre la mediana....

## *Ma che sono i percentili e i quartili!?*

Un **X Percentile** è la modalità che lascia alla propria sinistra (prima di sé, al di sotto di sé, a sinistra di sé ecc.) una **percentuale X** dei dati

Example: You are the fourth tallest person in a group of 20

80% of people are shorter than you:



That means you are at the **80th percentile**.

If your height is 1.85m then "1.85m" is the 80th percentile height in that group.

# Oltre la mediana....

## *Ma che sono i percentili e i quartili!?*

Un **quartile** è la modalità che lascia alla propria sinistra (prima di sé, al di sotto di sé, a sinistra di sé ecc.) una **percentuale** dei dati pari a **25%, 50 % o 75%**.

Si chiama quartile proprio perché divide in 4 parti di pari frequenza la distribuzione.

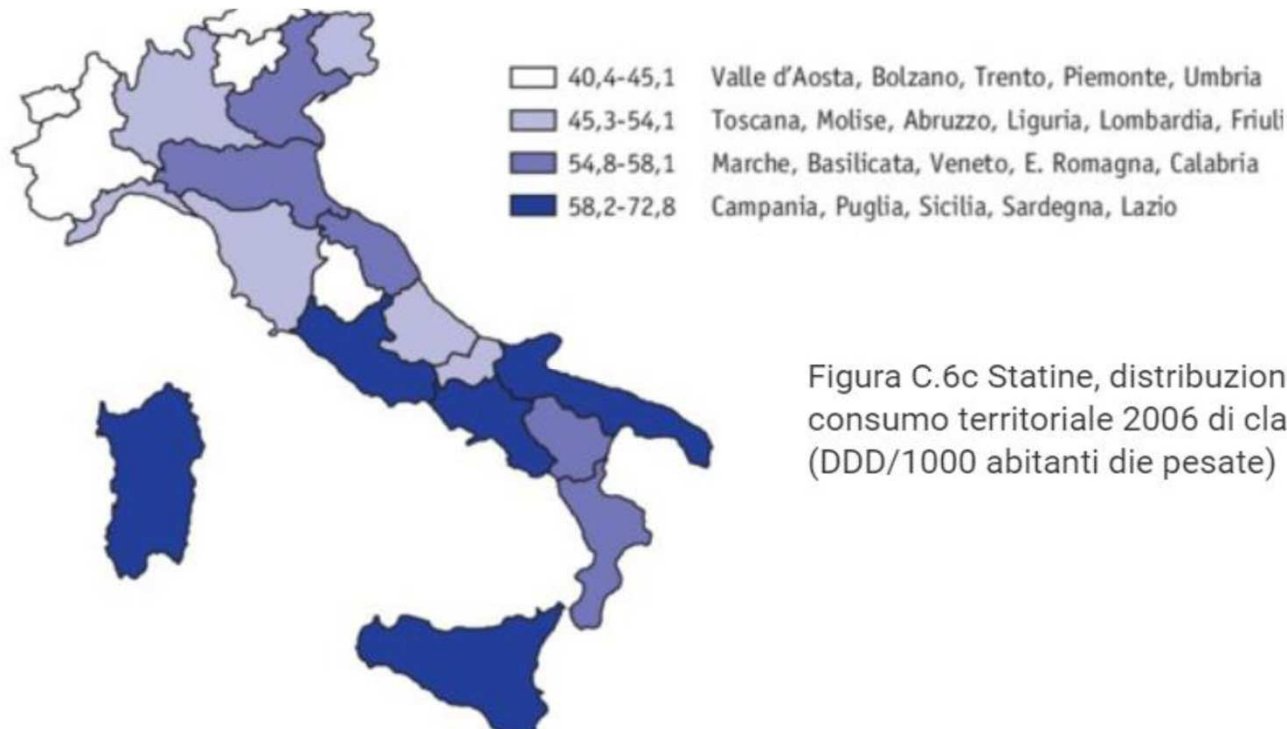


Figura C.6c Statine, distribuzione in quartili del consumo territoriale 2006 di classe A-SSN (DDD/1000 abitanti die pesate)

# Quartili

Sia  $x_1, x_2, \dots, x_N$  una distribuzione disaggregata.

Sia  $y_1, y_2, \dots, y_N$  la corrispondente distribuzione di termini ordinati, con  $y_1 \leq y_2 \leq \dots \leq y_N$ .

- Il **primo quartile**,  $q_1$ , è la quantità che non è superata da un quarto (25%) dei termini ordinati della distribuzione
- Il **secondo quartile**,  $q_2$ , è la quantità che non è superata dalla metà (50%) dei termini ordinati.
- Il **terzo quartile**,  $q_3$ , è la quantità che non è superata dai tre quarti (75%) dei termini ordinati della distribuzione.

**N.B.: Il secondo quartile coincide con la mediana**

# Troviamo le posizioni dei quartili

La formula base di tutti i quartili è:  
1 ( $Q_1$ ), 2 ( $Q_2$ ) e 3 ( $Q_3$ )

$$X\left(\frac{N+1}{4}\right)$$

poi di volta in volta si moltiplica per

Primo quartile  $\rightarrow Q_1 \rightarrow X\left(\frac{N+1}{4} * 1\right)$

Secondo quartile  $\rightarrow Q_2 \rightarrow$  Mediana  $\rightarrow X\left(\frac{N+1}{4} * 2\right) = X\left(\frac{N+1}{2}\right)$

Terzo quartile  $\rightarrow Q_3 \rightarrow X\left(\frac{N+1}{4} * 3\right)$

# Decili

In termini discorsivi, i decili si possono definire come medie di posizione tali che:

**Il primo decile:** è la quantità che non è superata da un decimo (10%) dei termini ordinati

**Il secondo decile:** è la quantità che non è superata da due decimi (20%) dei termini ordinati

...

**N.B.: i decili sono 9.**

**N.B.: Il quinto decile coincide con la mediana**

## Mediana e quartili: definizione operativa basata sulle frequenze percentuali cumulate

|         | $x_i$         | $n_i$      | $N_i$ | $P_i$  |     |
|---------|---------------|------------|-------|--------|-----|
|         | 41            | 3          | 3     | 2.1%   |     |
|         | 42            | 2          | 5     | 3.6%   |     |
|         | 43            | 6          | 11    | 7.9%   |     |
|         | 44            | 11         | 22    | 15.7%  |     |
|         | 45            | 8          | 30    | 21.4%  |     |
| $q_1$   | <b>46</b>     | 17         | 47    | 33.6%  | 25% |
|         | 47            | 21         | 68    | 48.6%  |     |
| $m=q_2$ | <b>48</b>     | 14         | 82    | 58.6%  | 50% |
|         | 49            | 17         | 99    | 70.7%  |     |
| $q_3$   | <b>50</b>     | 15         | 114   | 81.4%  | 75% |
|         | 51            | 10         | 124   | 88.6%  |     |
|         | 52            | 10         | 134   | 95.7%  |     |
|         | 53            | 5          | 139   | 99.3%  |     |
|         | 54            | 1          | 140   | 100.0% |     |
|         | <b>Totale</b> | <b>140</b> |       |        |     |

## Decili: definizione operativa basata sulle frequenze percentuali cumulate

|                       | $x_i$         | $n_i$      | $N_i$ | $P_i$  |         |
|-----------------------|---------------|------------|-------|--------|---------|
|                       | 41            | 3          | 3     | 2.1%   |         |
|                       | 42            | 2          | 5     | 3.6%   |         |
|                       | 43            | 6          | 11    | 7.9%   | 10%     |
| $d_1 \rightarrow$     | 44            | 11         | 22    | 15.7%  | 20%     |
| $d_2 \rightarrow$     | 45            | 8          | 30    | 21.4%  | 30%     |
| $d_3 \rightarrow$     | 46            | 17         | 47    | 33.6%  | 40%     |
| $d_4 \rightarrow$     | 47            | 21         | 68    | 48.6%  |         |
| $m=d_5 \rightarrow$   | 48            | 14         | 82    | 58.6%  | 60%;70% |
| $d_6;d_7 \rightarrow$ | 49            | 17         | 99    | 70.7%  | 80%     |
| $d_8 \rightarrow$     | 50            | 15         | 114   | 81.4%  | 90%     |
|                       | 51            | 10         | 124   | 88.6%  |         |
| $d_9 \rightarrow$     | 52            | 10         | 134   | 95.7%  |         |
|                       | 53            | 5          | 139   | 99.3%  |         |
|                       | 54            | 1          | 140   | 100.0% |         |
|                       | <b>Totale</b> | <b>140</b> |       |        |         |

## Esempio

Distribuzione di frequenze dei voti di 150 studenti all'esame di statistica

|              | $x_i$  | $n_i$ | $N_i$ | $P_i$  |     |
|--------------|--------|-------|-------|--------|-----|
|              | 18     | 3     | 3     | 2.0%   |     |
|              | 19     | 2     | 5     | 3.3%   |     |
| $d_1=20.5$   | 20     | 10    | 15    | 10.0%  |     |
|              | 21     | 7     | 22    | 14.7%  | 20% |
| $d_2$        | 22     | 11    | 33    | 22.0%  | 30% |
| $d_3$        | 23     | 13    | 46    | 30.7%  | 40% |
| $m=d_5=24.5$ | 24     | 29    | 75    | 50.0%  | 60% |
|              | 25     | 17    | 92    | 61.3%  |     |
|              | 26     | 8     | 100   | 66.7%  | 70% |
| $d_7$        | 27     | 15    | 115   | 76.7%  |     |
| $d_8$        | 28     | 10    | 125   | 83.3%  | 80% |
|              | 29     | 10    | 135   | 90.0%  |     |
| $d_9=29.5$   | 30     | 15    | 150   | 100.0% |     |
|              | Totale | 150   |       |        |     |

# Media aritmetica

La media aritmetica è quel valore di sintesi che sostituito alle modalità lascia inalterata la loro somma

- Insieme alle percentuali e ai grafici, la media aritmetica è lo strumento statistico più largamente utilizzato
- La media aritmetica di una **distribuzione statistica disaggregata** si calcola come la somma dei termini  $x_1, x_2, \dots, x_N$  divisa per  $N$

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

## Media aritmetica per una distribuzione disaggregata: calcolo

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i.$$

Serie storica degli studenti stranieri

| 2009/2010 | 2010/2011 | 2011/2012 | 2012/2013 | 2013/2014 | 2014/2015 |
|-----------|-----------|-----------|-----------|-----------|-----------|
| 673800    | 710263    | 755939    | 786630    | 803053    | 814208    |

La media annua degli studenti stranieri

$$\mu = \frac{673800+710263+755939+786630+803053+814208}{6} = 739036,1$$

# Media aritmetica per le distribuzioni di frequenze

**Modalità  
singole**

$$\mu = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{N} = \frac{1}{N} \sum_{i=1}^k x_i \cdot n_i$$
$$= x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k = \sum_{i=1}^k x_i \cdot f_i$$

**Modalità  
raggruppate in  
classi**

$$\mu = \frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2 + \dots + \bar{x}_k \cdot n_k}{N} = \frac{1}{N} \sum_{i=1}^k \bar{x}_i \cdot n_i$$

dove  $\bar{x}_i = \frac{C_{i-1} + C_i}{2}$  è il **valore centrale** della generica classe.

Media aritmetica per una distribuzione di frequenze a modalità singole: calcolo

$$\mu = \frac{1}{N} \sum_{i=1}^k x_i \cdot n_i$$

| # OF HOURS ON<br>THE INTERNET IN<br>A WEEK | frequenza<br>assoluta | modalità*<br>frequenza |
|--------------------------------------------|-----------------------|------------------------|
| $x_i$                                      | $n_i$                 | $x_i \cdot n_i$        |
| 1                                          | 1                     | 1                      |
| 2                                          | 2                     | 4                      |
| 3                                          | 3                     | 9                      |
| 4                                          | 3                     | 12                     |
| 5                                          | 1                     | 5                      |
| 6                                          | 2                     | 12                     |
| 7                                          | 1                     | 7                      |
| 8                                          | 2                     | 16                     |
| 9                                          | 3                     | 27                     |
| 10                                         | 1                     | 10                     |
| totale                                     | 19                    | 103                    |

□ La media aritmetica della distribuzione è data da:

$$\mu = \frac{1 \cdot 1 + 2 \cdot 2 + \dots + 10 \cdot 1}{140}$$

$$= \frac{103}{19} = 5.42$$

Media aritmetica per una distribuzione di frequenze a modalità raggruppate in classi: calcolo

$$\mu = \frac{1}{N} \sum_{i=1}^k \bar{x}_i \cdot n_i$$

### Distribuzione di frequenze

secondo l'età:

| Classe | Valore centrale $\bar{x}_i$ | $n_i$ | $\bar{x}_i \cdot n_i$ |
|--------|-----------------------------|-------|-----------------------|
| 19-21  | 20.0                        | 31    | 620                   |
| 21-24  | 22.5                        | 45    | 1012.5                |
| 24-27  | 25.5                        | 5     | 127.5                 |
| 27-30  | 28.5                        | 1     | 28.5                  |
| Totale |                             | 82    | 1788.5                |

□ La media aritmetica della distribuzione è data da:

$$\mu = \frac{20.0 \cdot 31 + 22.5 \cdot 45 + 25.5 \cdot 5 + 28.5 \cdot 1}{82}$$
$$= \frac{1788.5}{82} = 21.8$$

La somma algebrica degli scarti dalla media aritmetica è uguale a zero

|   |   |   |   |
|---|---|---|---|
| 1 | 6 | 4 | 1 |
|---|---|---|---|

$$\sum_{i=1}^4 (x_i - \bar{x}) = (1-3) + (6-3) + (4-3) + (1-3) = -2 + 3 + 1 - 2 = 0$$

| Voto esame di matematica ( $x_i$ ) | $n_i$     | $x_i n_i$ | $(x_i - \mu)$        | $(x_i - \mu) * n_i$ |
|------------------------------------|-----------|-----------|----------------------|---------------------|
| 18                                 | 5         | 90        | $18 - 24,76 = -6,76$ | -33,8               |
| 20                                 | 7         | 140       | $20 - 24,76 = -4,76$ | -33,32              |
| 23                                 | 8         | 184       | $23 - 24,76 = -1,76$ | -14,08              |
| 26                                 | 8         | 208       | $26 - 24,76 = 1,24$  | 9,92                |
| 27                                 | 10        | 270       | $27 - 24,76 = 2,24$  | 22,4                |
| 28                                 | 7         | 196       | $28 - 24,76 = 3,24$  | 22,68               |
| 30                                 | 5         | 150       | $30 - 24,76 = 5,24$  | 26,2                |
| <b>Totale</b>                      | <b>50</b> | 1238      |                      | <b>0</b>            |

$$\sum_{i=1}^7 (x_i - \bar{x}) n_i = 0$$

La somma degli scarti dalla media aritmetica al quadrato è un minimo

|   |   |   |   |
|---|---|---|---|
| 1 | 6 | 4 | 1 |
|---|---|---|---|

$$\mu=3$$

$$\sum_{i=1}^4 (x_i - \mu)^2 = (1-3)^2 + (6-3)^2 + (4-3)^2 + (1-3)^2 = 4 + 9 + 1 + 4 = 18$$

Se al posto di 3, che corrisponde alla media, metto un qualsiasi altro valore, questa somma sarà sempre >18

- **Proprietà di linearità** : Se ogni singolo termine della distribuzione,  $x_i$ , viene sottoposto alla trasformazione

$$a + bx_i,$$

con  $a$  costante qualsiasi e  $b \neq 0$ , la nuova media è legata a quella originaria dalla medesima trasformazione  $a + b\mu$

|                                                                                   |     |    |     |     |            |
|-----------------------------------------------------------------------------------|-----|----|-----|-----|------------|
| Voti ottenuti su 4 esami                                                          | 18  | 22 | 26  | 28  | $\mu=23.5$ |
| Se ad ogni esame lo studente avesse ottenuto 2 punti in più la media sarebbe 25.5 | 20  | 24 | 28  | 30  | $\mu=25.5$ |
| Guadagno su 4 giorni                                                              | 50  | 25 | 75  | 50  | $\mu=50$   |
| Se ogni giorno il guadagno fosse il doppio la media sarebbe 100                   | 100 | 50 | 150 | 100 | $\mu=100$  |

- **Proprietà di associatività** Se un collettivo statistico di  $N$  unità viene suddiviso in  $L$  sottoinsiemi disgiunti aventi numerosità  $N^{(1)}, N^{(2)}, \dots, N^{(L)}$  e medie  $\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(L)}$ , allora la media aritmetica del collettivo può essere così calcolata

$$\mu = \frac{1}{N} (\mu^{(1)} \cdot N^{(1)} + \mu^{(2)} \cdot N^{(2)} + \dots + \mu^{(L)} \cdot N^{(L)}) = \frac{1}{N} \sum_{l=1}^L \mu^{(l)} \cdot N^{(l)}$$

Durata media in minuti delle interrogazioni: 12.8, 13.0, 13.4, 13.4, 13.6, 13.5, 13.6, 13.7

$$\mu = \frac{12.8 + 13.0 + 13.4 + 13.4 + 13.6 + 13.5 + 13.6 + 13.7}{8} = 13.375 .$$

se suddividiamo la distribuzione data nelle due seguenti:

A. 12.8, 13.0, 13.4, 13.4, 13.6     $\mu_A = 13.240$

B. 13.5, 13.6, 13.7     $\mu_B = 13.600$

la media aritmetica della distribuzione **può essere ottenuta come**

$$\mu = \frac{13.240 \cdot 5 + 13.600 \cdot 3}{8} = 13.375 .$$

# Media aritmetica ponderata

Siano  $x_1, x_2, \dots, x_k$  le osservazioni e  $w_1, w_2, \dots, w_k$  i rispettivi pesi. Allora, la **media aritmetica ponderata** di  $x_1, x_2, \dots, x_k$  è data dal rapporto tra la somma delle osservazioni moltiplicate per i rispettivi pesi e la somma dei pesi

$$\mu = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_k \cdot w_k}{w_1 + w_2 + \dots + w_k} = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}$$

Media aritmetica ponderata:  
esempio

$$\mu = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}$$

| voto<br>esame | CFU   | <i>modalità*<br/>peso</i> |
|---------------|-------|---------------------------|
| $x_i$         | $w_i$ | $x_i * w_i$               |
| 24            | 9     | 216                       |
| 26            | 6     | 156                       |
| 30            | 12    | 360                       |
| 30            | 6     | 180                       |
| 22            | 9     | 198                       |
| 18            | 12    | 216                       |
| 18            | 6     | 108                       |
| totale        | 60    | 1434                      |

$$\mu = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i} = \frac{1434}{60} = 23.9$$

## Media aritmetica ponderata (1)

Poniamo il caso di avere i dati delle regioni italiane relative alle prove Invalsi di Italiano delle regioni settentrionali:

|                |     |
|----------------|-----|
| V Aosta        | 201 |
| Piemonte       | 203 |
| Liguria        | 200 |
| Lombardia      | 200 |
| Bolzano        | 196 |
| Trento         | 200 |
| Veneto         | 200 |
| Friuli-Ven.G.  | 202 |
| Emilia-Romagna | 198 |

Calcoliamo media Nord Italia: sommiamo valori e dividiamo per il numero di regioni

$$201 + 203 + 200...=1800 \longrightarrow 1800/N=1800/9=200$$

$$\mu = \frac{\sum x_i}{N} = 200$$

## Media aritmetica ponderata (2)

Un ricercatore dice che in realtà la media del test Invalsi (prova di italiano) del Nord Italia è maggiore (seppur di poco) rispetto a quella da noi calcolata, considerando tutti gli studenti nel loro complesso:

$$\mu = \frac{\sum x_i}{N} = 200$$

$$\mu = 200,18$$

Come mai? Che cosa è successo? Abbiamo dimenticato qualcosa di utile...?

## Media aritmetica ponderata (3)

| <b>Regione</b> | <b>Voto</b> | <b>Popolazione in milioni</b> |
|----------------|-------------|-------------------------------|
| V Aosta        | 201         | 0,1                           |
| Piemonte       | 203         | 4,4                           |
| Liguria        | 200         | 1,5                           |
| Lombardia      | 200         | 10                            |
| Bolzano        | 196         | 0,5                           |
| Trento         | 200         | 0,5                           |
| Veneto         | 200         | 4,9                           |
| Friuli-Ven.G.  | 202         | 1,2                           |
| Emilia-Romagna | 198         | 4,4                           |

Quando si calcola una media **ponderata (o pesata)** ogni unità statistica **può** avere peso differente, a seconda del volume, della popolazione, del reddito ecc.

In questo caso ponderiamo rispetto alla **popolazione**, se fosse stato uno studio sul numero di alberi magari avremmo ponderato sui km quadri di aree verdi ecc.

# Meglio mediana o media?

Ci dicono cose diverse, spesso vengono utilizzate entrambe  
A volte coincidono, a volte sono vicine, a volte sono molto distanti...

La media risponde alla domanda «se dovessi assegnare ad ogni individuo lo stesso valore per ottenere la stessa somma totale, quale valore dovrei assegnare?»

La mediana risponde alla domanda «se dovessi trovare il valore che si trova al centro e che divide in due parti di uguale frequenza i dati, a quale valore dovrei guardare?»

# Valori anomali, o outlier

Sono valori **decisamente diversi** da tutti gli altri, o perché troppo grandi o perché troppo piccoli

Possono essere frutto di:

- ✓ **Errata digitazione o differente scala di misura** (**valori anomali erronei**) Esempio: volevo scrivere 10 e ho scritto 100 digitando sul foglio excel, ho messo altezza in cm a tutti e ad una sola persona in metri ecc.
- ✓ **Valori di individui con caratteristiche particolari** (**valori anomali in senso stretto**). Esempio: analisi sui redditi delle famiglie italiane e mi capita in un campione di 100 famiglie la famiglia Agnelli...

Se ci si accorge di **valori anomali erronei** tendenzialmente li si corregge/elimina, altrimenti **valori anomali in senso stretto** restano inclusi nell'analisi.

Si cerca di utilizzare indici in grado di «contenere» comunque questi outlier, in modo che non sballino tutte le analisi.

## Mediana più robusta di media!

Quando un indice è particolarmente in grado di «reggere» alla presenza di valori anomali, allora esso si dire **robusto**.

Poniamo il caso di avere 10 valori: 2-4-5-5-6-7-8-9-9-106

A quanto è uguale la mediana?



$$\text{Mediana} = X_{\left(\frac{N+1}{2}\right)} = X_{5,5} = \frac{6+7}{2} = 6,5$$

A quanto è uguale la media?

$$\frac{1}{N} \sum_{i=1}^n x_i = 161/10 = 16,1$$

### **DOMANDA**

Quale dei due è stato influenzato dall'**outlier** 106? Che spiegazione ne diamo?

Un «meme» statistico...per ricordarvi della robustezza!



# Media, moda o mediana? Perché non tutte e 3?



|                       | Pro                                                               | Contro                                                                |
|-----------------------|-------------------------------------------------------------------|-----------------------------------------------------------------------|
| <u><b>MEDIA</b></u>   | Considera tutte le osservazioni (completezza)                     | Considera tutte le osservazioni (robustezza)                          |
| <u><b>MEDIANA</b></u> | 1. Robustezza<br>2. Divide dati in due segmenti di pari frequenza | Tiene conto solo di 1 o 2 osservazioni centrali nel calcolo effettivo |
| <u><b>MODA</b></u>    | 1. Velocità<br>2. Indica dove si trova massima concentrazione     | Nel caso di tante modalità perde di significato                       |

# Domande (1)

- La distribuzione delle altezze degli adulti in Italia è unimodale?
- Data la distribuzione disaggregata  $x=\{21,16, 21,21,23,23,17\}$ , calcolare
  - a.  $x_1+x_3$
  - b. la media della distribuzione disaggregata
  - c. la mediana della distribuzione disaggregata
  - d. la moda della distribuzione disaggregata

## Domande (2)

Tasso di abbandono alla fine del primo biennio delle scuole secondarie superiori nelle regioni italiane (anno di riferimento 2012)

Calcolare i quartili della del tasso di abbandono scolastico per il periodo 2012 e interpretare i risultati.

| REGIONI                      | %    |
|------------------------------|------|
| Piemonte                     | 6,9  |
| Valle d'Aosta/Vallée d'Aoste | 10,8 |
| Lombardia                    | 6,6  |
| Trentino-Alto Adige/Südtirol | 3,2  |
| Veneto                       | 4,3  |
| Friuli-Venezia Giulia        | 4,6  |
| Liguria                      | 7,5  |
| Emilia-Romagna               | 6,6  |
| Toscana                      | 7,4  |
| Umbria                       | 4,7  |
| Marche                       | 4,8  |
| Lazio                        | 5,8  |
| Abruzzo                      | 6,1  |
| Molise                       | 5,7  |
| Campania                     | 9,3  |
| Puglia                       | 5,3  |
| Basilicata                   | 5,3  |
| Calabria                     | 5,3  |
| Sicilia                      | 9,2  |
| Sardegna                     | 10,4 |
| Bolzano/Bozen                | 2,9  |
| Trento                       | 3,6  |

\* Valori %, anno di riferimento 2012

## Domande (3)

| Abbandoni scolastici nella scuola secondaria, 2006–12 |                |                |               |               |               |               |               |                |
|-------------------------------------------------------|----------------|----------------|---------------|---------------|---------------|---------------|---------------|----------------|
| Numero                                                |                |                |               |               |               |               |               |                |
|                                                       | 2006           | 2007           | 2008          | 2009          | 2010          | 2011          | 2012          | totale         |
| Germany                                               | 64,865         | 64,093         | 55,905        | 50,965        | 49,622        | 50,791        | 47,667        | 383,908        |
| France                                                | 5,792          | 5,797          | 6,128         | 6,007         | 5,869         | 5,928         | 4,821         | 40,342         |
| Italy                                                 | 32,306         | 34,439         | 34,082        | 34,101        | 32,761        | 34,034        | 33,852        | 235,575        |
| <b>Totale</b>                                         | <b>102,963</b> | <b>104,329</b> | <b>96,115</b> | <b>91,073</b> | <b>88,252</b> | <b>90,753</b> | <b>86,340</b> | <b>659,825</b> |

- Calcolare ed interpretare le medie aritmetiche per ogni stato e per ogni anno.

## Domande (4)

- A seguito di un'analisi su 10 scuole, sono stati rilevati i punteggi medi ottenuti dagli studenti in un test standardizzato.
- Di queste:
- **1 scuola** ha ottenuto un punteggio medio molto elevato pari a **100**
- le altre **9 scuole** hanno ottenuto punteggi medi pari a **60**

Qual è l'indice che consente di sintetizzare meglio i risultati degli studenti?

- la **media**
- oppure la **mediana**?

## Domande (5)

- I voti riportati da uno studente in 5 esami sono 18, 25, 30, 26, 27. Che voto deve prendere al prossimo esame per avere
  - una media aritmetica dei voti uguale a 26?
  - un voto mediano uguale a 26?
  - un voto modale uguale a 26?
- La media aritmetica di 5 valori è 6. La media aritmetica di 3 di questi 5 valori è 8. Qual è la media dei restanti due valori?
- La media di 121 numeri è 59. Se ogni numero è moltiplicato per 4, quale sarà la media della nuova distribuzione? E se ad ogni numero sommiamo 10?



The average human has one breast  
and one testicle.

— *Des MacHale* —

AZ QUOTES