

# Promozione e diffusione della cultura statistica

## *La statistica e le nuove tecnologie a supporto della didattica*

Antonella Bianchino  
21 novembre 2013



# Premessa

Dopo aver raccolto e sistemato le informazioni relative ad un dato fenomeno, il passo successivo che viene compiuto è quello di individuare dei valori di sintesi che:

- rappresentino le caratteristiche del fenomeno
- consentano di effettuare confronti con valori riferiti a fenomeni analoghi osservati in tempi e luoghi diversi



# Premessa

- Nella ricerca scientifica e tecnologica così come nelle scienze sociali e politiche, è importante misurare la reale efficacia di interventi e modifiche sul sistema oggetto di studio.
- Per questo occorrono misure sintetiche che posizionino la distribuzione di frequenza di un certo fenomeno e consentano il passaggio da una pluralità di informazioni ad un solo numero



# Indici di posizione

- Obiettivo di una misura di posizione è quello di *sintetizzare in un singolo valore numerico l'intera distribuzione di frequenza per effettuare confronti nel tempo, nello spazio o tra circostanze differenti*
- La metodologia statistica mette a disposizione degli indicatori sintetici, detti indici di posizione, che possono essere utilizzati per sintetizzare in un valore significativo i molti valori di un fenomeno raccolti durante l'indagine



# Indici di posizione

Tra le diverse tecniche di sintesi dei dati proposti dalla statistica descrittiva, esamineremo:

- La moda
- La mediana
- I quantili
- La media aritmetica

La moda, la mediana e i quantili rientrano tra le medie lasche.



# Medie lasche

- Si definiscono medie lasche quelle che utilizzano, per la individuazione sintetica della posizione di una variabile, alcuni valori specifici della distribuzione di frequenza individuati sulla base della loro collocazione relativa rispetto a tutti gli altri, ma senza coinvolgere nel calcolo tutte le modalità della variabile
- Esse si ottengono a seguito della scelta, all'interno della serie ordinata o della distribuzione, di valori particolarmente rilevanti ai fini dell'interpretazione dell'intera serie o distribuzione



# Moda

- Si definisce moda **la modalità** che presenta la frequenza, assoluta o relativa, più elevata
- È un indicatore che può essere utilizzato anche per le mutabili statistiche
- Il calcolo della moda risulta immediato dall'osservazione della distribuzione o della sua rappresentazione grafica



# Moda

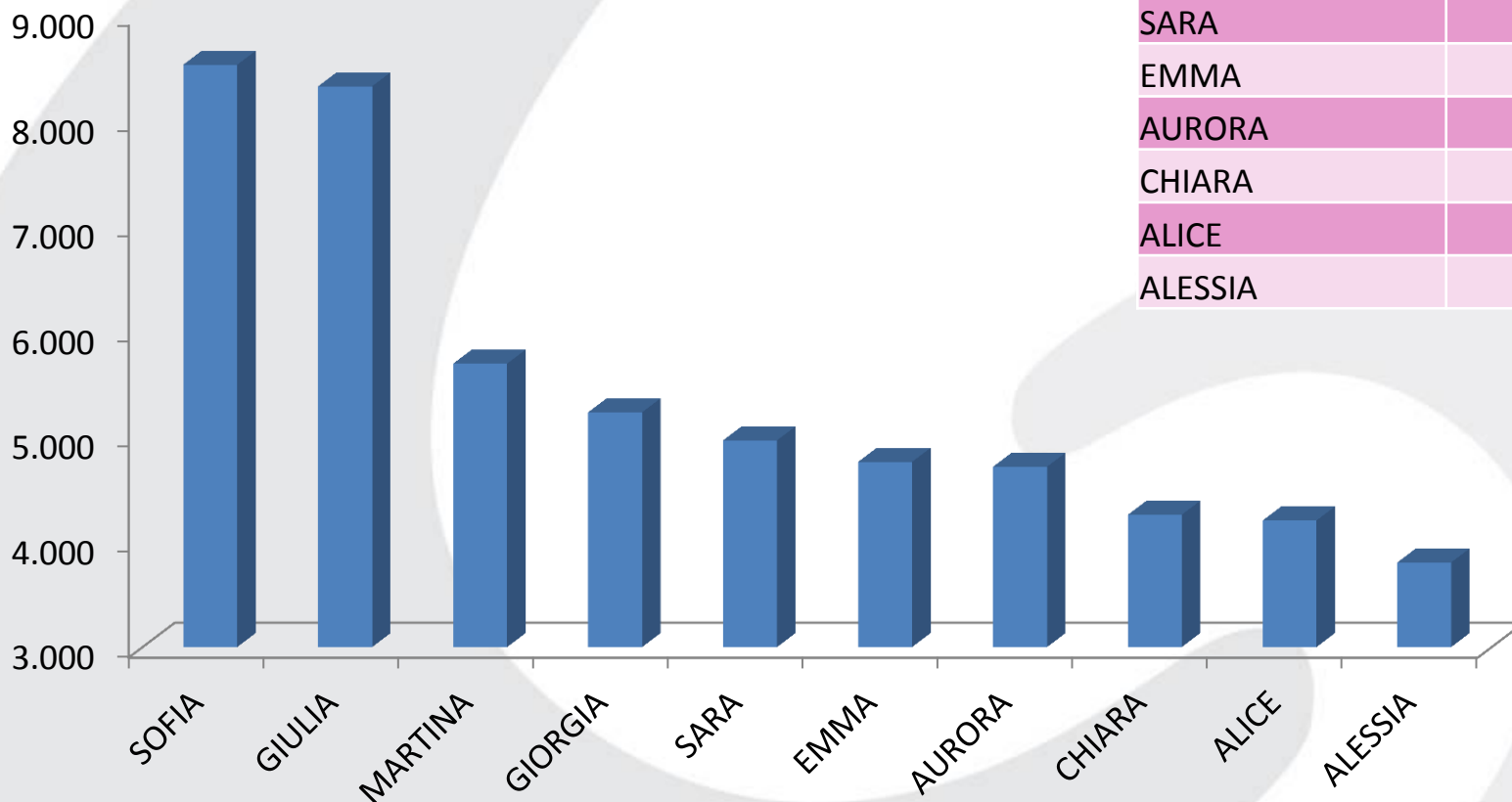
- Sintetizzare una variabile tramite la sua moda significa assumere come valore più rappresentativo della distribuzione quello che si è verificato più spesso di tutti gli altri
- Questo implica che se si sceglie tra più preferenze il valore modale si crea il minor numero di «scontenti»





# Esempio - mutabile

**Dieci nomi più utilizzati per le femmine nate nel 2011**



NOMI	NATE CON QUEL NOME
SOFIA	8.531
GIULIA	8.327
MARTINA	5.692
GIORGIA	5.230
SARA	4.964
EMMA	4.759
AURORA	4.713
CHIARA	4.257
ALICE	4.202
ALESSIA	3.804

Fonte: Istat - Rilevazione individuale degli iscritti in anagrafe per nascita. Anno 2011

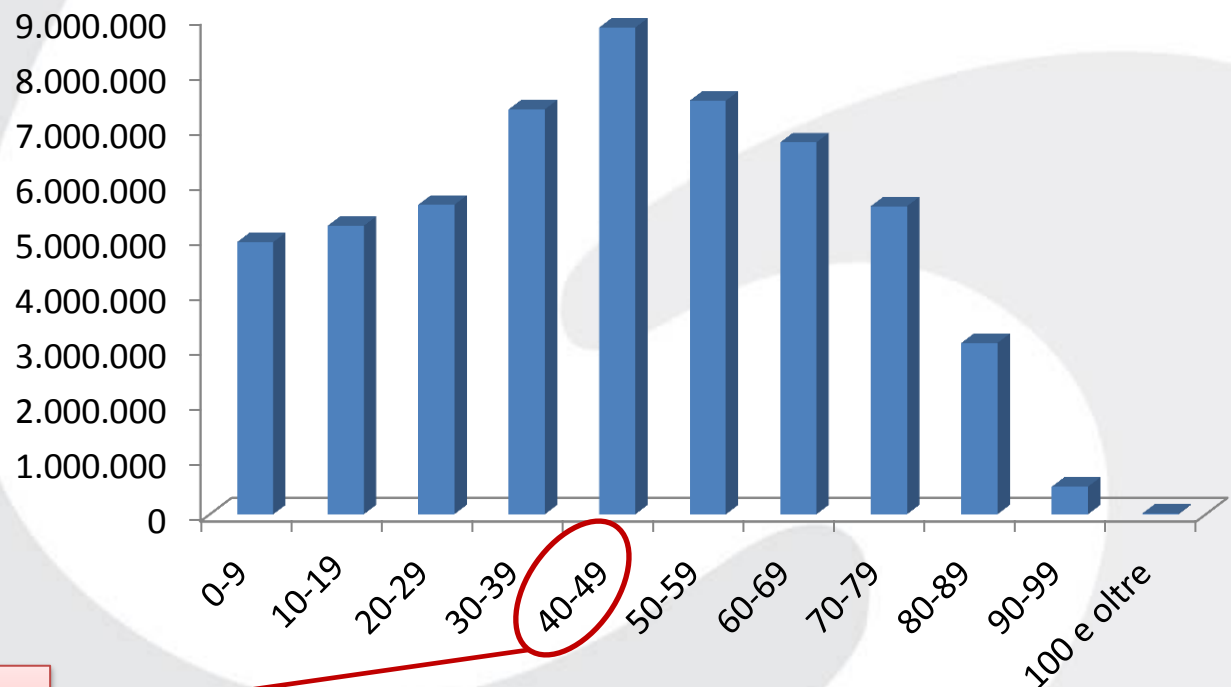


# Classe modale –Esempio

## Distribuzione con classi equi-ampie

Età	Popolazione
0-9	4.935.563
10-19	5.231.950
20-29	5.611.565
30-39	7.343.279
40-49	8.827.546
50-59	7.501.195
60-69	6.744.676
70-79	5.586.735
80-89	3.107.008
90-99	501.582
100 e oltre	15.018
<b>Totale</b>	<b>55.406.117</b>

**Popolazione residente in Italia di cittadinanza italiana per classe di età - Anno 2011**



**Classe  
modale**

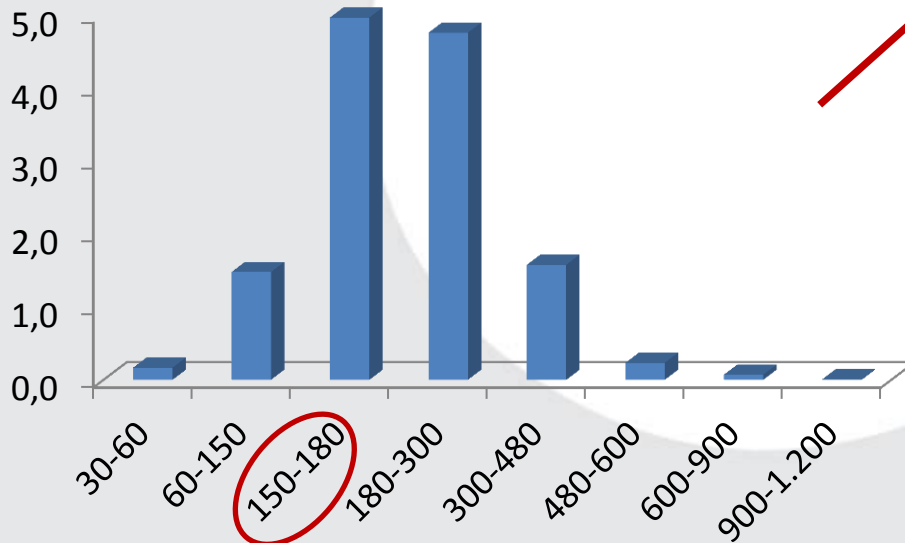
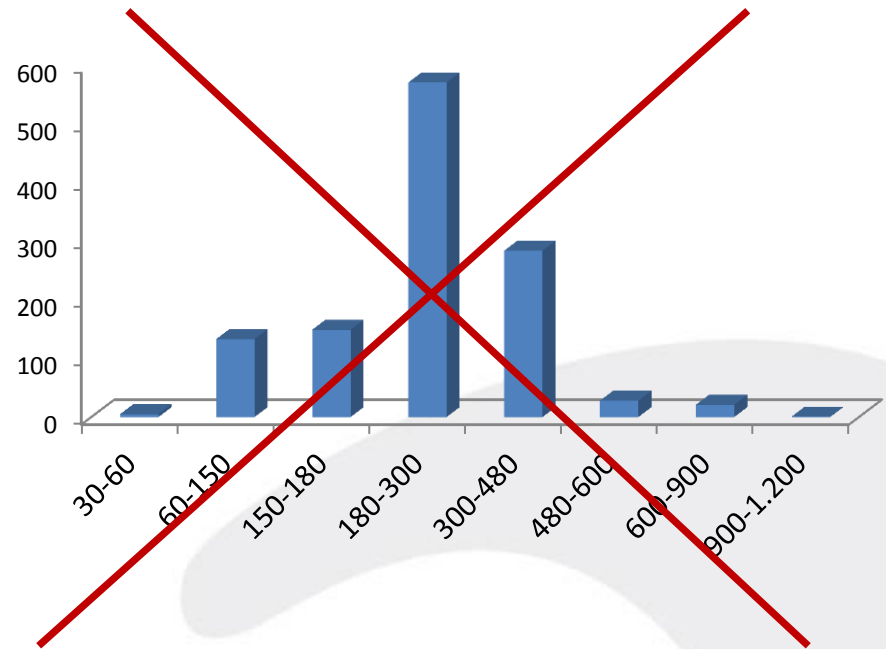
Fonte: 15° Censimento generale della Popolazione e delle abitazioni



# Classe modale – Esempio Distribuzione con classi di diversa ampiezza

Tab. 1- Durata in secondi di 1.192 brani musicali

Durata in secondi	Frequenze assolute	Frequenze relative	Densità di frequenza
30-60	5	0,00419	0,16667
60-150	133	0,11158	1,47778
150-180	149	0,12500	4,96667
180-300	571	0,47903	4,75833
300-480	284	0,23826	1,57778
480-600	28	0,2349	0,2333
600-900	21	0,01762	0,0700
900-1.200	1	0,00084	0,0082



# Moda

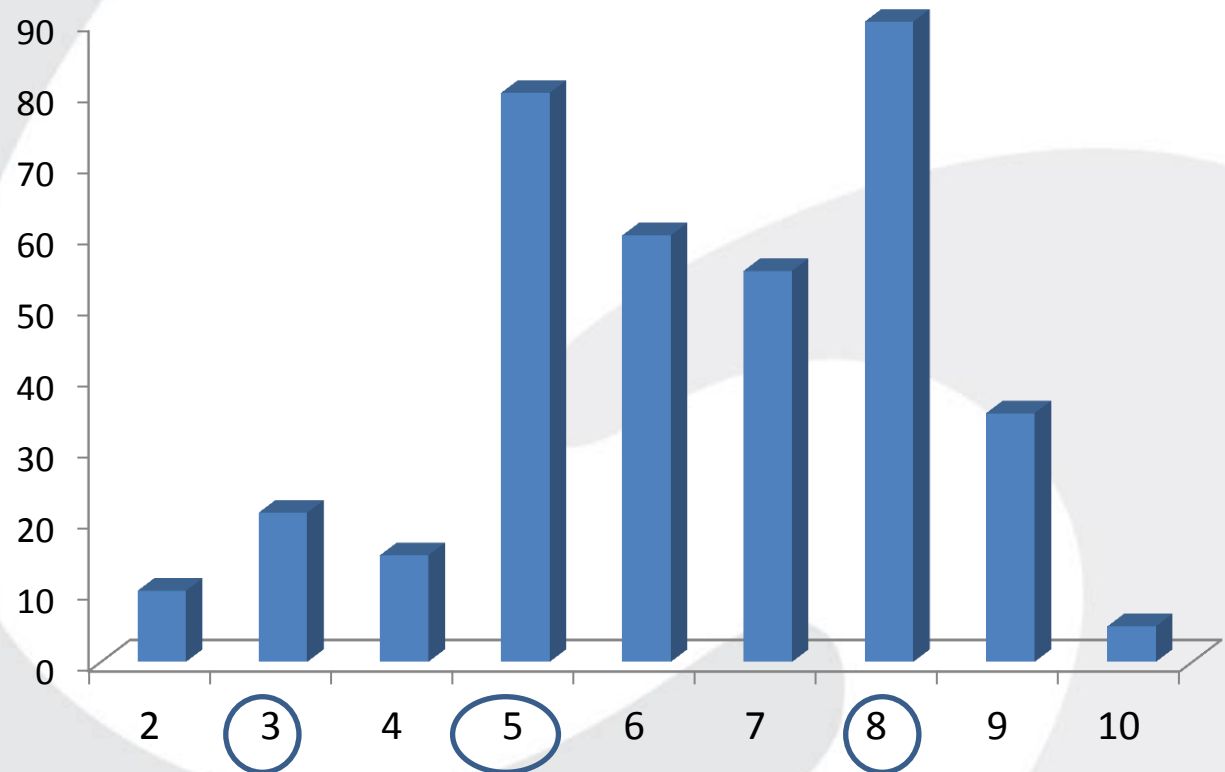
- Può verificarsi che più modalità abbiano una frequenza più elevata rispetto a quelle adiacenti. In tal caso la distribuzione presenta più valori modali e viene detta plurimodale
- Spesso quando si presentano più mode all'interno di una distribuzione di frequenza ciò deriva dalla presenza di sottopopolazioni che presentano posizioni differenti per ragioni strutturali



# Esempio

Voto in matematica degli alunni del corso B della scuola media X

Voto in matematica	Alunni
2	10
3	21
4	15
5	80
6	60
7	55
8	90
9	35
10	5
<b>Totale</b>	<b>371</b>



# Mediana

- Introdotta da Galton nel 1883 è definita come quel valore che divide una distribuzione statistica in due parti uguali.
- La mediana è la modalità dell'unità statistica che occupa il posto centrale nella distribuzione ordinata delle osservazioni, per cui metà della distribuzione ha valori inferiori alla mediana e l'altra metà ha valori superiori



# Calcolo della mediana

## Serie

Se  $x_1, x_2, x_3, \dots, x_n$ , sono i valori osservati, ciascuno con frequenza unitaria, disposti in ordine non decrescente, la mediana è data da:

se  $n$  è dispari  $Me = \frac{n+1}{2}$

–se  $n$  è pari si hanno due valori centrali

$$\begin{cases} Me_1 = \frac{n}{2} \\ Me_2 = \frac{n}{2} + 1 \end{cases}$$

$$Me = \frac{Me_1 + Me_2}{2}$$



# Calcolo della mediana - Esempio

- Retribuzione mensile di 5 dipendenti del comune X

–3.000 3.500 4.000 4.100 8.000

$$Me = (5+1)/2 = 3$$

–La mediana è rappresentata dalla retribuzione che occupa il 3° posto = 4.000

- Retribuzione mensile dei 6 dipendenti del comune X

–3.000 3.500 4.000 4.100 4.300 8.000

$$-Me_1 = 6/2 = 3 \Rightarrow 4.000$$

$$-Me_2 = (6/2)+1 = 4 \Rightarrow 4.100$$

$$Me = (4.000+4.100)/2 = 4.050$$





# Calcolo della mediana

## Distribuzione di frequenze

- Nel caso di distribuzioni di frequenze si opera sulle frequenze cumulate
- Se le modalità non sono raggruppate in classi la mediana è data dalla prima modalità la cui frequenza cumulata ( $N_i$ ) è immediatamente superiore a  $n / 2$



# Calcolo della mediana - Esempio

## Distribuzione di frequenza delle retribuzioni dei dipendenti del comune X

– Retribuzioni	Num	Freq. cumulate ( $N_i$ )
– 3.000	3	3
– 3.500	5	8
– 4.000	3	11
– 4.100	1	12
– 8.000	1	13 $\leftarrow n$

–Me  $\Rightarrow n/2 = 13/2 = 6,5 \Rightarrow 3.500$

Ampiezza  
collettiva



# Calcolo della mediana - Esempio

3.000 3.000 3.000 3.500 3.500 3.500 3.500 3.500 4.000  
4.000 4.000 4.100 4.100 8.000

$n=13$

$$Me = \frac{n+1}{2}$$

Me  $\Rightarrow (n+1)/2 = 14/2 = 7 \Rightarrow 3.500$



# Calcolo della mediana – frequenze assolute

## Distribuzione di frequenze

- Se le modalità sono raggruppate in classi bisogna prendere in considerazione le frequenze cumulate e ricercare la classe la cui frequenza cumulata è immediatamente superiore a  $n/2$ , questa è la classe mediana. Il valore mediano va ricercato per interpolazione:

$$Me = x_{i-1} + \frac{d_i}{n_i} \left( \frac{1}{2} - N_{i-1} \right)$$

dove:

- $x_{i-1}$  = estremo superiore della classe  $i-1$
- $N_{i-1}$  = frequenza cumulata della classe  $i-1$
- $n_i$  = frequenza della classe mediana
- $d_i$  = ampiezza della classe mediana



# Calcolo della mediana – frequenze relative

## Distribuzione di frequenze

$$Me = x_{i-1} + \frac{d_i}{f_i} \left( \frac{1}{2} - F_{i-1} \right)$$

dove:

- $x_{i-1}$  = estremo superiore della classe  $i-1$
- $F_{i-1}$  = frequenza relativa cumulata della classe  $i-1$
- $f_i$  = frequenza relativa della classe mediana
- $d_i$  = ampiezza della classe mediana



# Calcolo della mediana-Esempio

Età	Popolazione	Frequenze relative	Frequenze cumulate
0-9	4.935.563	0,0891	0,0891
10-19	5.231.950	0,0944	0,1835
20-29	5.611.565	0,1013	0,2848
30-39	7.343.279	0,1325	0,4173
40-49	8.827.546	0,1593	0,5766
50-59	7.501.195	0,1354	0,7120
60-69	6.744.676	0,1217	0,8338
70-79	5.586.735	0,1008	0,9346
80-89	3.107.008	0,0561	0,9907
90-99	501.582	0,0091	0,9997
100 e oltre	15.018	0,0003	1,0000
<b>Totale</b>	<b>55.406.117</b>	<b>1,0000</b>	

$$F_i \quad 0,5766 > 0,5$$

Classe mediana= 40-49

$$Me = x_{i-1} + \frac{d_i}{f_i} \left( \frac{1}{2} - F_{i-1} \right)$$

$$Me = 39 + \frac{10}{0,1593} (0,5 - 0,4173) = 44,1891$$

$M_e = 44$  anni e 68 giorni



# Proprietà della mediana

- È un indice robusto ossia, per il fatto che occupa una posizione centrale, non è influenzato dalla presenza di valori eccezionalmente grandi o eccezionalmente piccoli
- La somma dei valori assoluti degli scarti dalla mediana è un minimo



# Quantili

- Si definiscono quantili i  $q-1$  valori che ripartiscono una distribuzione statistica in  $q$  parti uguali
- Fra i quantili vengono utilizzati maggiormente i ***quartili*** e i ***centili***





# Quartili

- Sono i tre valori che dividono una distribuzione statistica ordinata in quattro parti di pari numerosità
- Pertanto:
  - 1° quartile=  $Q_1 = X_{(n/4)}$
  - 2° quartile=  $Q_2 = Me$
  - 3° quartile=  $Q_3 = X_{(3n/4)}$



# Quartili

Si noti che:

- il 25% dei valori della distribuzione è inferiore al primo quartile
- Il 50% è inferiore al secondo quartile
- Il 75% è inferiore al terzo quartile



# Calcolo dei quartili- Esempio (serie)

- Retribuzione mensile di 12 dipendenti del comune X

1.500 3.000 3.200 3.500 3.700 4.000 4.100 4.300 4.600 5.200 5.600 8.000

Primo quartile

$$X_{(n/4)} = X_{(12/4)} = 3 \Rightarrow 3.200$$

$$X_{(n/4)+1} = X_{(12/4)+1} = 4 \Rightarrow 3.500$$

$$Q_1 = (3.200 + 3.500) / 2 = 3.350$$

Secondo quartile=Me

$$X_{(n/2)} = X_{(12/2)} = 6 \Rightarrow 4.000$$

$$X_{(n/2)+1} = X_{(12/2)+1} = 7 \Rightarrow 4.100$$

$$Q_2 = (4.000 + 4.100) / 2 = 4.050$$



# Calcolo dei quartili- Esempio (serie)

Terzo quartile

$$X_{(3n/4)} = X_{(12/4)} = 3 \quad 4.600$$

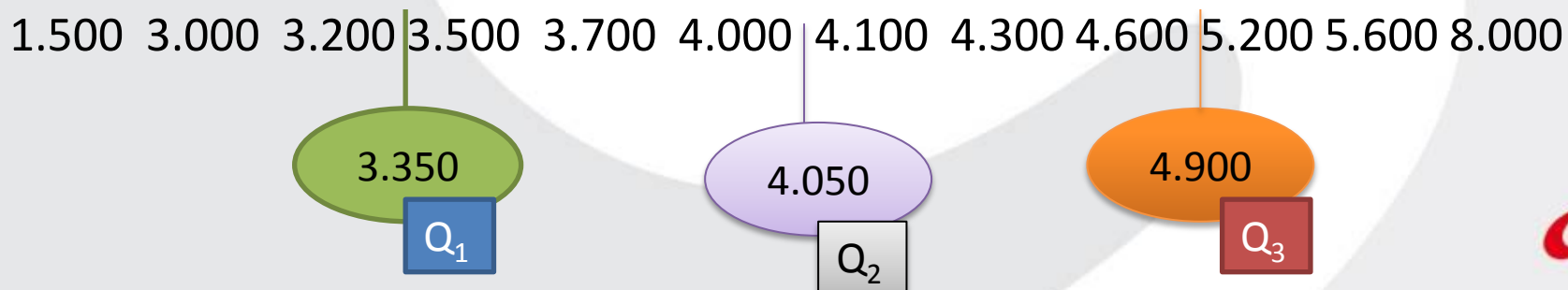
$$X_{(3n/4)+1} = X_{(36/4)+1} = 4 \quad \Rightarrow \quad 5.200$$

$$Q_3 = (4.600 + 5.200) / 2 = 4.900$$

$$Q_0 = \text{minimo} \quad 1.500$$

$$Q_4 = \text{massimo} \quad 8.000$$

- Retribuzione mensile di 12 dipendenti del comune X



# Calcolo dei quartili-Esempio (distribuzione)

Età	Popolazione	Frequenze relative	Frequenze cumulate
0-9	4.935.563	0,0891	0,0891
10-19	5.231.950	0,0944	0,1835
20-29	5.611.565	0,1013	0,2848
30-39	7.343.279	0,1325	0,4173
40-49	8.827.546	0,1593	0,5766
50-59	7.501.195	0,1354	0,7120
60-69	6.744.676	0,1217	0,8338
70-79	5.586.735	0,1008	0,9346
80-89	3.107.008	0,0561	0,9907
90-99	501.582	0,0091	0,9997
100 e oltre	15.018	0,0003	1,0000
<b>Italia</b>	<b>55.406.117</b>	<b>1,0000</b>	

$$3n/4=3/4=0,75$$

$$F_7= 0,0,8338>0,75$$

Classe = 60-69

$$Q_3 = x_{i-1} + \frac{d_i}{f_i} \left( \frac{3n}{4} - F_{i-1} \right)$$

$$n/4=1/4=0,25$$

$$F_3 0,2848>0,25$$

Classe = 20-29

$$Q_1 = x_{i-1} + \frac{d_i}{f_i} \left( \frac{n}{4} - F_{i-1} \right)$$

$$Q_1 = 19 + \frac{10}{0.1013} \left( \frac{1}{4} - 0.1835 \right) = 25,5647$$

$$Q_2 = Me = 44,1891$$

$$Q_3 = 59 + \frac{10}{0.1217} \left( \frac{3}{4} - 0.7120 \right) = 62,1224$$



# Calcolo dei quartili-Esempio

Età	Popolazione	Frequenze relative	Frequenze cumulate
0-9	4.935.563	0,0891	0,0891
10-19	5.231.950	0,0944	0,1835
20-29	5.611.565	0,1013	0,2848
30-39	7.343.279	0,1325	0,4173
40-49	8.827.546	0,1593	0,5766
50-59	7.501.195	0,1354	0,7120
60-69	6.744.676	0,1217	0,8338
70-79	5.586.735	0,1008	0,9346
80-89	3.107.008	0,0561	0,9907
90-99	501.582	0,0091	0,9997
100 e oltre	15.018	0,0003	1,0000
<b>Italia</b>	<b>55.406.117</b>	<b>1,0000</b>	

$$Q_1 = 19 + \frac{10}{0.1013} \left( \frac{1}{4} - 0.1835 \right) = 25,5647$$

$$Q_1 = 25,5647 = 25 \text{ anni e } 203 \text{ giorni}$$

$$Q_2 = Me = 44,1891$$

$$Q_2 = 44,1891 = 44 \text{ anni e } 68 \text{ giorni}$$

$$Q_3 = 59 + \frac{10}{0.1217} \left( \frac{3}{4} - 0.7120 \right) = 62,1224$$

$$Q_3 = 62,1224 = 62 \text{ anni e } 44 \text{ giorni}$$



# Media aritmetica

- L'indice di posizione per eccellenza è la media aritmetica delle osservazioni
- La media aritmetica (spesso indicata semplicemente come “media”) rappresenta l'indicatore sintetico più utilizzato nelle elaborazioni statistiche
- Costituisce un concetto primitivo per l'essere umano



# Media aritmetica

- La media aritmetica è data da:

$$\mu = \frac{\sum_{i=1}^n y_i}{n}$$

nel caso di serie

- E da:

$$\mu = \frac{\sum_{i=1}^u y_i n_i}{n}$$

nel caso di distribuzioni





# Calcolo della media -Esempio

Voti dell'alunno X nell'anno scolastico YY

– Italiano	8
– Latino	7
– Greco	8
– Storia	9
– Filosofia	10
– Matematica	8
– Scienze	6
– Fisica	6

$$\mu = \frac{\sum_{i=1}^n y_i}{n}$$

$$Media = \frac{8 + 7 + 8 + 9 + 10 + 8 + 6 + 6}{8} = 7,75$$



# Calcolo della media - Esempio

Età	Popolazione	Frequenze relative	valori centrali	$y_i n_i$
0-9	4.935.563	0,0891	4,5	22.210.034
10-19	5.231.950	0,0944	14,5	75.863.275
20-29	5.611.565	0,1013	24,5	137.483.343
30-39	7.343.279	0,1325	34,5	253.343.126
40-49	8.827.546	0,1593	44,5	392.825.797
50-59	7.501.195	0,1354	54,5	408.815.128
60-69	6.744.676	0,1217	64,5	435.031.602
70-79	5.586.735	0,1008	74,5	416.211.758
80-89	3.107.008	0,0561	84,5	262.542.176
90-99	501.582	0,0091	94,5	47.399.499
100 e oltre	15.018	0,0003	105,0	1.576.890
<b>Totale</b>	<b>55.406.117</b>	<b>1,0000</b>		<b>2.453.302.626</b>

$$\mu = \frac{\sum_{i=1}^u y_i n_i}{n}$$

$$\mu = \frac{2.453.302.626}{55.406.117} = 44,2785$$

$$\mu = 44 \text{ anni e } 100 \text{ giorni}$$



# Media

- La media aritmetica va interpretata come la quantità che sostituita ai singoli dati lascia invariato l'ammontare globale del carattere
- La media aritmetica è il baricentro della distribuzione nel senso che la somma degli scarti dei singoli dati dalla media è nulla
- Non è un indice robusto, essa viene influenzata dalla presenza di valori anomali



# Considerazioni finali

Non è possibile utilizzare un indice prescindendo dalle finalità della ricerca. Pertanto:

- La moda è utile quando bisogna minimizzare gli scontenti
- La mediana minimizza i costi complessivi
- La media è un indice di equilibrio generale



# Il teorema di Trilussa

- Se due persone hanno a disposizione due polli e uno li mangia tutti e due, in media ne hanno mangiato uno a testa

$$\text{media} = \frac{2 + 0}{2} = 1$$

## Il “dilemma” di TRILUSSA

“Me spiego: da li conti che se fanno  
seconno le statistiche d'adesso  
risurta che te tocca un pollo all'anno:  
e, se nun entra ne le spese tue,  
t'entra ne la statistica lo stesso  
perché c'è un antro che ne magna due”

$$\left[ \text{pollo} + \text{pollo} + 0 \right] / 2 = \text{pollo} (?)$$

- La metodologia statistica offre gli strumenti per individuare l'ineguale ripartizione dei ... polli



# La variabilità

In altri termini: se gli indici di posizione da un lato forniscono informazioni di sintesi su un collettivo, dall'altro da soli non riescono a dare informazioni complete sul fenomeno allo studio.

In particolare non fanno emergere l'aspetto legato alla  
**VARIABILITÀ**

cioè l'attitudine del fenomeno stesso ad assumere modalità differenti.



La metodologia statistica offre gli strumenti per individuare l'ineguale ripartizione del carattere tra le unità statistiche nel collettivo. Dunque, se un dato statistico è costruito in maniera corretta (tenendo conto della variabilità del fenomeno osservato) non si corre il rischio di fornire un'informazione sbagliata come citato ironicamente da Trilussa

# Indici di posizione e di variabilità

La capacità di un indice di posizione di sintetizzare in un solo valore una distribuzione, sarà tanto maggiore quanto più contenuta è la variabilità della distribuzione stessa

Una corretta informazione statistica dovrebbe sempre accompagnare ad un indice di posizione una misura della sua “accuratezza”, cioè un’indicazione della capacità dell’indice di riassumere le osservazioni



# Indici di posizione e di variabilità

- Per misurare la variabilità di un fenomeno si usano gli indici di variabilità
- Gli indici di variabilità più usati in statistica descrittiva sono:
  - il campo di variazione
  - la varianza
  - lo scarto quadratico medio
  - il coefficiente di variazione





# Campo di variazione

L'indice più semplice è il campo di variazione (range), rappresentato dalla differenza tra il valore massimo e il valore minimo delle modalità di  $Y$ , cioè:

$$\text{Range } (Y) = \max (Y) - \min (Y)$$



# Campo di variazione

Il campo di variazione è un indice piuttosto grossolano in quanto è influenzato dai valori estremi, che sono spesso valori eccezionali del fenomeno e poco indicativi dell'insieme dei dati

–Es.:

Reddito dei componenti la famiglia XXX

1300, 1600, 2000, 3500, 4000

Campo di variazione =  $4000 - 1300 = 2700$

Reddito dei componenti la famiglia XXX

1300, 1320, 1330, 1335, 4000

Campo di variazione =  $4000 - 1300 = 2700$



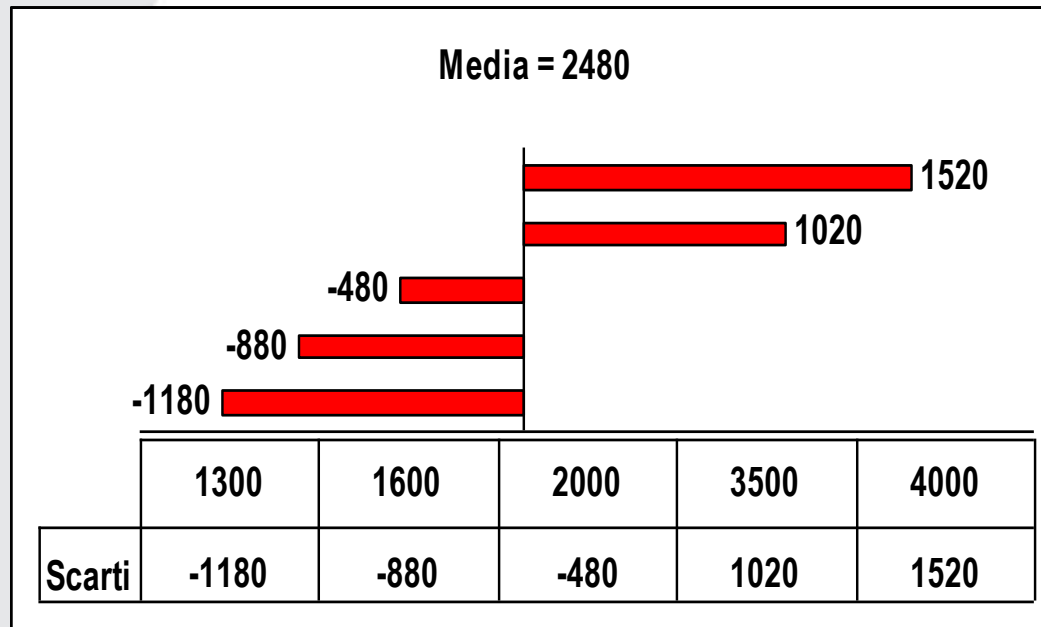
# Varianza

Per misurare la variabilità occorre misurare se le singole unità statistiche presentano modalità più o meno distanti dall'indice di posizione assunto come rappresentativo della distribuzione di frequenza



# Varianza

La varianza è un indice di variabilità che misura le distanze o scarti (al quadrato) dalla media aritmetica di ciascun valore della distribuzione



# Varianza

La varianza, introdotta da Fisher nel 1918, viene indicata con il simbolo  $\sigma^2$  e rappresenta l'indice più importante per misurare la variabilità (o dispersione) della serie o della distribuzione

La varianza è espressa dalla media degli scarti al quadrato ed è data da:

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \mu)^2}{n}$$

per le serie



# Varianza

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \mu)^2 n_i}{n}$$

per le distribuzioni

La varianza è una quantità sempre positiva e cresce all'aumentare della dispersione dei dati della distribuzione intorno alla media aritmetica



# L'ineguale ripartizione dei polli

Pers. 1	Pers. 2	$x_i - \mu$		$(x_i - \mu)^2$		$\Sigma (x_i - \mu)^2$	$\sigma^2$
2	0	1	-1	1	1	2	1
1,75	0,25	0,75	-0,75	0,56	0,56	1,13	0,56
1,50	0,50	0,50	-0,50	0,25	0,25	0,50	0,25
1	1	0	0	0	0	0	0

La varianza decresce man mano che la ripartizione dei polli diventa più equa



# Varianza

- La varianza può essere più facilmente calcolata con la seguente formulazione

$$\sigma^2 = \frac{\sum_{i=1}^n y_i^2}{n} - \left( \frac{\sum_{i=1}^n y_i}{n} \right)^2$$





# Calcolo della varianza - Esempio

Età	Popolazione	Frequenze relative	valori centrali	$y_i^2$	$y_i^2 f_i$
0-9	4.935.563	0,0891	4,5	20,25	1,80
10-19	5.231.950	0,0944	14,5	210,25	19,85
20-29	5.611.565	0,1013	24,5	600,25	60,79
30-39	7.343.279	0,1325	34,5	1.190,25	157,75
40-49	8.827.546	0,1593	44,5	1.980,25	315,50
50-59	7.501.195	0,1354	54,5	2.970,25	402,13
60-69	6.744.676	0,1217	64,5	4.160,25	506,43
70-79	5.586.735	0,1008	74,5	5.550,25	559,65
80-89	3.107.008	0,0561	84,5	7.140,25	400,40
90-99	501.582	0,0091	94,5	8.930,25	80,84
100 e oltre	15.018	0,0003	105,0	11.025,00	2,99
<b>Totale</b>	<b>55.406.117</b>	<b>1,0000</b>		<b>43.777,50</b>	<b>2.508,15</b>

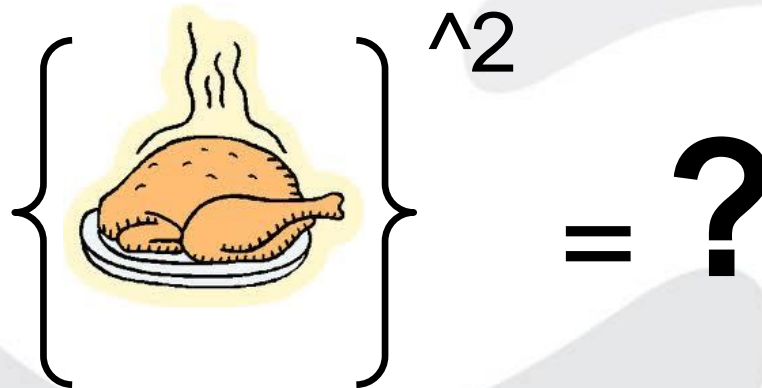
$$\sigma^2 = \frac{\sum_{i=1}^n y_i^2}{n} - \left( \frac{\sum_{i=1}^n y_i}{n} \right)^2 = \frac{\sum_{i=1}^n y_i^2 f_i}{1} - \left( \frac{\sum_{i=1}^n y_i f_i}{1} \right)^2$$

n=1

$$\sigma^2 = \frac{2.508,15}{1} - (44,29)^2 = 546,546$$

# Scarto quadratico medio

L'unità di misura della varianza è rappresentata dal quadrato dell'unità di misura del fenomeno in esame, risulta quindi di non immediata interpretazione (...pollo al quadrato?)



# Scarto quadratico medio

- Per questo motivo Pearson propose come misura della variabilità lo scarto quadratico medio (s.q.m.) o deviazione standard (d.s.) che, essendo dato dalla radice quadrata della varianza, è espresso nella stessa unità di misura dei dati.



# Scarto quadratico medio

Lo scarto quadratico medio è dato da

per le serie

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (y_i - \mu)^2}{n}}$$

per le distribuzioni

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (y_i - \mu)^2 n_i}{n}}$$



# Scarto quadratico medio

Età	Popolazione	Frequenze relative	valori centrali	$y_i^2$	$y_i^2 f_i$
0-9	4.935.563	0,0891	4,5	20,25	1,80
10-19	5.231.950	0,0944	14,5	210,25	19,85
20-29	5.611.565	0,1013	24,5	600,25	60,79
30-39	7.343.279	0,1325	34,5	1.190,25	157,75
40-49	8.827.546	0,1593	44,5	1.980,25	315,50
50-59	7.501.195	0,1354	54,5	2.970,25	402,13
60-69	6.744.676	0,1217	64,5	4.160,25	506,43
70-79	5.586.735	0,1008	74,5	5.550,25	559,65
80-89	3.107.008	0,0561	84,5	7.140,25	400,40
90-99	501.582	0,0091	94,5	8.930,25	80,84
100 e oltre	15.018	0,0003	105,0	11.025,00	2,99
<b>Totale</b>	<b>55.406.117</b>	<b>1,0000</b>		<b>43.777,50</b>	<b>2.508,15</b>

$$\mu = 44 \text{ anni e } 100 \text{ giorni}$$

$$\sigma = 23 \text{ anni e } 136 \text{ giorni}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^n (y_i - \mu)^2 n_i}{n}}$$

$$\sigma = \sqrt{\frac{2.508,15}{1} - (44,29)^2} = 23,38$$

# Coefficiente di variazione

- Lo scarto quadratico medio e la varianza sono indici assoluti, espressi rispettivamente in termini di unità di misura del fenomeno e del suo quadrato
- Per confrontare la variabilità di fenomeni espressi in unità di misura diverse, occorre un indice che non dipenda dalle unità di misura, sia cioè un numero puro



# Coefficiente di variazione

A tal fine si ricorre al coefficiente di variazione (CV), proposto da K. Pearson nel 1905 e dato da:

$$CV = \frac{\sigma}{\mu}$$

dove  $\mu > 0$



# Statistiche: avvertenze per l'uso

- Fare riferimento alle tabelle e non ai grafici
- Leggere sempre le note e l'intestazione
- Individuare la fonte
- Leggere i metadati
- Spesso un solo indicatore può non essere sufficiente per descrivere un fenomeno complesso

Tab. 1- Popolazione residente per ripartizione geografica-Anno2011

Ripartizioni geografiche	Popolazione residente
Nord	27.283.974
Centro	11.657.380
Mezzogiorno	20.629.227
<b>Italia</b>	<b>59.570.581</b>

Fonte: 15° Censimento generale della Popolazione e delle abitazioni

## Il "dilemma" di TRILUSSA

"Me spiego: da li conti che se fanno seconno le statistiche d'adesso risurta che te tocca un pollo all'anno: e, se nun entra ne le spese tue, t'entra ne la statistica lo stesso perché c'è un antro che ne magna due"

$$\left[ \text{pollo} + 0 \right] / 2 = \text{pollo} (?)$$





*Tu mangi un pollo. Io no.  
La media è mezzo pollo a testa,  
ma io ho ancora fame...  
anche questa è statistica...(simon)\**

\*da «La statistica è...» concorso lanciato dall'Istat in occasione della prima Giornata mondiale della statistica

