

# Tecniche statistiche di analisi del cambiamento

01-Ripasso statistica descrittiva  
(vers. 1.1a, 7 ottobre 2016)

Germano Rossi<sup>1</sup>

`germano.rossi@unimib.it`

<sup>1</sup>Dipartimento di Psicologia, Università di Milano-Bicocca

2016-17

# Link utili

## Ripassare i livelli di misura

<http://www.germanorossi.it/mi/file/elem/Elem01-Introduzione.pdf>

## Ripassare le rappresentazioni grafiche

<http://www.germanorossi.it/mi/file/elem/Elem02-GraficiDistribuzioni.pdf>

## Ripassare i percentili

<http://www.germanorossi.it/mi/file/elem/Elem03-Percentili.pdf>

## Ripassare le statistiche della tendenza centrale

<http://www.germanorossi.it/mi/file/elem/Elem04-TendCentrale.pdf>

## Ripassare le statistiche di variabilità

<http://www.germanorossi.it/mi/file/elem/Elem05-IndiciVariabilita.pdf>

# Livelli di misura

Sapete il significato di:

- livelli di misura o scale di misura
- variabile categoriale vs. punteggi
- variabile qualitativa vs. quantitativa
- variabile discreta vs. continua
- variabile nominale vs. ordinale vs. intervallo/rapporto

*(Alzate la mano se NON sapete una di queste cose)*

# Livelli di misura

Sapete il significato di:

- livelli di misura o scale di misura
- variabile categoriale vs. punteggi
- variabile qualitativa vs. quantitativa
- variabile discreta vs. continua
- variabile nominale vs. ordinale vs. intervallo/rapporto

*(Alzate la mano se NON sapete una di queste cose)*

Se conoscete questi concetti, saltiamo al lucido **9**

Ripassare i livelli di misura

<http://www.germanorossi.it/mi/file/elem/Elem01-Introduzione.pdf>

# Livelli di misura

- I diversi modi di misurare si chiamano **livelli di misura**
- I livelli di misura possono essere classificati con diverse modalità

categorie vs. punteggi

qualitative vs. quantitative

discrete vs. continue

nominali vs. ordinali vs. intervallo/rapporto

## Attenzione

- Excel e SPSS possono lavorare con numeri e con stringhe (testo)
- Excel e SPSS non sanno nulla delle scale di misura
- Se sono numeri, li trattano come numeri; se sono testi, come testi

# Qualitative vs. quantitative

- Un livello di misura è **qualitativo** quando tiene conto solo e soltanto di una qualche **qualità** di ciò che si sta misurando (ad. es. il colore dei capelli, la razza di appartenenza, il titolo di studio raggiunto)
- Le scale qualitative vengono anche chiamate **categoriali** perchè i loro valori sono categorie
- Un livello di misura è **quantitativo** quando è possibile usare i numeri per indicare i diversi livelli di ciò che si sta misurando (ad. l'età, l'altezza di una persona, il numero di risposte giuste ad un questionario) e il numero esprime “effettivamente” un'**unità di misura**
- Le scale qualitative utilizzano **punteggi**

# Discrete vs. continue

- Un livello di misura è **continuo** quando un valore utilizzato “scivola” nell’altro, cioè quando vi sono infinite possibili misurazioni (ad es. l’età). Il valore che viene misurato è sempre un’*approssimazione del valore reale* (l’approssimazione dipende dalla sensibilità dello strumento di misura)



- Un livello di misura è **discreto** quando un valore utilizzato è completamente separato da qualunque altro e non vi sono valori intermedi; le variabili qualitative sono sempre discrete, mentre quelle quantitative sono discrete se non è possibile effettuare una misurazione intermedia fra due valori contigui (ad es. numero di persone che entrano in un negozio)



# Nominali vs. ordinali vs. intervallo/rapporto

- È il sistema di classificazione più usato ed è stato proposto da Stevens
- Un livello di misura è **nominale** quando è soltanto possibile dare un nome (o etichetta) alle categorie di ciò che si sta misurando (ad. es. il colore dei capelli)
- Un livello di misura è **ordinale** quando è possibile ordinare i diversi livelli di categorie (ad. per il titolo di studio, la maturità è superiore alla licenza media, la laurea è superiore alla maturità, ...) in un modo qualunque
- Un livello di misura è ad **intervallo/rapporto** quando i valori assunti dalla variabile possono essere espressi tramite numeri che fanno riferimento ad una specifica **unità di misura** (anni di studio, altezza in cm, ...)

# Implicazioni

Per stabilire il livello di misurazione di una variabile devo considerare:

- la natura della variabile stessa
- il modo in cui è stata rilevata
- quali sono le relazioni valide all'interno della variabile
- **Attenzione:** il nome della variabile potrebbe non indicare la sua scala di misura (se non è un nome codificato)

## Esempio

Il **reddito** è una variabile quantitativa che, in teoria, oscilla fra 0 e  $+\infty$ ; se rilevo il valore esatto del reddito, sarà misurata a livello *rapporto* (unità di misura, origine assoluta); se rilevo delle fasce di reddito, sarà misurata a livello *ordinale* (non esiste un'unità di misura)

Chi **non conosce SPSS**, alzi la mano!

# SPSS

Chi **non conosce SPSS**, alzi la mano!

Se lo conoscete, saltiamo al lucido **14**

# Cos'è SPSS

**SPSS** significa **Statistical Package for Social Sciences** (Pacchetto statistico per le scienze sociali) ed è un programma per gestire dati e calcolare statistiche. Ma adesso si chiama **IBM Spss**

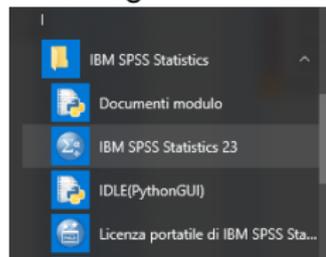
- Permette di inserire, nominare e gestire le misurazioni statistiche
- Permette di manipolare le variabili (in modo analogo ai fogli elettronici)
- Permette di calcolare le statistiche descrittive, di fare grafici
- Permette di fare l'analisi dei dati

Esiste un clone “open source” chiamato PSPP

(<http://savannah.gnu.org/projects/pspp>, gratuito ma non ancora completo)

# Primo avvio

## ■ Per eseguire SPSS

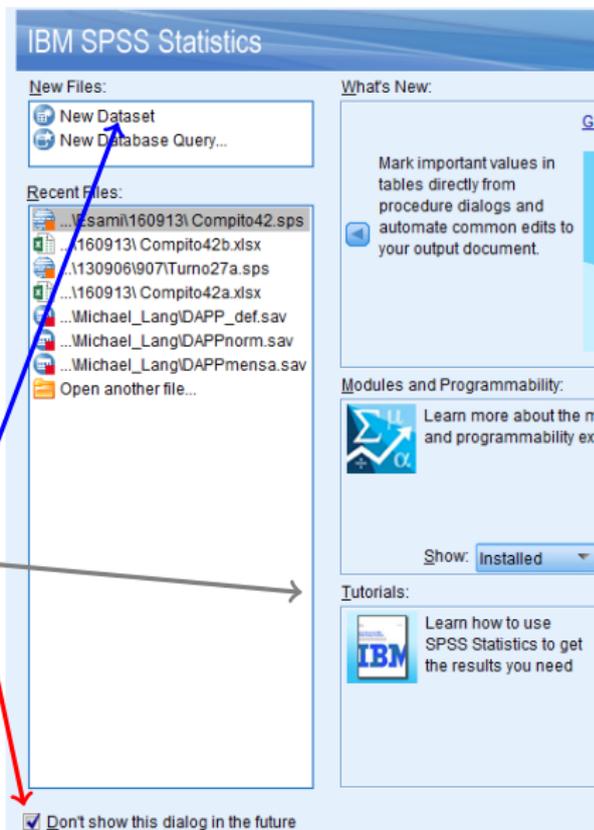


■ Alla prima esecuzione compare una finestra di dialogo che permette di effettuare alcune scelte. Per disattivarla, click-ate qui

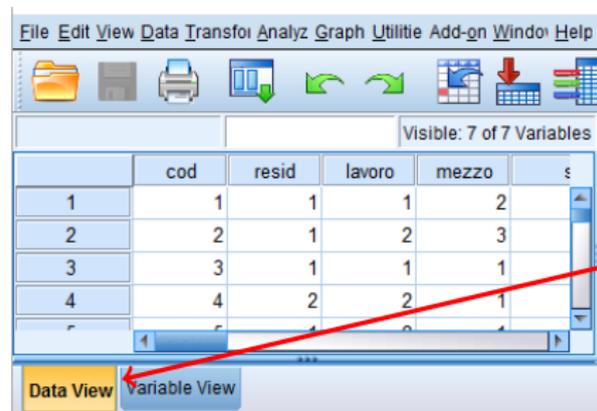
■ Potete selezionare “Tutorials”

■ Potete selezionare “New Dataset”

■ poi 



# Finestra di base



File Edit View Data Transform Analyze Graph Utilities Add-on Windows Help

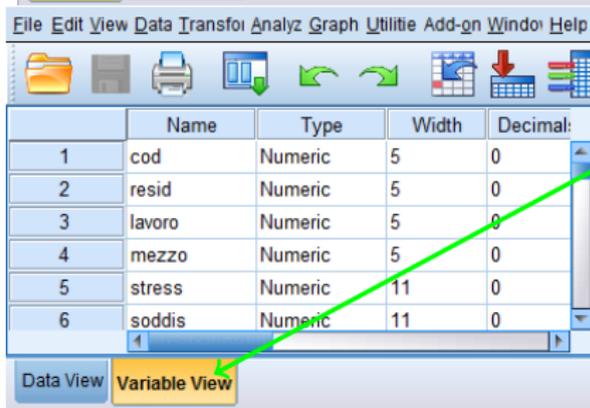
Visible: 7 of 7 Variables

	cod	resid	lavoro	mezzo	s
1	1	1	1	2	
2	2	1	2	3	
3	3	1	1	1	
4	4	2	2	1	
5	5	1	1	1	

Data View Variable View

A red arrow points from the 'Variable View' tab to the 'Variable View' window below.

- La finestra principale è formata da due pannelli
- uno per la visualizzazione dei dati
  - Ogni colonna è una variabile
  - Ogni riga un caso statistico
- uno per la descrizione delle variabili



File Edit View Data Transform Analyze Graph Utilities Add-on Windows Help

	Name	Type	Width	Decimal
1	cod	Numeric	5	0
2	resid	Numeric	5	0
3	lavoro	Numeric	5	0
4	mezzo	Numeric	5	0
5	stress	Numeric	11	0
6	soddis	Numeric	11	0

Data View Variable View

A green arrow points from the 'Variable View' tab to the 'Variable View' window above.

- dopo il nome, informazioni sul tipo (numerico, stringa, data...), l'ampiezza, i decimali (se numero)...

# Finestra di base



Le voci di menù più utilizzate sono:

- **Dati**: permette di agire sul file dei dati (ordinare, selezionare, filtrare...)
- **Trasforma**: permette di manipolare le variabili (calcolare nuove variabili, ricodificare, contare valori...)
- **Analizza**: È il menù più utilizzato perché contiene tutte le procedure statistiche disponibili

La prima volta che usate SPSS vi conviene fare l'**esercitazione** disponibile nell'Help (Guida).

In tutti i lucidi, per indicare un percorso di menù, userò questa convenzione voce principale | sottomenù | sotto-sottomenù:  
ad es. Guida | Esercitazione

# Cos'è la statistica

La **statistica** come scienza si può suddividere in

- Statistica descrittiva (il primo passo) vs. statistica inferenziale (il passo successivo)
- **Statistica descrittiva**
  - Riassume e descrive tramite numeri le caratteristiche principali di un insieme di “misurazioni” (che rappresentano “informazioni”) raccolte sulla realtà
- **Statistica inferenziale**
  - Permette di stimare la “sicurezza” di alcune statistiche calcolate su un piccolo insieme (campione) rispetto alla popolazione oppure di confrontare tra loro dei gruppi (insiemi) differenti

# Statistica descrittiva

Avendo un insieme di dati, possiamo fare una distribuzione di frequenze, delle rappresentazioni grafiche, calcolare delle statistiche descrittive

I tipi di statistiche descrittive possibili sono:

- **Rappresentazioni** riassuntive
- Statistiche della **tendenza centrale**: danno indicazioni sulla prevalenza dei dati
- Statistiche di **posizione**: associano certi valori con la posizione che occupano nella distribuzione e viceversa
- Statistiche di **variabilità**: come i dati sono dispersi
- Statistiche di **associazione**: ci informano su quanto due (o più) variabili sono collegate fra loro [prossima lezione]

# Rappresentazioni riassuntive

Conoscete queste **Rappresentazioni** riassuntive?

- Distribuzioni di frequenza (semplice, cumulata)
- Rappresentazioni grafiche
  - Grafici a barre
  - Istogrammi a barre

*(Alzate la mano se NON sapete una di queste cose)*

# Rappresentazioni riassuntive

Conoscete queste **Rappresentazioni** riassuntive?

- Distribuzioni di frequenza (semplice, cumulata)
- Rappresentazioni grafiche
  - Grafici a barre
  - Istogrammi a barre

*(Alzate la mano se NON sapete una di queste cose)*

Se le conoscete, saltiamo al lucido **30**

Ripassare le rappresentazioni grafiche

<http://www.germanorossi.it/mi/file/elem/Elem02-GraficiDistribuzioni.pdf>

# Distribuzione di frequenza semplice

Si tratta di contare quanti elementi appartengono ad una stessa categoria presente in una variabile.

## Esempio

Se abbiamo la seguente distribuzione di dati, la distribuzione di frequenza sarà:  
M F M F F M M M F F M M F F M  
M F F M M M M F F M F M F M M

	f		
F	13	←	$f_f$
M	17	←	$f_m$
Tot	30	←	$N$

# Distribuzione di frequenza semplice

Si tratta di contare quanti elementi appartengono ad una stessa categoria presente in una variabile.

## Esempio

Se abbiamo la seguente distribuzione di dati, la distribuzione di frequenza sarà:

M **F** M **F F** M M M **F F** M M **F F** M  
M **F F** M M M M **F F** M **F** M **F** M M

	<b>f</b>		
<b>F</b>	<b>13</b>	←	$f_f$
M	17	←	$f_m$
Tot	30	←	$N$

- Contiamo le F; il loro numero è la frequenza delle femmine ( $f_f$ )

# Distribuzione di frequenza semplice

Si tratta di contare quanti elementi appartengono ad una stessa categoria presente in una variabile.

## Esempio

Se abbiamo la seguente distribuzione di dati, la distribuzione di frequenza sarà:

M F M F F M M M F F M M F F M  
M F F M M M M F F M F M F M M

		f		
F	13	←	$f_f$	
M	17	←	$f_m$	
Tot	30	←	$N$	

- Contiamo le F; il loro numero è la frequenza delle femmine ( $f_f$ )
- Contiamo le M; il loro numero è la frequenza dei maschi ( $f_m$ )

# Distribuzione di frequenza semplice

Si tratta di contare quanti elementi appartengono ad una stessa categoria presente in una variabile.

## Esempio

Se abbiamo la seguente distribuzione di dati, la distribuzione di frequenza sarà:  
M F M F F M M M F F M M F F M  
M F F M M M M F F M F M F M M

	f		
F	13	←	$f_f$
M	17	←	$f_m$
Tot	30	←	$N$

- Contiamo le F; il loro numero è la frequenza delle femmine ( $f_f$ )
- Contiamo le M; il loro numero è la frequenza dei maschi ( $f_m$ )
- La somma di tutte le frequenze equivale al numero dei casi statistici (indicato con N)

# Distribuzione di frequenza semplice

- La somma di tutte le frequenze, deve necessariamente equivalere a N (chiamata **numerosità** ovvero il numero di casi statistici).
- Se mancano delle risposte, dovrà esserci una categoria “mancanti” o simili
- la distribuzione di frequenza permette di riassumere velocemente molti dati
- si applica sicuramente a scale Nominali, Ordinali ed è possibile anche a livello Intervallo/Rapporto (ma non sempre è utile)

# Distribuzione di frequenza cumulata

- È la somma delle frequenze che precedono una determinata categoria
- si applica a scale Ordinali e Intervallo/Rapporto

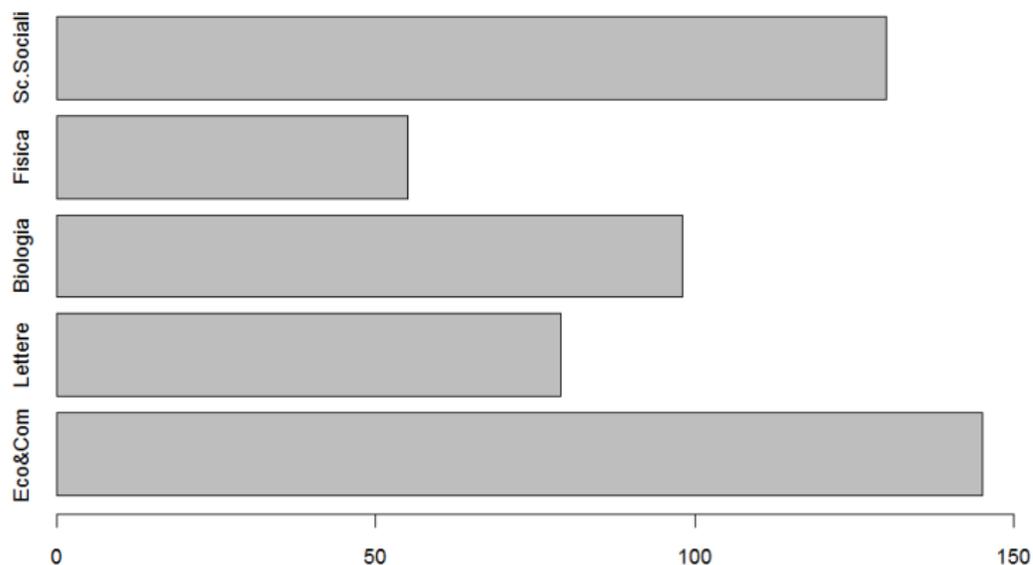
	f	fc		
Nessuna	4	4	4	
1 scelta	2	6	4+2	
2-4 scelte	52	58	6+52	4+2+52
5 o più scelte	18	76	58+18	4+2+52+18
Totale	76			

# Grafici/Istogrammi a barre

- Grafici e istogrammi a barre sono molti simili
- I **grafici a barre** sono indicati per variabili nominali e ordinali: le barre sono separate l'una dall'altra
- Gli **istogrammi a barre** sono indicati per variabili intervallo/rapporto: le barre sono contigue fra loro
- Ogni frequenza è rappresentata da una barra
- La lunghezza di ogni barra è proporzionale alla frequenza: barre più lunghe indicano frequenze più elevate
- Negli istogrammi anche l'area di una barra è proporzionale alla frequenza

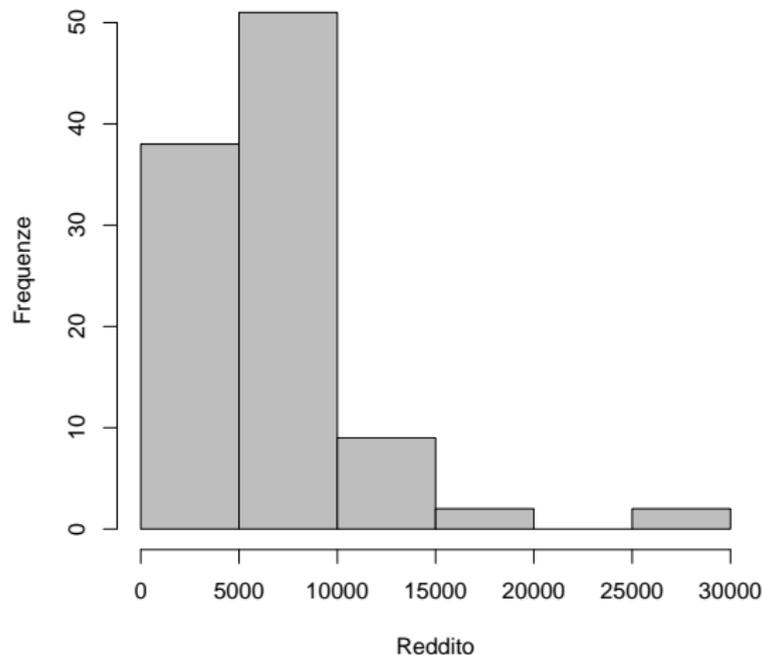
# Grafico a barre (Nominale/Ordinale)

Studenti iscritti in varie Facoltà



# Istogramma (I/R)

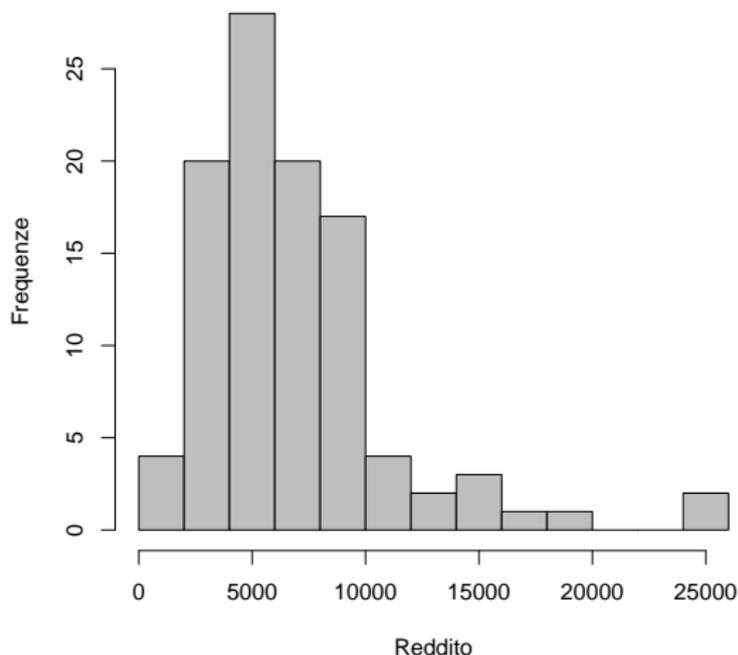
Istogramma



- la variabili quantitative (in particolare quelle continue) vengono rappresentate tramite intervalli di classe (SPSS lo fa automaticamente)

# Istogramma con classi più piccole (I/R)

Istogramma con classi più piccole



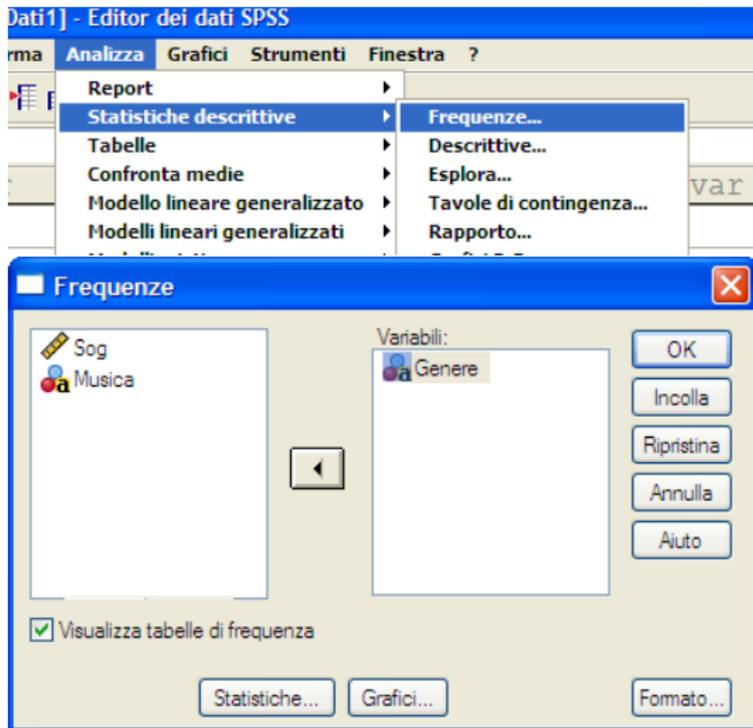
- se la variabile non è molto simmetrica, l'ampiezza degli intervalli può cambiare moltissimo la distribuzione raffigurata

# Distribuzione e grafici in SPSS

- La maggior parte delle statistiche che abbiamo visto e che vedremo e i relativi grafici, si ottengono in SPSS tramite due comandi
  - `Analizza | Statistiche descrittive | Frequenze...`
  - `Analizza | Statistiche descrittive | Esplora...`
- In Spss si può fare solo quello che è previsto dal software; i grafici non sono bellissimi
- In Excel si possono fare molti più grafici (anche più “belli”)
- In R (usato in questi lucidi) si può fare praticamente tutto, ma è molto più complicato

# Spss: frequenze

- Calcoliamo le frequenze con il comando  
Analizza | Statistiche descrittive | Frequenze...
- Poi spostiamo una variabile nominale/ordinale fra le Variabili
- E premiamo OK



# Spss: frequenze

**Frequenze**

[InsiemeDati1] C:\Documenti\TeX\lucidi\Elem\Fig\esempio\_dati\_x\_lucidi.sav

**Statistiche**

Genere

N	Validi	30
	Mancanti	0

**Genere**

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	F	13	43,3	43,3	43,3
	M	17	56,7	56,7	100,0
<b>Total</b>	<b>e</b>	<b>30</b>	<b>100,0</b>	<b>100,0</b>	

# Spss: istogramma (Frequenze)

- Analizza | Statistiche descrittive | Frequenze...

■ Pulsante

■ Selezioniamo Istogramma

■ eventualmente anche Con curva normale

■ Pulsante

■ e



# Spss: istogramma (Esplora)

- Analizza | Statistiche descrittive | Esplora...

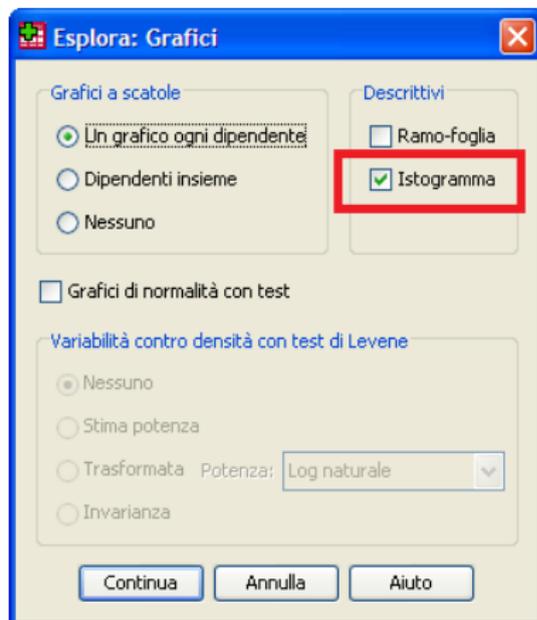
- Pulsante Grafici...

- Selezioniamo Istogramma

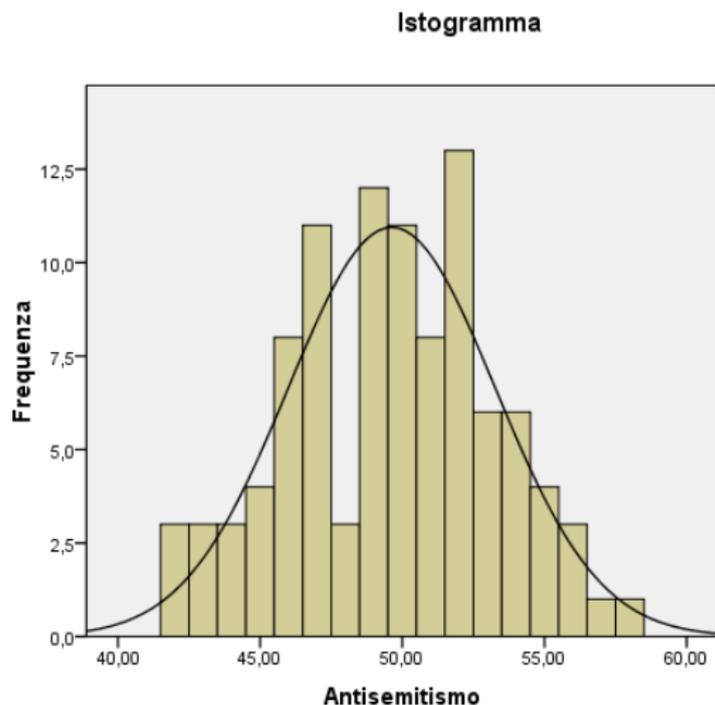
- (non è possibile avere la curva normale)

- Pulsante Continua

- e OK



# Spss: istogramma (Frequenze/Esplora)



Media =49,63  
Dev. stand. =3,645  
N =100

- Spss usa automaticamente gli intervalli di classe
- non è possibile modificare l'ampiezza di classe
- La curva normale non c'è con Esplora

# Indici di posizione

Conoscete questi **Indici di posizione**?

- Percentuali
- N-tili
- Quartili
- Percentili
- IQR (o differenza interquartilica)

*(Alzate la mano se NON conoscete una di queste)*

# Indici di posizione

Conoscete questi **Indici di posizione**?

- Percentuali
- N-tili
- Quartili
- Percentili
- IQR (o differenza interquartilica)

*(Alzate la mano se NON conoscete una di queste)*

Se sapete tutto, saltiamo al lucido **43**

Ripassare gli Indici di posizione

`http://www.germanorossi.it/mi/file/elem/Elem03-Percentili.pdf`

# Proporzioni e percentuali

Proporzione

$$\frac{f}{N}$$

Percentuale

$$\frac{f}{N} \times 100$$

- Le proporzioni e le percentuali si possono calcolare a tutti i livelli (N, O, I/R), anche se è preferibile non usarli per variabili intervallo/rapporto continue e con molti valori
- Quando la N su cui si calcolano le percentuali è molto piccola ( $N \leq 20$ ), la percentuale perde di significato
- Infatti 1 unità su 20 equivale al 5%, mentre 1 unità su 100 equivale all'1%
- Se per un qualunque motivo, si è verificato un errore di misurazione e una unità è stata “categorizzata” male, se N è basso, le percentuali cambieranno parecchio dando una visione falsata della realtà

# Tabella delle frequenze

Una **tabella delle frequenze** indica quanti valori abbiamo per ogni categoria della variabile.

Sesso	freq	prop	%
Maschi	17	.57	56.67
Femmine	13	.43	43.33
Totale	30	1.00	100

- Di solito si indicano **frequenze e percentuali** oppure **frequenze e proporzioni**
- A livello ordinale, si possono calcolare anche le *percentuali cumulate*

# Frequenze cumulate

D12

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	1	8	6,8	6,9	6,9
	2	22	18,6	19,0	25,9
	3	18	15,3	15,5	41,4
	4	14	11,9	12,1	53,4
	5	15	12,7	12,9	66,4
	6	13	11,0	11,2	77,6
	7	5	4,2	4,3	81,9
	8	8	6,8	6,9	88,8
	9	2	1,7	1,7	90,5
	10	11	9,3	9,5	100,0
	Totale	116	98,3	100,0	
Mancanti	Mancante di sistema	2	1,7		
	Totale	118	100,0		

Il 53.4% dei soggetti ha dato come risposta valori inferiori a 5.

# Frequenze cumulate

**Autoritarismo**

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	20,00	1	,8	,9	,9
	24,00	1	,8	,9	1,8
	26,00	1	,8	,9	2,7
	27,00	2	1,7	1,8	4,5
	31,00	3	2,5	2,7	7,1
	32,00	2	1,7	1,8	8,9
	33,00	2	1,7	1,8	10,7
	34,00	1	,8	,9	11,6
	35,00	1	,8	,9	12,5
	36,00	2	1,7	1,8	14,3
	37,00	4	3,4	3,6	17,9
	38,00	3	2,5	2,7	20,5

Su variabili intervallo/rapporto (spesso) non hanno molto senso le tabelle di frequenza (soprattutto se ci sono molti valori diversi con frequenze simili).

# Indici di posizione: N-tili

- Dopo aver ordinato i valori di una distribuzione, possiamo suddividere l'intera distribuzione di frequenza in  $n$  parti uguali.
- Se divisa in 100 parti, **Centili** ( $C_1, C_2, \dots, C_{99}$ ) o **Percentili** ( $P_1, P_2, \dots$ )
- se divisa in 10 parti, **Decili** ( $D_1, D_2, \dots, D_9$ )

■ se divisa in 4 parti, **Quartili** ( $Q_1, Q_2, Q_3$ )

■ se divisa in 3 parti, **Terzili**

■ Notate che

- $D_1 = P_{10}$ , e così via
- $Q_1 = P_{25}$ ,
- $Q_2 = P_{50} = D_5$  è anche chiamato "Mediana",
- $Q_3 = P_{75}$

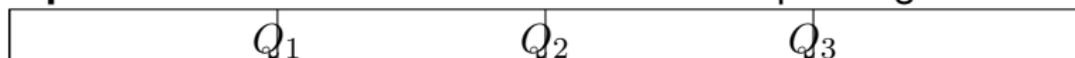
la formula generica è:

$$\frac{N}{\text{parti}} * \text{ntile}$$

ma cambia un pochino  
in base al tipo di n-tile

# Quartili, Percentili

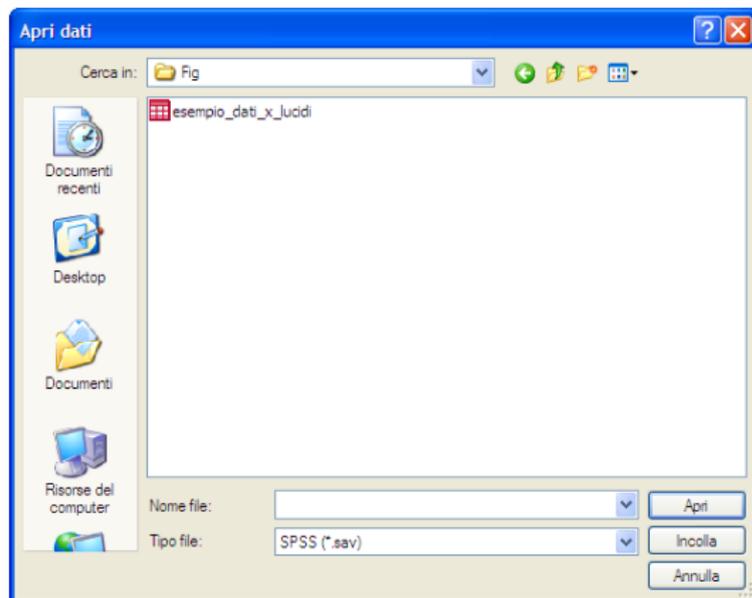
- Un quant-tile è **un valore corrispondente ad una posizione**
- I **quartili** suddividono la distribuzione in 4 parti uguali



- Quindi ogni quartile delimita uno dei 4 blocchi corrispondenti ad un 25% dei dati
- Si usano solitamente il primo e il terzo quartile ( $Q_1$  e  $Q_3$ )
- Hanno senso con variabili almeno a livello Ordinale (ma se non ci sono poche categorie)
- I **percentili** corrispondono (grosso modo) alle percentuali

Usando il file `esempio1.sav` vediamo le statistiche presentate.

- Usando **File** | **Apri** | **Dati** cerchiamo e apriamo il file

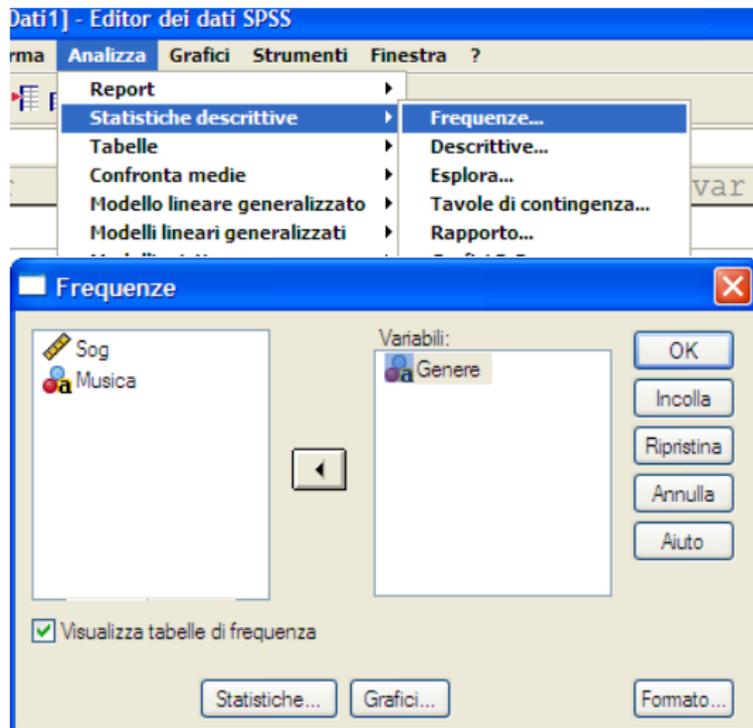


# Spss: frequenze

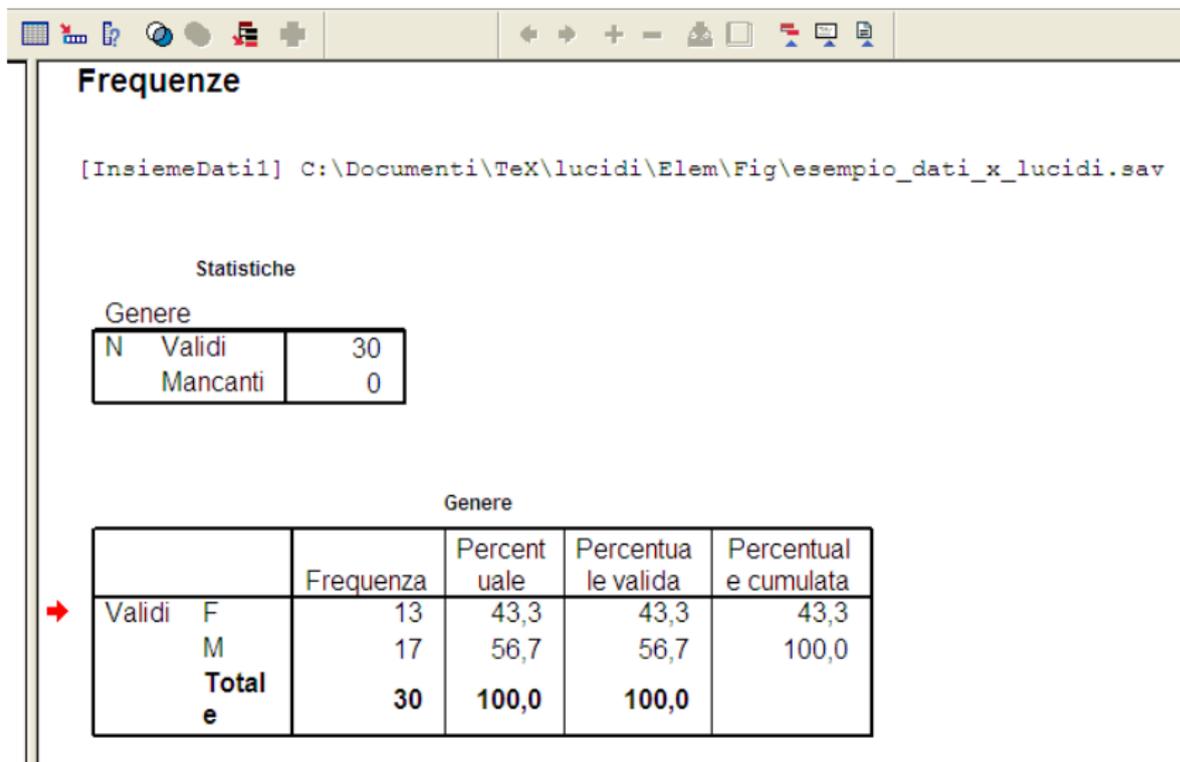
- Calcoliamo le frequenze con il comando

Analizza |  
Statistiche descrittive |  
Frequenze...

- Poi spostiamo Genere fra le Variabili
- E premiamo OK



# Spss: frequenze



The screenshot shows the SPSS 'Frequenze' (Frequencies) output window. At the top, the file path is displayed: [InsiemeDati1] C:\Documenti\TeX\lucidi\Elem\Fig\esempio\_dati\_x\_lucidi.sav. Below this, the 'Statistiche' (Statistics) section is shown, with a table for 'Genere' (Gender) indicating 30 valid cases and 0 missing cases. The main part of the output is a table titled 'Genere' with columns for 'Frequenza' (Frequency), 'Percentuale' (Percentage), 'Percentuale valida' (Valid Percentage), and 'Percentuale cumulata' (Cumulative Percentage). The rows represent 'Validi F' (13 cases, 43.3%), 'Validi M' (17 cases, 56.7%), and 'Total e' (30 cases, 100.0%). A red arrow points to the first row of the main table.

**Frequenze**

[InsiemeDati1] C:\Documenti\TeX\lucidi\Elem\Fig\esempio\_dati\_x\_lucidi.sav

**Statistiche**

Genere

N	Validi	30
	Mancanti	0

**Genere**

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	F	13	43,3	43,3	43,3
	M	17	56,7	56,7	100,0
<b>Total e</b>		<b>30</b>	<b>100,0</b>	<b>100,0</b>	

# Spss: n-tili (Frequenze)

- In **Statistiche...**
- **Quartili** calcola i quartili
- **Punti di divisione** divide in **n parti uguali**
- **Percentili**: scrivete il **percentile che volete e aggiungete**

Fasce di età

N	Validi	100
	Mancanti	0
Percentili	25	1,00
	33,33333333	2,00
	45	2,00
	50	2,00
	66,66666667	3,00
	75	3,00

Valori percentili

Quartili

Punti di divisione per:  gruppi uguali

Percentili:

Aggiungi

Cambia

Rimuovi

Qui abbiamo chiesto: i quartili, i terzili e il 45esimo percentile  
← Qui i risultati espressi come percentili

# Spss: Esplora...

- Dopo aver scelto le variabili, metterle in Variabili dipendenti



- poi

- click-are su

Statistiche...

e

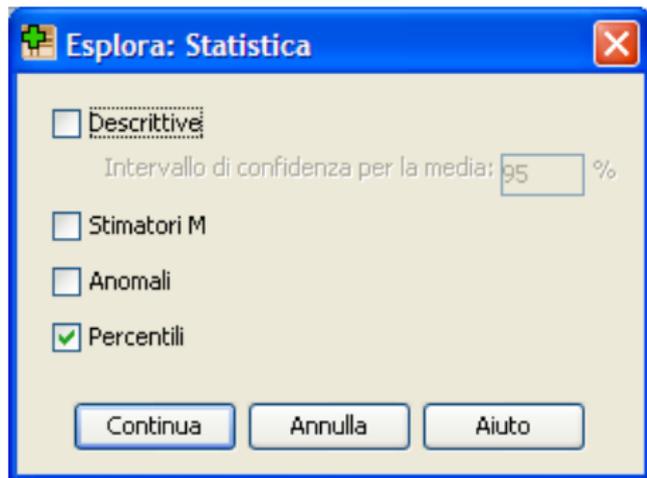
scegliere Percentili

- Quindi, click-are su

Continua

- Poi

OK



Esplora non permette di scegliere, ma fornisce alcuni n-tili notevoli

Percentili

		Percentili						
		5	10	25	50	75	90	95
Media ponderata (Definizione 1)	Fasce di età	1,00	1,00	1,00	2,00	3,00	3,00	3,00
Cardini di Tukey	Fasce di età			1,00	2,00	3,00		

I tre quartili e i valori corrispondenti al 5% e 10% su entrambi i lati. Questi valori hanno particolarmente senso con variabili “normali” (capiremo più avanti).

# Tendenza centrale

Conoscete queste Statistiche della **tendenza centrale**?

- Moda
- Mediana
- Media

*(Alzate la mano se NON sapete una di queste cose)*

# Tendenza centrale

Conoscete queste Statistiche della **tendenza centrale**?

- Moda
- Mediana
- Media

*(Alzate la mano se NON sapete una di queste cose)*

Se sapete tutto, saltiamo al lucido **30**

Ripassare le statistiche della tendenza centrale

<http://www.germanorossi.it/mi/file/elem/Elem04-TendCentrale.pdf>

# Tendenza centrale

- La tendenza centrale è un'indicazione generica di come sta andando la distribuzione della variabile
- Ci sono diversi indici che “misurano” la tendenza centrale, alcuni poco informativi, altri molto informativi

Livello	Statistica
Nominale	Moda
Ordinale	Mediana
Intervallo/rapporto	Media

- Ricordiamo che ogni livello “eredita” dai livelli precedenti

# Tendenza centrale: Moda

- La **Moda** ( $Mo$ ) è la frequenza più elevata di una distribuzione
- Se c'è una sola moda, la distribuzione si dice **Unimodale**

## Esempio

$M=17, F=13 \Rightarrow$  **Maschi** perché ha frequenza 17

Se ci sono molte categorie, oppure poche categorie tutte con frequenze simili, la moda non ha molto senso.

## Esempio

1 1 1 2 2 2 2 3 3 3       $Mo=2$  (ma non ha molto senso)

# Spss: moda

## ■ Tramite

Analizza |  
Statistiche descrittive |  
Frequenze...  
pulsante

Statistiche

,  
possiamo far  
calcolare la moda.

## ■ poi Continua e OK

Frequenze: Statistiche

Valori percentili

Quartili

Punti di divisione per: 10 gruppi uguali

Percentili:

Aggiungi

Cambia

Rimuovi

Tendenza centrale

Media

Mediana

**Moda**

Somma

I valori sono punti centrali di gruppi

Dispersione

Deviazione stand.  Minimo

Varianza  Massimo

Intervallo  Errore standard della media

Distribuzione

Asimmetria

Curtosi

Continua

Annulla

Aiuto

Statistiche		
Genere		
→ N	Validi	30
	Mancanti	0
	Moda	1

# Tendenza centrale: Mediana

- La **mediana** (Mdn) è il valore che divide esattamente la distribuzione a metà (corrisponde, ma è teoricamente diversa, a  $Q_2$ )
- Se la variabile è ordinale la mediana dovrebbe essere un valore esistente (il valore centrale se N è dispari; i due valori centrali se pari)
- Se la variabile è quantitativa la mediana è un valore stimato come interpolazione lineare

## Attenzione

Spss (e la maggior parte dei software statistici) fanno **sempre** l'interpolazione!

# Tendenza centrale: Media [aritmetica]

La **media aritmetica** ( $\bar{X}$ , Md, M) è la somma ( $\sum$ ) di tutti i valori di una distribuzione, divisa per la numerosità (N o N-1)

$$\bar{X} = \frac{\sum_{i=1}^N X_i}{N} = \frac{\sum X}{N} \text{ oppure } \frac{\sum X}{N - 1}$$

## Esempio

M(10, 15, 16, 18, 20, 24, 32, 35, 38, 40) = 24.8

$$\frac{10 + 15 + 16 + 18 + 20 + 24 + 32 + 35 + 38 + 40}{10} = \frac{248}{10}$$

# Spss: media

La media viene visualizzata da Spss in molte procedure. Quelle specifiche sono:

- `Analizza | Statistiche descrittive | Frequenze...` (fra le varie statistiche che è possibile stampare vi è anche la media)
- `Analizza | Statistiche descrittive | Descrittive...` (è la procedura specifica per le statistiche descrittive)
- `Analizza | Statistiche descrittive | Esplora...` (stampa la media come una delle diverse statistiche per capire l'andamento e la distribuzione di una variabile)

# Spss: Media con Frequenze...

- Dopo aver scelto le variabili, click-are su  e selezionare Media
- Quindi, click-are su
- Con variabili quantitative conviene de-selezionare anche  Visualizza tabelle di frequenza
- oppure in 
  - Sopprimi le tabelle con più di n categorie
  - Numero massimo di categorie:
- Poi

Tendenza centrale

Media

Mediana

Moda

Somma

# Spss: Media con Descrittive...

- Dopo aver scelto le variabili, click-are su **Opzioni...**
- Normalmente **Media** è già selezionato
- Potete ordinare i risultati in vari modi
- Poi **OK**

**Statistiche descrittive**

	N	Media
gruppo	100	,44
Antisemitismo	100	49,6300
Validi (listwise)	100	

**Descrittive: Opzioni**

Media  Somma

**Dispersione**

Deviazione stand.  Minimo

Varianza  Massimo

Intervallo  Errore standard della media

**Distribuzione**

Curtosi  Asimmetria

**Ordine di visualizzazione**

Variabili nel file

Alfabetico

Medie crescenti

Medie decrescenti

Continua Annulla Aiuto

# Variabilità

Conoscete queste Statistiche della **variabilità**?

- Intervallo (o gamma di oscillazione)
- Varianza
- Deviazione standard

*(Alzate la mano se NON sapete una di queste cose)*

# Variabilità

Conoscete queste Statistiche della **variabilità**?

- Intervallo (o gamma di oscillazione)
- Varianza
- Deviazione standard

*(Alzate la mano se NON sapete una di queste cose)*

Se sapete tutto, Abbiamo finito

Ripassare le statistiche di variabilità

<http://www.germanorossi.it/mi/file/elem/Elem05-IndiciVariabilita.pdf>

# Misure di variabilità

Gli indici di variabilità ci dicono quanto i valori sono dispersi attorno alla tendenza centrale.

## Esempio

<i>valori</i>	$\sum X/N = M$
7 7 6 5 4 4 4 3	40/8=5
10 10 9 7 5 4 3 2 0 0	50/10=5

A livello di scala intervallo/rapporto ci sono diversi indici di variabilità:

- Campo di variazione, gamma (di oscillazione), *range*, intervallo
- Differenza interquartilica (DI o IQR)
- Deviazione media o scostamento semplice medio
- Varianza e deviazione standard

# Misure di variabilità: intervallo

- Il **campo di variazione** o **gamma** (di oscillazione) o **range** o **intervallo** (per SPSS) è la differenza fra il valore massimo e quello minimo

$$\text{gamma} = \text{max} - \text{min}$$

## Esempio

<i>valori</i>	<i>campo var.</i>
3 4 4 4 5 6 7 7	7-3=4
0 0 2 3 4 5 7 9 10 10	10-0=10

# Misure di variabilità: differenza interquartilica

- La **differenza interquartilica** (DI, IQR) è la differenza fra il terzo e il primo quartile

$$IQR = Q_3 - Q_1$$

e corrisponde al 50% centrale dei valori della distribuzione

## Esempio

<i>valori</i>	N	Q3-Q1	<i>IQR</i>
3 <b>4</b> 4 4 5 <b>6</b> 7 7	8	6-4	2
0 <b>0</b> 2 3 4 5 7 <b>9</b> 10 10	10	9-0	9

- La semi-differenza interquartilica è la metà dell'IQR e corrisponde al 25% dei valori sopra o sotto la mediana

- **Varianza ( $s^2$ , var)**: elevare gli scarti a quadrato e fare la loro media (N o N-1)

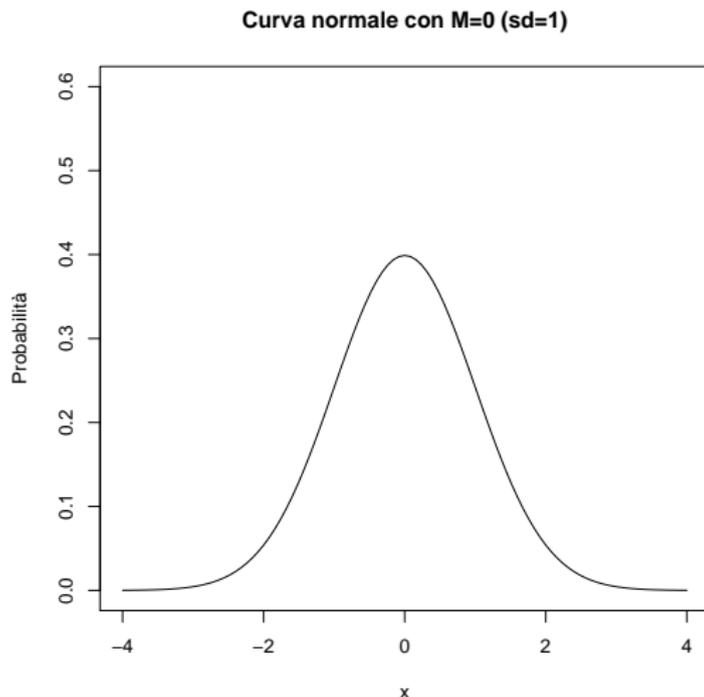
$$s^2 = var = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

- Però la varianza è un quadrato (un'area) e quindi si introduce anche una versione lineare (che è una distanza), lo scarto quadratico medio.
- Lo **scarto quadratico medio (sqm)** o **deviazione standard (ds, s)** è la radice quadrata della varianza

$$s = \sqrt{s^2} = \sqrt{var} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N - 1}}$$

# Misure di variabilità: curtosi e asimmetria

- **Curtosi:** è il grado di piattezza della curva della densità delle frequenze rispetto ad una curva particolare che è la **normale** (valore di riferimento = 0);
- **Asimmetria:** è il grado di asimmetria della curva della densità delle frequenze rispetto ad una curva particolare che è la **normale** (valore di riferimento = 0);



# Spss: Variabilità

Anche varianza e dev. st. sono visualizzate da Spss (sempre N-1) in molte procedure. Quelle specifiche sono:

- `Analizza | Statistiche descrittive | Frequenze...` (fra le varie statistiche vi è anche quelle di variabilità)
- `Analizza | Statistiche descrittive | Descrittive...` (è la procedura specifica per le statistiche descrittive)
- `Analizza | Statistiche descrittive | Esplora...` (stampa le misure di variabilità come parte delle diverse statistiche per capire l'andamento e la distribuzione di una variabile)

# Spss: Variabilità con Frequenze...

- Dopo aver scelto le variabili, click-are su **Statistiche...** e selezionare quelle che servono

Dispersione	Distribuzione
<input type="checkbox"/> Deviazione stand.	<input type="checkbox"/> Minimo
<input type="checkbox"/> Varianza	<input type="checkbox"/> Massimo
<input type="checkbox"/> Intervallo	<input type="checkbox"/> Errore standard della media
	<input type="checkbox"/> Asimmetria
	<input type="checkbox"/> Curtosi

- Quindi, click-are su **Continua**
- Con variabili quantitative conviene selezionare anche

<input checked="" type="checkbox"/> Visualizza tabelle di frequenza
<input type="checkbox"/> Sopprimi le tabelle con più di n categorie
Numero massimo di categorie: <input type="text" value="10"/>

- oppure in

**Formato...**

- Poi **OK**

# Spss: Variabilità con Descrittive...

- Dopo aver scelto le variabili, click-are su

Opzioni...

## Dispersione

- Deviazione stand.     Minimo  
 Varianza                 Massimo  
 Intervallo                 Errore standard della media

## Distribuzione

- Curtosi                     Asimmetria

- Poi  e

Statistiche descrittive

	N	Intervallo	Minimo	Massimo	Deviazione std.	Varianza	Asimmetria		Curtosi	
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Errore std	Statistica	Errore std
gruppo	100	1	0	1	,499	,249	,245	,241	-1,980	,478
Antisemitismo	100	16,00	42,00	58,00	3,64499	13,286	-,092	,241	-,523	,478
Validi (listwise)	100									