

# **ANALISI DEI DATI + LABORATORIO**

**prof. Claudio Barbaranelli**

**Dip. di Psicologia, via dei Marsi 78, 00185 Roma  
tel. 06/49917623 – Stanza 212  
claudio.barbaranelli@uniroma1.it**

**Ricevimento studenti:**

**Fino alla cessazione dell'emergenza Covid, il ricevimento va concordato con il docente tramite email ed avverrà sulla piattaforma Google/Meet**

**Eventuali variazioni rispetto a quanto specificato sopra verranno comunicate per tempo su elearning2.**

# ANALISI DEI DATI + LABORATORIO

prof. Claudio Barbaranelli

## Orario Lezioni

Aula	Orario
<b>Aula 8</b>	<b>Lunedì 11:30-14:00</b>
	<b>Giovedì 8:30-10:30</b>
	<b>Venerdì 9:00-11:30</b>

<b>inizio 1° semestre</b>	<b>28/09/2020</b>
<b>fine 1° semestre</b>	<b>18/12/2018</b>

# ANALISI DEI DATI + LABORATORIO

## CONTENUTO DEL CORSO

**Il corso riguarderà i seguenti argomenti:**

- I trattamenti preliminari dei dati**
- La regressione lineare multipla**
- L'analisi Fattoriale Esplorativa**
- I modelli di equazioni strutturali**
- L'Analisi della Varianza (ANOVA)**

**Le applicazioni informatiche verranno effettuate con i programmi SPSS e MPLUS**

# ANALISI DEI DATI + LABORATORIO

## TESTI DI RIFERIMENTO

- a) Barbaranelli, C. (2007). **Analisi dei dati. II edizione. Milano: Led. (capitoli 1, 2, 3, 4, appendici 1 e 2).**
- b) Barbaranelli, C. (2006). **Analisi dei dati con SPSS: Le analisi multivariate. Milano: Led. (capitoli 1, 2 e 3).**
- c) Barbaranelli, C. e D'Olimpio, F. (2007). **Analisi dei dati con SPSS: Le analisi di base. Milano: Led. (capitoli 1, 2, 3, 4 e 6).**
- d) **Lucidi e materiale integrativo presentato a lezione. Questo materiale è disponibile sul sito [www.elearning2.uniroma1.it](http://www.elearning2.uniroma1.it).**

**IL TESTO "INTRODUZIONE AI MODELLI DI EQUAZIONI STRUTTURALI" NON SARA' DISPONIBILE ED E' SOSTITUITO DALLO SCRITTO "NOTE SUI SEM" SCARICABILE DAL SITO [www.elearning2.uniroma1.it](http://www.elearning2.uniroma1.it).**



# ANALISI DEI DATI + LABORATORIO

## METODI DIDATTICI

**Gli argomenti del corso verranno presentati attraverso lezioni prevalentemente frontali sollecitando un ruolo attivo da parte degli studenti.**

**Le ore di laboratorio si alterneranno con le lezioni teoriche e prevedranno esercitazioni su MPLUS e su SPSS.**

**Gli studenti possono scaricare la *DEMO version* di MPLUS dal sito: <http://www.statmodel.com>**

**Tale versione è gratuita ed ha una licenza perpetua.**

**Gli studenti possono scaricare SPSS (licenza autorizzata per gli studenti e il personale della Sapienza) dal sito della Sapienza. Istruzioni alternative per scaricare e installare il software sono presenti nella pagina elearning del corso.**

## ANALISI DEI DATI + LABORATORIO

**MODALITÀ DI FREQUENZA:** La frequenza alle lezioni e ai laboratori non è obbligatoria, ma raccomandata.

**MODALITÀ D'ESAME:** L'esame prevede una prova scritta costituita da:

- domande a risposta chiusa e aperta relative ai testi in programma;
- esercizi sull'interpretazione di output dei programmi SPSS e MPLUS;
- esercizi sulla programmazione in linguaggio MPLUS.

Esempi di esercizi su MPLUS sono scaricabili dal sito

<http://elearning2.uniroma.it>

Per sostenere la prova è necessario prenotarsi entro i termini definiti sul sito della Facoltà.

Le modalità d'esame NON saranno differenziate per studenti frequentanti e non frequentanti.

# TRATTAMENTI PRELIMINARI DEI DATI

# Trattamenti preliminari dei dati

## Sommario

- \* **Forma della distribuzione**
- \* **Valori anomali (outliers) univariati**
- \* **Normalità bivariata e multivariata**
- \* **Outlier multivariati**
- \* **Le informazioni mancanti (*missing values*)**

## Forma della distribuzione

### Distribuzione Normale Univariata

**Forma "a campana", unimodale, simmetrica rispetto alla media (quindi media e mediana coincidono, e coincidono anche con la moda), presenta due punti di flesso per  $x = \mu - \sigma$ , e  $x = \mu + \sigma$ .**

**Famiglia di distribuzioni normali univariate: diverse distribuzioni normali sono definite da due parametri, la **media** ( $\mu$ ) e la **deviazione standard** ( $\sigma$ ) della distribuzione.**

**Funzione di probabilità della distribuzione normale:**

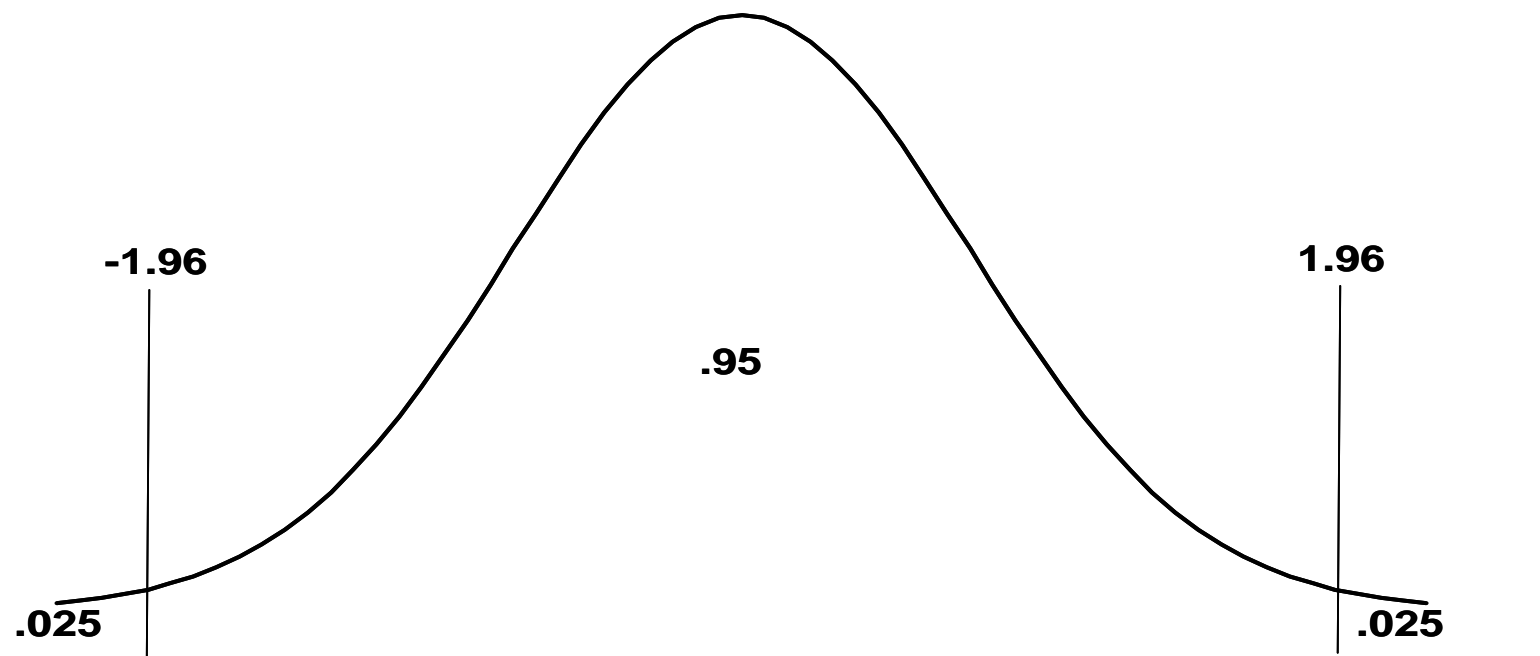
$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

# Forma della distribuzione

## Normale standardizzata

La probabilità dei suoi valori è stata tabulata: ciò la rende particolarmente utile nella verifica delle ipotesi statistiche.

$$P\left(\left[\frac{X-V}{\sigma_E}\right] \in [-1.96, 1.96]\right) = .95$$



# Forma della distribuzione

## Esame della normalità della distribuzione

**Diversi metodi per esaminare se una variabile è normale. Le informazioni di questi diversi metodi vanno integrate.**

- Indici di forma della distribuzione**
- Test statistici**
- Metodi grafici**

## Forma della distribuzione

**Indici che valutano la forma della distribuzione:  
Curtosi e Asimmetria (o *skewness*)**

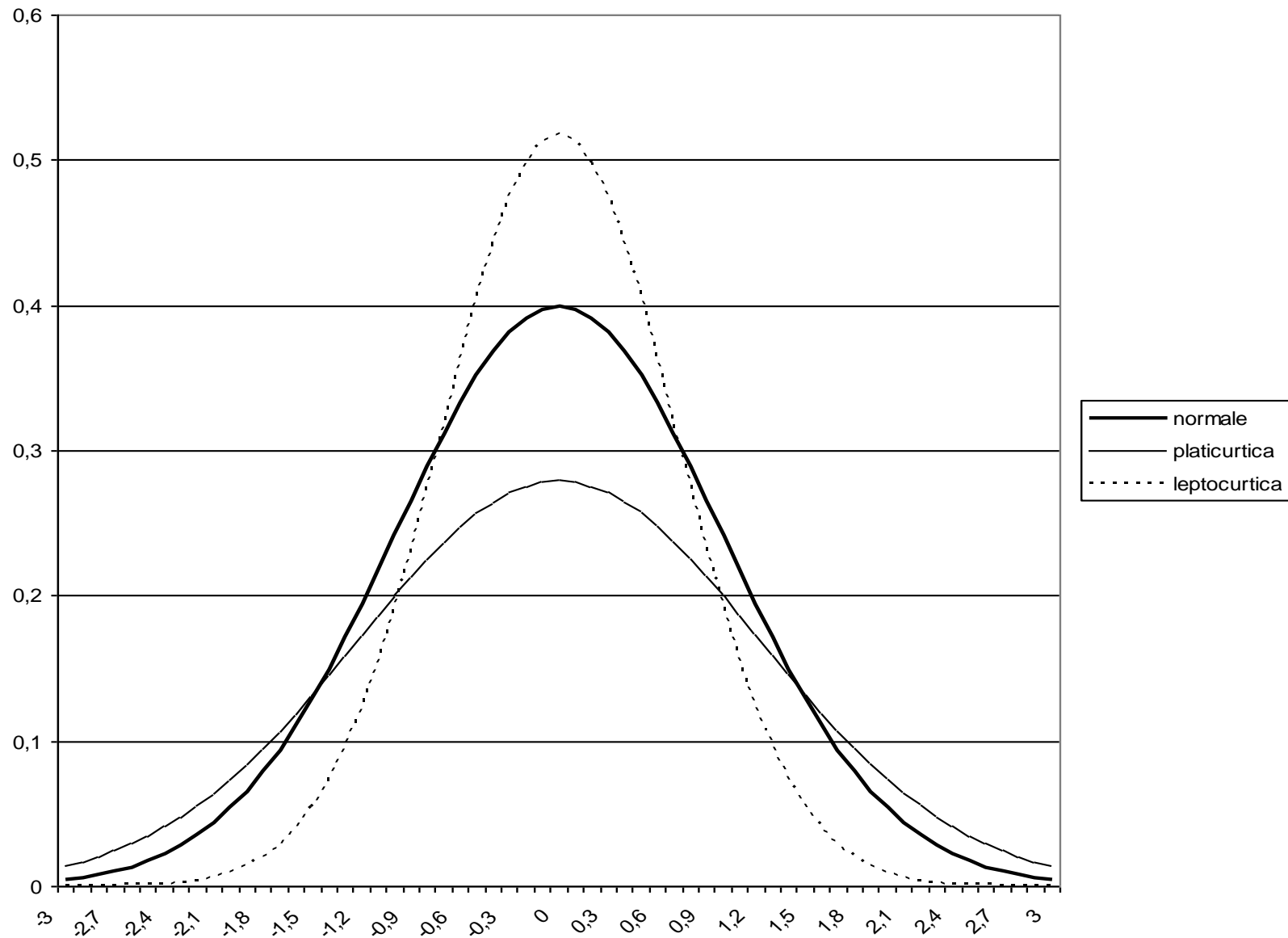
**Curtosi:** riflette il grado in cui i punteggi sono distribuiti nelle code piuttosto che nelle zone centrali della distribuzione. Uguale a 0 quando la distribuzione è perfettamente normale.

**Curtosi Negativa:** distribuzione platicurtica, “più schiacciata”, i valori estremi sono più frequenti rispetto alla normale.

**Curtosi Positiva:** distribuzione leptocurtica, “più appuntita”, i valori estremi sono meno frequenti.



# Forma della distribuzione



## Forma della distribuzione

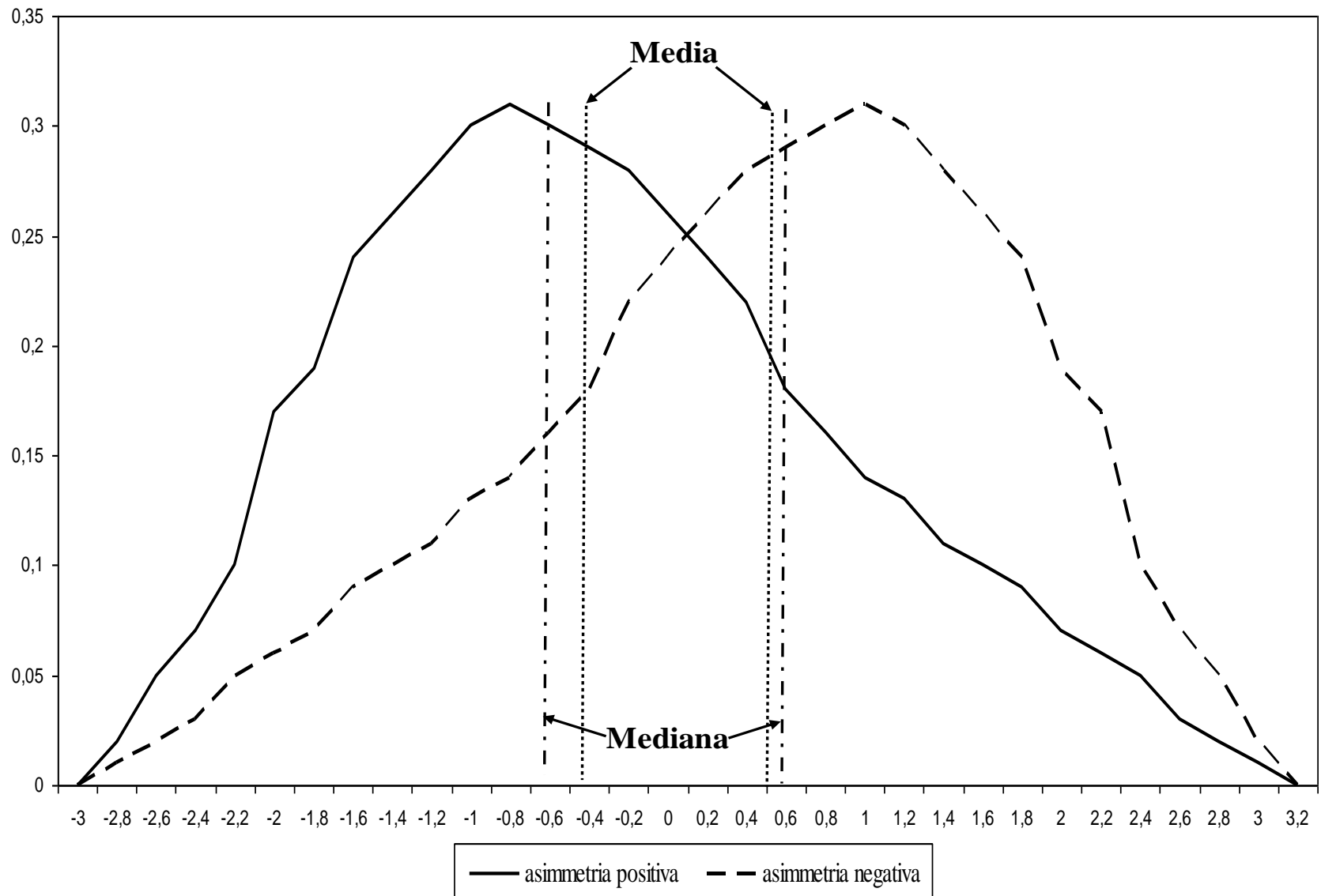
**Indici che valutano la forma della distribuzione:  
Curtosi e Asimmetria (o *skewness*)**

**Asimmetria:** riflette il grado in cui la distribuzione è disposta simmetricamente attorno ai valori di tendenza centrale. Uguale a 0 quando la distribuzione è perfettamente normale.

**Asimmetria positiva:** i valori bassi hanno frequenza maggiore, la media risulta maggiore della mediana.

**Asimmetria negativa:** i valori alti sono più frequenti, la media risulta inferiore alla mediana.

# Forma della distribuzione



## Forma della distribuzione

### Formula per la curtosi

$$\frac{\sum_{i=1}^N (x_i - \bar{X})^4}{N} \bigg/ \left( \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N} \right)^2$$

**Errore standard della curtosi =  $(24/N)^{1/2}$**

**Di solito viene sottratto il valore 3 per rendere la curtosi uguale a 0 nel caso di perfetta distribuzione normale.**

# Forma della distribuzione

## Formule per l'asimmetria

$$\left( \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{N} \right)^2 \bigg/ \left( \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} \right)^3$$

**Errore standard della asimmetria =  $(6/N)^{1/2}$**

$$3 \frac{(\bar{x} - \text{Mediana})}{S_x}$$

## Forma della distribuzione

**Verifica delle ipotesi per asimmetria e curtosi:** dividere il singolo indice (di asimmetria o di curtosi) per il suo errore standard, ed utilizzare come distribuzione di riferimento la normale standardizzata. Test troppo potente, ovvero risulta significativo quasi sempre.

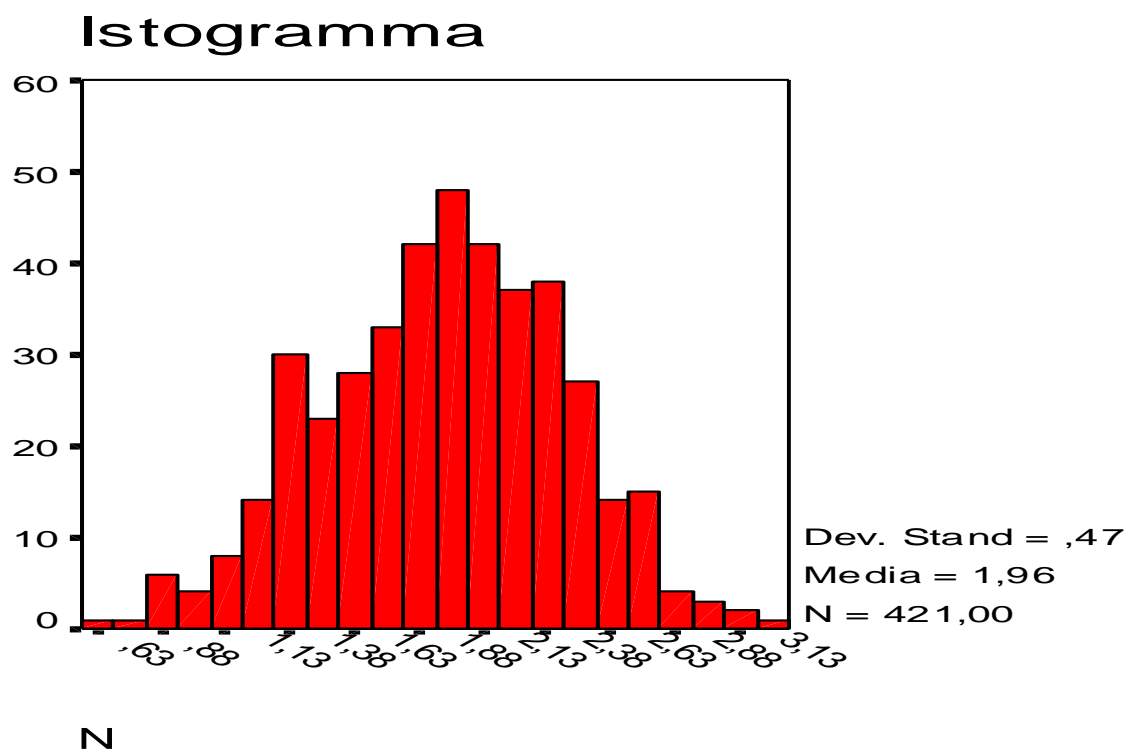
**Criterio empirico:** accettabili valori compresi tra  $-1$  e  $1$   
Nella verifica delle ipotesi su asimmetria e curtosi utilizzare un livello di alpha più basso ( $.01$  o  $.001$ ).

**Test statistici di Normalità: Kolmogorov-Smirnov e Shapiro-Wilk.** Se risultano significativi si deve rifiutare l'ipotesi nulla che la distribuzione sia normale. Test molto potenti che conducono troppo spesso al rifiuto dell'ipotesi nulla.

# Forma della distribuzione

## Grafici per l'esame della normalità

### Istogramma della distribuzione di frequenze della variabile



## Forma della distribuzione

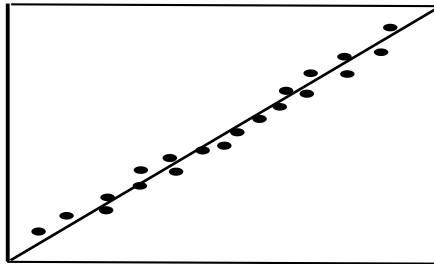
### Grafici per l'esame della normalità

#### **Plot dei quantili, o *Q-Q Plot* o *Cumulative Normal Plot***

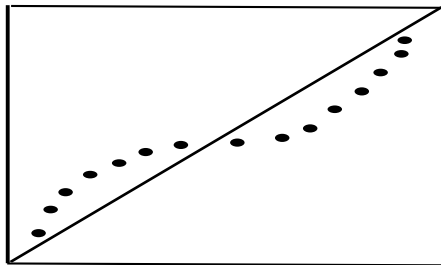
**Si confrontano i quantili della distribuzione della variabile, rispetto ai quantili della distribuzione normale. In ascissa sono riportati i valori osservati, in ordinata i valori attesi se la distribuzione è normale. Se la variabile si distribuisce in forma normale, i punti di tale distribuzione congiunta sono addensati sulla diagonale.**



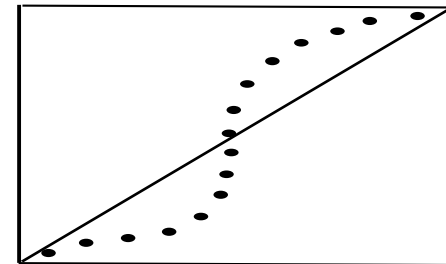
# Forma della distribuzione



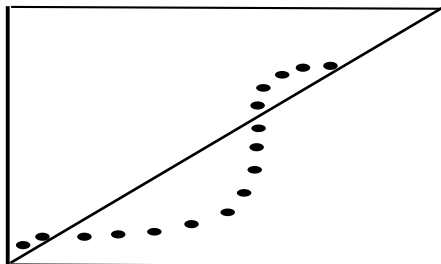
1. Normale



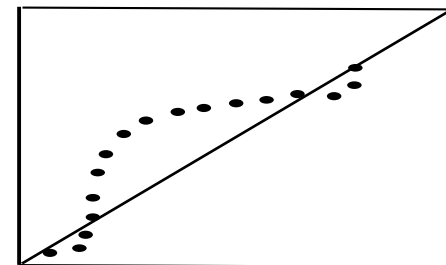
2. Platicurtica



3. Leptocurtica



4. Asimmetria negativa



5. Asimmetria Positiva

## Valori anomali (outliers) univariati

**Valori che si distinguono in maniera particolare rispetto agli altri valori nella distribuzione.**

**Outliers univariati:** casi che in una variabile presentano valori estremamente elevati o estremamente bassi.

**Gli outliers possono influenzare: la media, la deviazione standard, l'asimmetria e la curtosi, il coefficiente di correlazione di Pearson.**

**Indici che risultano meno influenzati dagli outliers: mediana e moda;**

**Statistiche "robuste" (es., media "*trimmed*" calcolata eliminando il 5% dei casi con punteggi più elevati e più bassi).**

## Valori anomali (outliers) univariati

### Individuare gli outliers univariati

**Standardizzare i punteggi relativi alla variabile in esame e calcolare una distribuzione delle frequenze.**

**Sono possibili outliers i casi che presentano un punteggio  $z$  maggiore di  $|3|$ .**

**Esame della distribuzione per vedere se i punteggi troppo elevati sono casi isolati dal resto dei punteggi.**

## Valori anomali (outliers) univariati

### Individuare gli outliers univariati

In alternativa è possibile utilizzare il valore assoluto mediano (MAD) secondo una formula più complessa:

$$MAD_i = |X_i - Mdn| / (1.483 * MAD)$$

**MAD = mediana della variabile  $|X_i - Mdn|$**

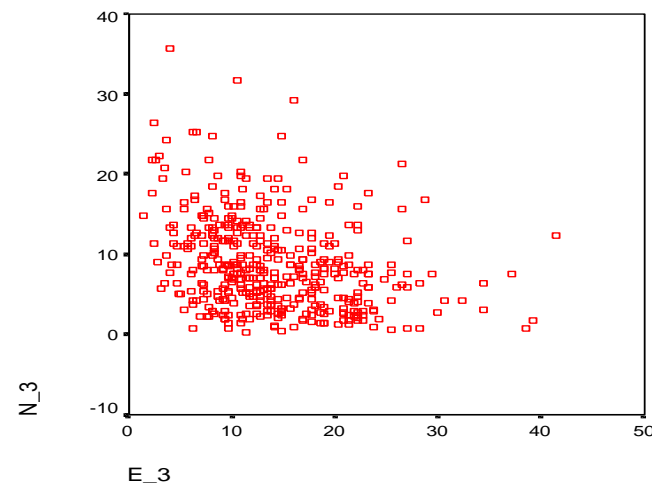
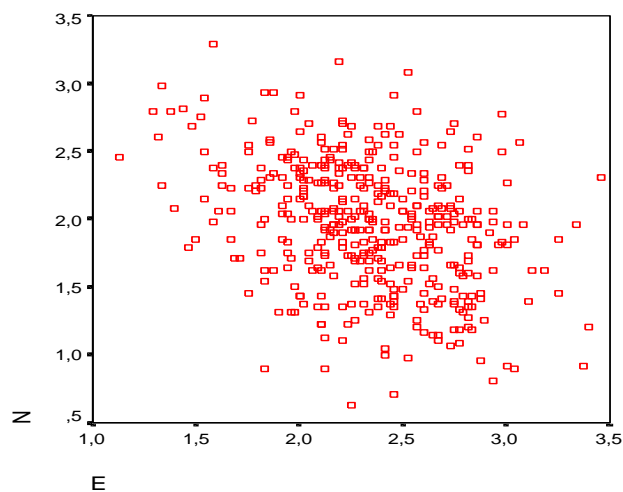
**Miller (1991) propone di considerare come outlier i soggetti che presentano punteggi  $MAD_i$  maggiori di  $|3|$  o di  $|2.5|$  (scelta moderatamente conservativa)**

# Linearità

## Linearità della relazione tra due variabili

**Relazione lineare tra X e Y: la variazione nei punteggi in Y attesa in concomitanza di una variazione di punteggi in X è costante per tutti i valori di X.**

## Diagramma di dispersione (o scatterplot)



## Trasformazioni delle variabili 1

**Non-linearità e non-normalità: fenomeni collegati.  
Tecniche per rendere "normale" la distribuzione  
(Tabachnick e Fidell, 1994, 2013)**

<b>Problema</b>	<b>Trasformazione [<math>X^* = f(X)</math>]</b>
<b>Asimmetria Positiva Estrema (valori &gt;2)</b>	<b>Reciproco: <math>X^* = 1/X</math></b>
<b>Asimmetria Positiva Sostanziale (valori tra 1 e 2)</b>	<b>Logaritmo: <math>X^* = \text{Log}_{10}(X)</math></b>
<b>Asimmetria Positiva Moderata (valori tra .5 e 1)</b>	<b>Radice Quadrata <math>X^* = \sqrt{X}</math></b>
<b>Asimmetria Negativa Moderata (valori tra -.5 e -1)</b>	<b>Radice Quadrata <math>X^* = \sqrt{(K - X)}</math></b>
<b>Asimmetria Negativa Sostanziale (valori tra -1 e -2)</b>	<b>Logaritmo = <math>X^* = \text{Log}_{10}(K - X)</math></b>
<b>Asimmetria Negativa Estrema (valori &lt;-2)</b>	<b>Reciproco = <math>X^* = 1/(K - X)</math></b>

**Nb. K è uguale al valore più elevato della variabile X, +1**

## Trasformazioni delle variabili 2

### Trasformazioni di Box-Cox (Box e Cox, 1964)

**Si tratta di una serie di trasformazioni che servono a normalizzare una distribuzione ma sono più complicate da calcolare**

$$X^{(\lambda)} = \begin{cases} \frac{X^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ \log X, & \text{if } \lambda = 0. \end{cases}$$

**La costante  $\lambda$  serve a normalizzare i punteggi: il suo valore ottimale (che massimizza la correlazione tra i punteggi originali e quelli trasformati) può essere individuato con appositi algoritmi (possibile anche in SPSS, vedi Osborne, 2010,**

***<http://pareonline.net/getvn.asp?v=15&n=12>***

## **Trasformazioni delle variabili 3**

### **POMS (Little, 2013)**

**Percentage of Proportion of maximum scoring (POMS) è un tipo di trasformazione che NON serve per normalizzare i punteggi, ma per renderne omogenea l'unità di misura, senza ricorrere alla standardizzazione (che annulla le differenze nella variabilità). Dopo la trasformazione le variabili hanno tutte la stessa unità di misura.**

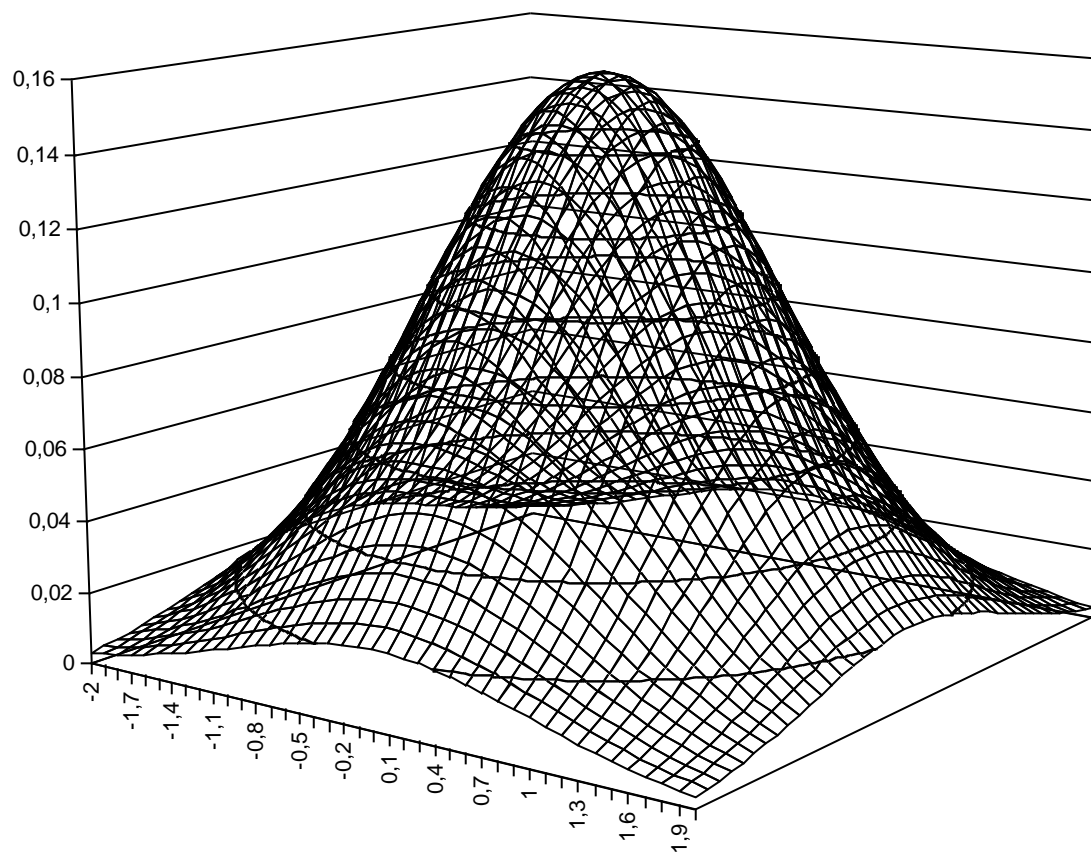
$$\text{Punteggio trasformato} = \frac{\text{Punteggio} - \text{minimo}}{\text{Massimo} - \text{minimo}}$$

**Il risultato è quello di trasformare le variabili su una scala che va da 0 (minimo) a 1 (massimo)**



## Normalità bivariata

Distribuzione normale bivariata: ciascuna delle 2 variabili è distribuita normalmente rispetto all'altra. La loro distribuzione congiunta ha la seguente forma:



## Normalità bivariata

Distribuzione normale bivariata: ciascuna delle 2 variabili è distribuita normalmente rispetto all'altra.

Funzione di probabilità della d.n.b.:

$$f(\mathbf{x}, \mathbf{y}; \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho_{xy}) =$$

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\left[\frac{(\mathbf{x}-\mu_x)^2}{2\sigma_x^2} + \frac{(\mathbf{y}-\mu_y)^2}{2\sigma_y^2}\right] - 2\rho\left[\frac{(\mathbf{x}-\mu_x)^2}{2\sigma_x^2} + \frac{(\mathbf{y}-\mu_y)^2}{2\sigma_y^2}\right]}$$

dove  $\mu_x$  e  $\mu_y$  sono le medie di  $x$  e  $y$ ,  $\sigma_x^2$  e  $\sigma_y^2$  sono le varianze di  $x$  e  $y$ , e  $\rho_{xy}$  è la correlazione tra  $x$  e  $y$ .

## Normalità multivariata

**Distribuzione normale multivariata**: generalizzazione della normale bivariata per  $k > 2$  variabili.

**Normalità multivariata**: assunzione che riguarda l'insieme delle variabili che vengono considerate in analisi.

**Funzione di probabilità della normale multivariata:**

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \text{EXP}\left(\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right)$$

dove  $p$  è il numero di variabili nel vettore  $\mathbf{y}$ ,  $\boldsymbol{\Sigma}$  è la matrice di varianze e covarianze tra le  $p$  variabili,  $|\boldsymbol{\Sigma}|$  è il suo determinante,  $\boldsymbol{\mu}$  è il centroide delle  $p$  variabili, e EXP è l'operatore della funzione esponenziale  $e^x$ .

**La funzione ha in tutto  $p(p+3)/2$  parametri.**

## Normalità multivariata

**La distribuzione multivariata di  $p$  variabili è normale se:**

- tutte le distribuzioni univariate delle variabili sono normali;**
- le distribuzioni congiunte di tutte le coppie di variabili seguono la distribuzione normale bivariata;**
- tutte le combinazioni lineari delle variabili sono normali.**

**Di solito se la distribuzione univariata di ogni singola variabile è normale, anche la distribuzione multivariata delle variabili lo è. Se c'è normalità multivariata, le relazioni tra le variabili considerate sono sicuramente lineari.**

## Normalità multivariata

**Valutare la normalità multivariata: Test grafico basato sui quantili della distribuzione del chi quadrato.**

**Distanza generalizzata o **distanza di Mahalanobis** per ogni singolo caso:**

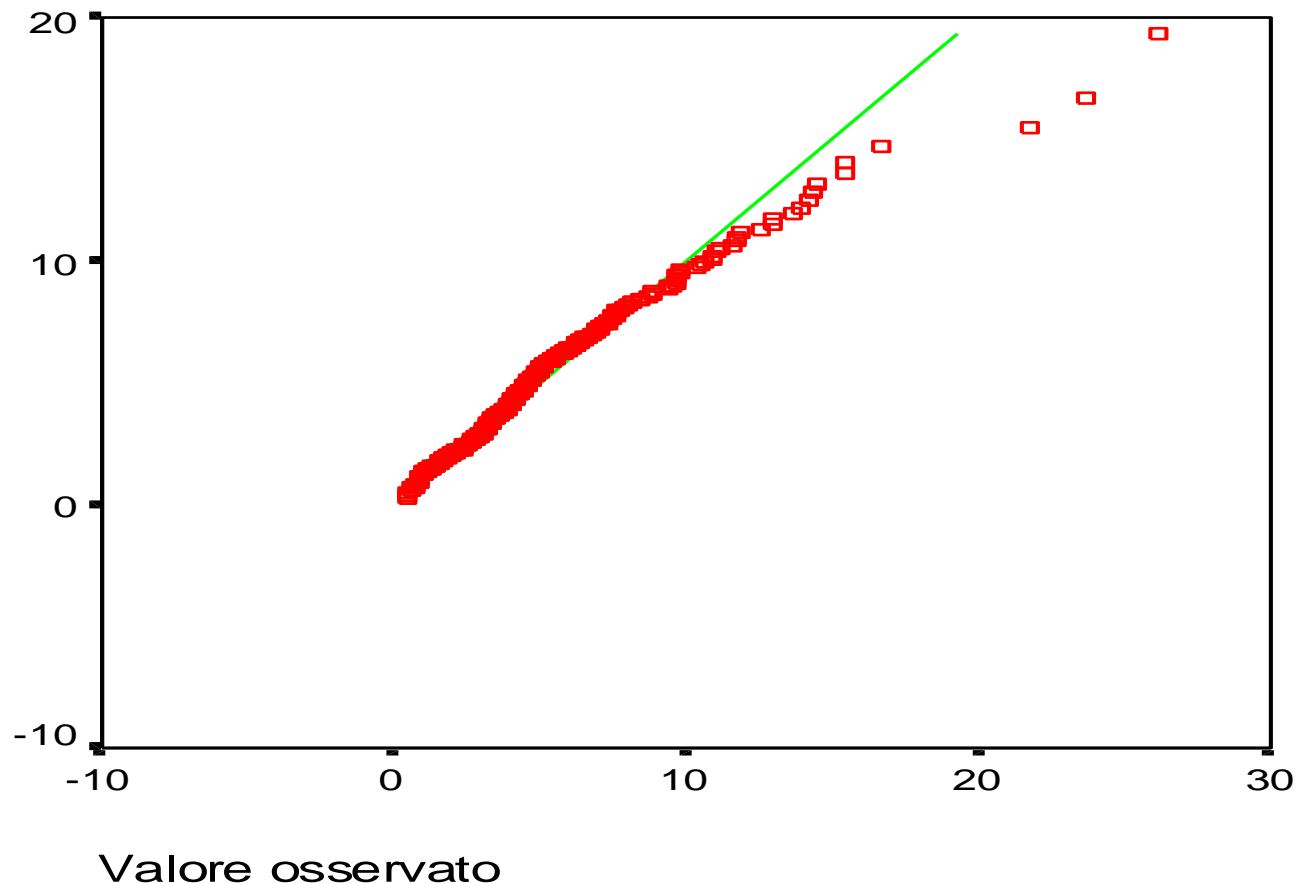
$$D_i^2 = (\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$$

**Rappresenta la distanza del vettore di punteggi di un soggetto ( $\mathbf{X}_i$ ) dal centroide del campione  $\bar{\mathbf{X}}$ , pesata per le var/covarianze ( $\mathbf{S}$ ).**

**Se la distribuzione delle variabili è normale multivariata e il numero dei casi meno il numero di variabili è maggiore di 25, la distanza generalizzata segue la **distribuzione del chi-quadrato**.**

# Normalità multivariata

## Grafico Q-Q Chi-quadrato di Mahal



Valore osservato

**In ascissa sono riportati i valori osservati ( $D^2$ ), in ordinata i valori attesi della distribuzione del chi-quadrato. Se la distribuzione è normale multivariata il grafico ha un andamento lineare.**

## Normalità multivariata

### Coefficiente di curtosi multivariata di Mardia

$$k = \frac{\sum_{i=1}^N (D_i^2)^2}{N}$$

**Se la distribuzione delle  $p$  variabili è normale multivariata, e se  $n > 50$  soggetti) il coefficiente di curtosi multivariata di Mardia è  $\leq p(p+2)$ .**

$$Z_k = \frac{k - E(k)}{\sqrt{\text{VAR}(k)}}$$

**$Z_k$  si distribuisce approssimativamente come una variabile normale standardizzata se il campione è sufficientemente ampio. Esame dell'ipotesi nulla che  $k < p(p+2)$ , con un test a due code per un livello di probabilità pari a  $\alpha/2$ .**

## Outlier multivariati

**Combinazioni dei punteggi delle singole variabili che risultano particolarmente "strani".**

**Casi che hanno una combinazione di punteggi particolarmente rara rispetto al resto del campione.**

**Si possono considerare outliers multivariati i casi in cui la distanza di Mahalanobis  $D^2$  risulta significativa al livello  $p < .001$  (Tabachnick e Fidell, 2007), prendendo come distribuzione di riferimento quella del chi-quadrato con  $p$  gradi di libertà (dove  $p$  = numero di variabili).**



## Le informazioni mancanti (*missing values*)

**In fase di codifica dei dati è bene che i valori mancanti siano opportunamente codificati, in modo da distinguerli dai valori *effettivi* che possono assumere le variabili.**

**In fase di analisi è necessario che il ricercatore decida cosa fare dei valori mancanti.**

**Ci sono diverse strategie possibili.**

## Le informazioni mancanti (*missing values*)

- a) la limitazione dell'analisi ai soli casi che presentano valori validi per tutte le variabili in esame (esclusione *listwise*);
- b) la limitazione dell'analisi ai casi che di volta in volta presentano valori validi nella coppia di variabili che viene considerata (esclusione *pairwise*);
- c) la **sostituzione** del valore mancante con la media della variabile nel campione, o con la media ottenuta dal soggetto nelle variabili considerate;
- d) la **sostituzione** del valore mancante con una sua stima ricavata tramite procedure statistiche (regressione, EM) effettuate sui soggetti che presentano dati completi.

# Statistical Package for Social Sciences



<https://www.spss.it/>

<https://web.uniroma1.it/infosapienza/>

# SPSS

## INTRODUZIONE A SPSS

- Le componenti fondamentali di SPSS
- Lo screening dei dati (es. valutare la normalità della distribuzione; come trattare i dati mancanti)
- L'analisi dei dati (statistiche descrittive, attendibilità, analisi degli item, analisi della varianza, correlazione e regressione, analisi fattoriale)

# SPSS

## LE COMPONENTI FONDAMENTALI DI SPSS

- **1. LE FINESTRE**
- **2. I MENÙ**
- **3. LE FINESTRE DI DIALOGO**
- **4. LE BARRE DEGLI STRUMENTI**
- **5. LA BARRA DI STATO**

# SPSS

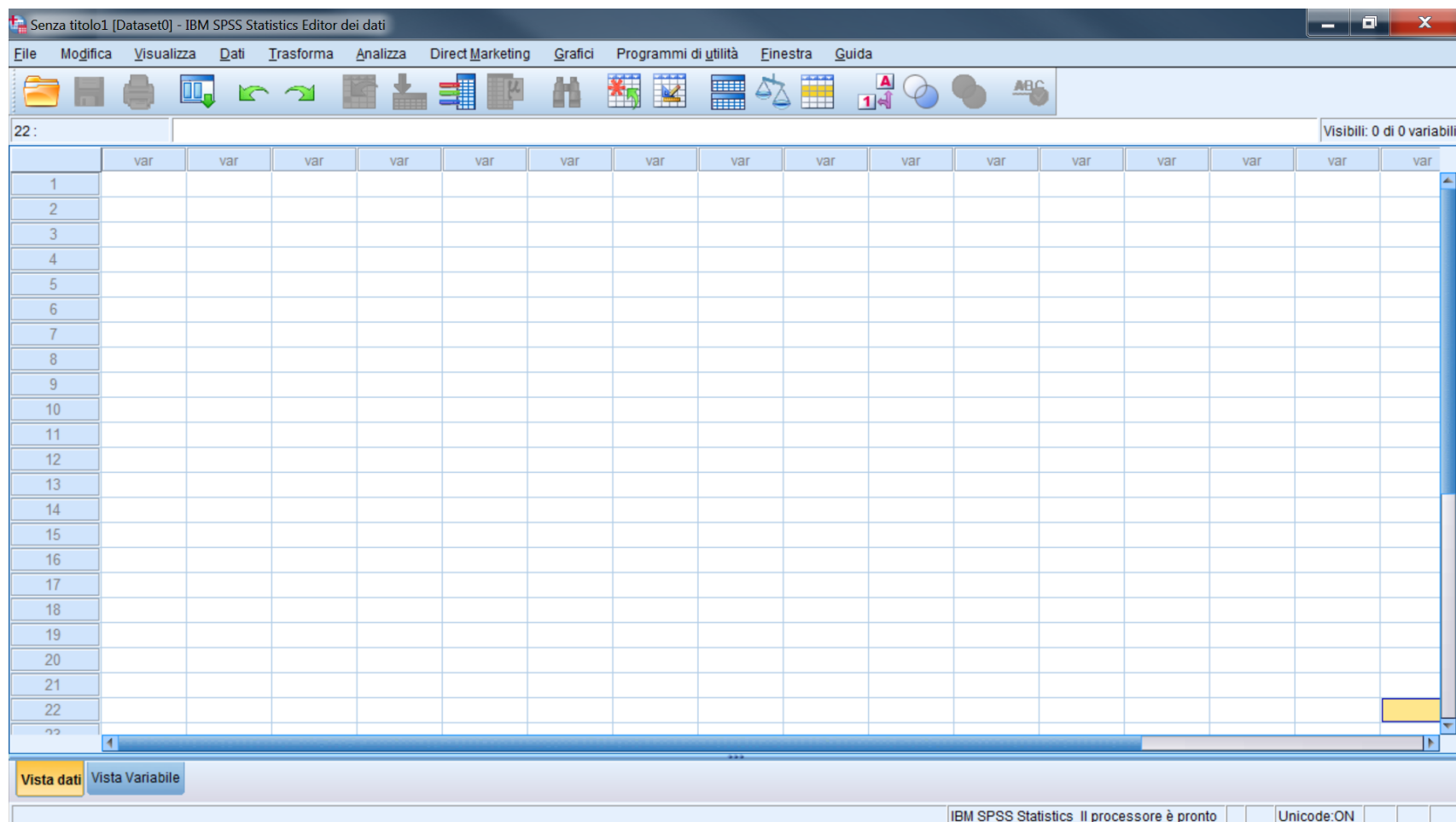
## LE FINESTRE DI SPSS

- **1. LA FINESTRA EDITOR DEI DATI**
- **2. LA FINESTRA VISUALIZZATORE**
- **3. LA FINESTRA SINTASSI**

# SPSS

## La finestra Editor dei dati

**QUESTA FINESTRA MOSTRA I CONTENUTI DEL FILE DEI DATI**



**La finestra DATA EDITOR si apre automaticamente ogni volta che ha inizio una sessione SPSS. Si possono aprire più data files alla volta.**

# SPSS

## La finestra Editor dei Dati

**Molte caratteristiche della finestra data editor sono simili a quelle dei fogli elettronici (es. excel). Vi sono comunque alcune importanti differenze.**

- Le righe corrispondono ai casi (unità). Ciascuna riga rappresenta un caso o un'osservazione. Ad esempio ciascun individuo che compila un questionario è un caso.
- Le colonne sono le variabili. Ciascuna colonna rappresenta una variabile o una caratteristica rilevata. Ad esempio ciascun item di un questionario è una variabile.
- Le celle contengono i valori. Ogni cella contiene un singolo valore di una variabile relativa ad un caso. La cella è l'intersezione di un caso con una variabile. **Diversamente da Excel, le celle contengono solo valori, e non possono contenere formule.**



# SPSS

## La finestra Editor dei Dati

- Il data file è rettangolare

**Le dimensioni del data file sono determinate dal numero di casi e di variabili**

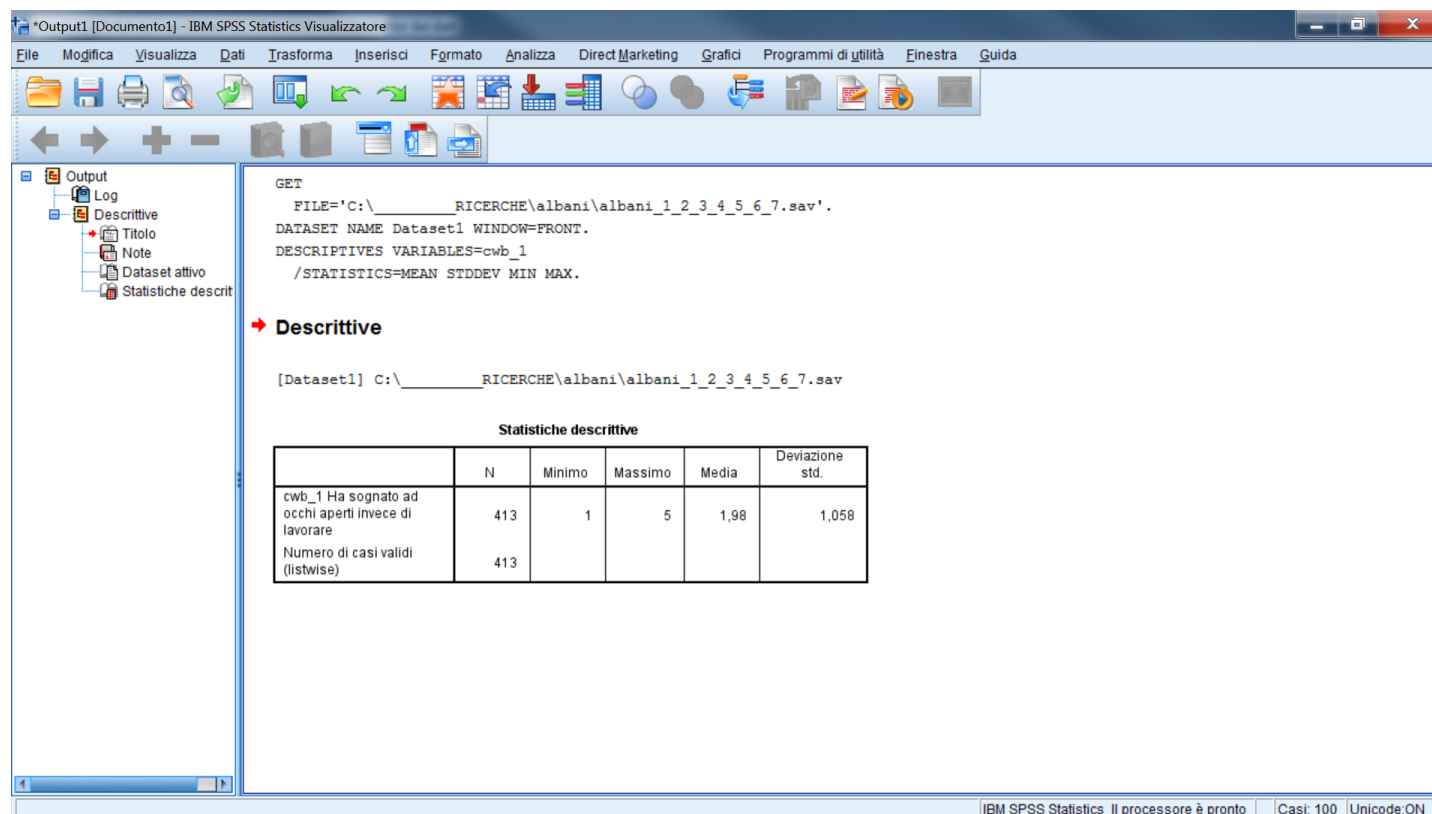
**Si possono inserire dati in ogni cella. Se si inseriscono dati in una cella che è al di fuori dei confini che delimitano il data file, il rettangolo dei dati si estende fino ad includere ogni riga e/o colonna tra questa cella e i confini del file**

**Le celle vuote all'interno dei confini del data file vengono considerate come un valore mancante, ovvero vengono convertite in valori mancanti "di sistema"**

# SPSS

## La finestra Visualizzatore

Questa finestra si apre automaticamente la prima volta che viene eseguita una procedura che genera un output  
Nella finestra Visualizzatore vengono mostrati tutti i risultati statistici, le tabelle e i grafici (output)



The screenshot shows the IBM SPSS Statistics Visualizzatore window. The window title is "\*Output1 [Documento1] - IBM SPSS Statistics Visualizzatore". The menu bar includes File, Modifica, Visualizza, Dati, Trasforma, Inserisci, Formato, Analizza, Direct Marketing, Grafici, Programmi di utilità, Finestra, and Guida. The toolbar contains various icons for file operations and analysis. The left pane shows a tree view with "Output" expanded, containing "Log", "Descrittive", "Titolo", "Note", "Dataset attivo", and "Statistiche descrittive". The main area displays the following content:

```
GET
FILE='C:\_____RICERCHE\albani\albani_1_2_3_4_5_6_7.sav'.
DATASET NAME Dataset1 WINDOW=FRONT.
DESCRIPTIVES VARIABLES=cw_b_1
/STATISTICS=MEAN STDDEV MIN MAX.
```

→ Descrittive

[Dataset1] C:\\_\_\_\_\_RICERCHE\albani\albani\_1\_2\_3\_4\_5\_6\_7.sav

Statistiche descrittive

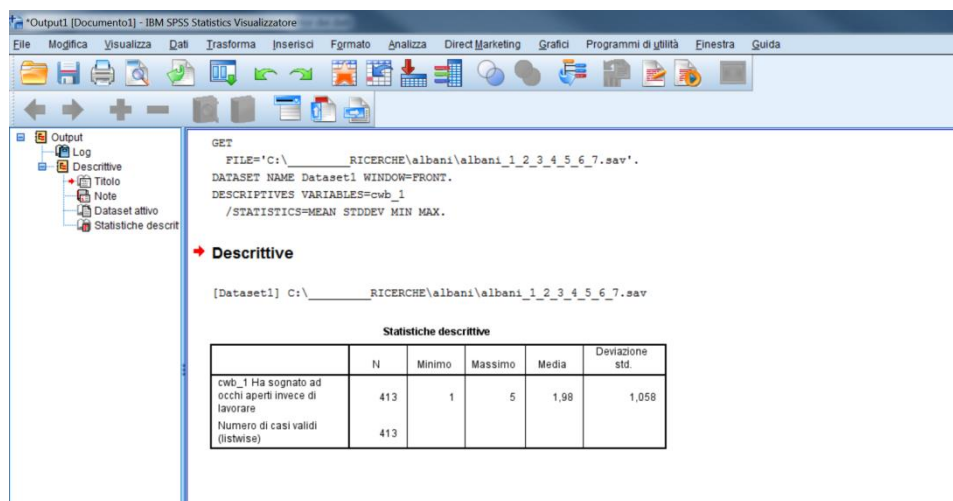
	N	Minimo	Massimo	Media	Deviazione std.
cw_b_1 Ha sognato ad occhi aperti invece di lavorare	413	1	5	1,98	1,058
Numero di casi validi (listwise)	413				

IBM SPSS Statistics Il processore è pronto | Cast: 100 | Unicode:ON

# SPSS

## La finestra Visualizzatore

La finestra visualizzatore è suddivisa in due parti:



- Il quadro di sinistra fornisce una visione d'insieme dei contenuti dell'output.

-Il quadro di destra contiene gli elementi veri e propri dell'output (tabelle, grafici e testo).

La maggior parte delle caratteristiche dell'output (es. il colore o l'ampiezza del testo) sono modificabili

# SPSS

## La finestra Sintassi

**I comandi SPSS possono essere eseguiti utilizzando il linguaggio di programmazione di SPSS in un file di sintassi**



**Un file "sintassi" è un file di testo che contiene dei comandi.**

**I comandi scritti nel linguaggio di programmazione di SPSS possono essere salvati in modo tale da rendere possibile la ripetizione delle analisi in un momento successivo.**

**Una interessante risorsa per file di sintassi è:**  
<http://www.spsstools.net/en/>

# SPSS

## I menu di SPSS

**Ciascuna finestra in SPSS ha la propria barra dei menù, che consente la selezione dei menu appropriati per quel tipo di finestra.**

**I menu ANALIZZA e GRAFICI sono disponibili su tutte le finestre, rendendo più semplice la creazione di nuovi output senza dover passare ad altre finestre.**



# SPSS

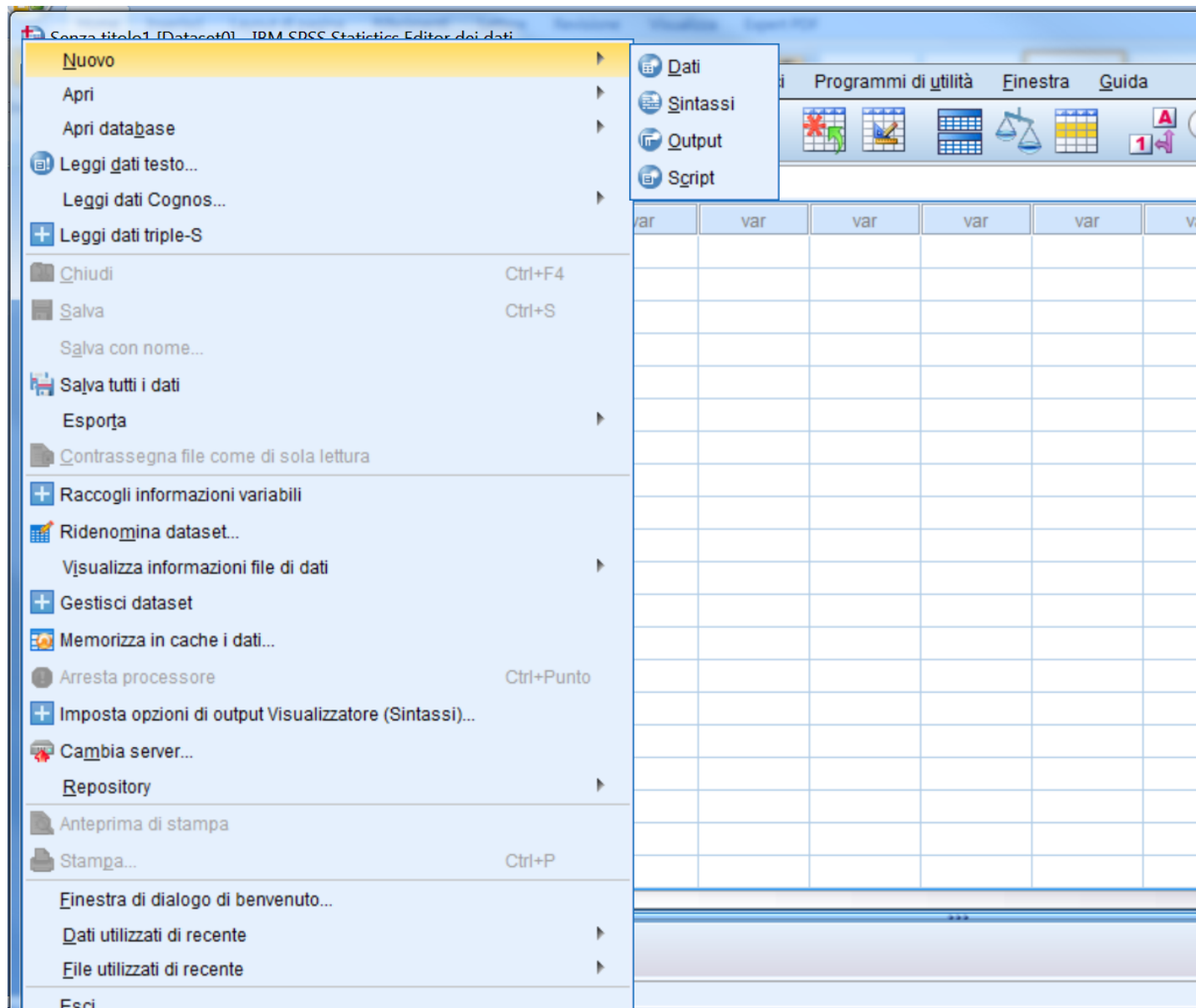
## Il menu file

**Si tratta di un menù particolarmente importante poiché mette in comunicazione il programma SPSS con l'esterno.**

**Il menù File può essere utilizzato per creare un nuovo file scegliendo File/Nuovo. A seconda del tipo di file desiderato è possibile scegliere tra file di dati (Dati), file testo per i comandi nel linguaggio di programmazione (Sintassi), file che contengono risultati sia in formato SPSS (Output), file che consentono di automatizzare alcune operazioni tramite appositi programmi (Script).**

# SPSS

## Il menu file



# SPSS

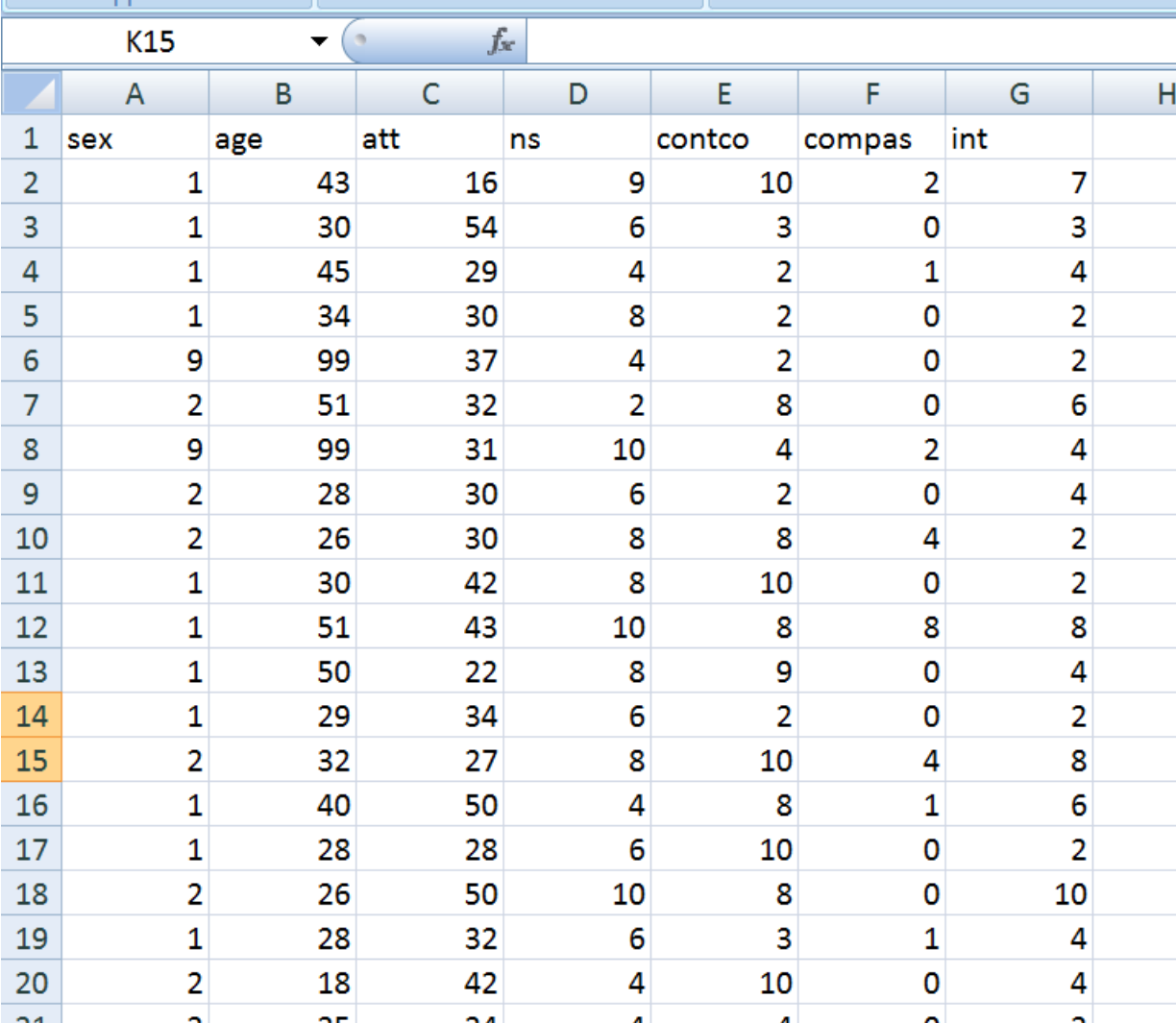
## Il menu file

Tra i diversi tipi di file di dati che è possibile aprire segnaliamo i seguenti, rimandando il lettore al manuale di SPSS e alle funzioni di aiuto in linea per le ulteriori opzioni relative al menù:

- File SPSS (\*.sav), SPSS/PC+ (\*.sys) e Portabile SPSS (\*.por);
- File testo "Tab delimitati", ovvero con i valori separati da tabulazioni, o fissi (\*.dat);
- File di fogli elettronici come Excel (\*.xls, xlsx), o Lotus (\*.wk3, \*.wk1, \*.wks).
- File SYLK - Symbolic Link per fogli elettronici di Microsoft Excel e Multiplan (\*.slk).
- File dBASE IV, III o II (\*.dbf)
- File SAS (\*.sd2, \*.ssd01, \*.ssd04, \*.sd7, \*.sas7bdat, \*.ssd01, \*.xpt).



## Aprire un file di dati in formato excel

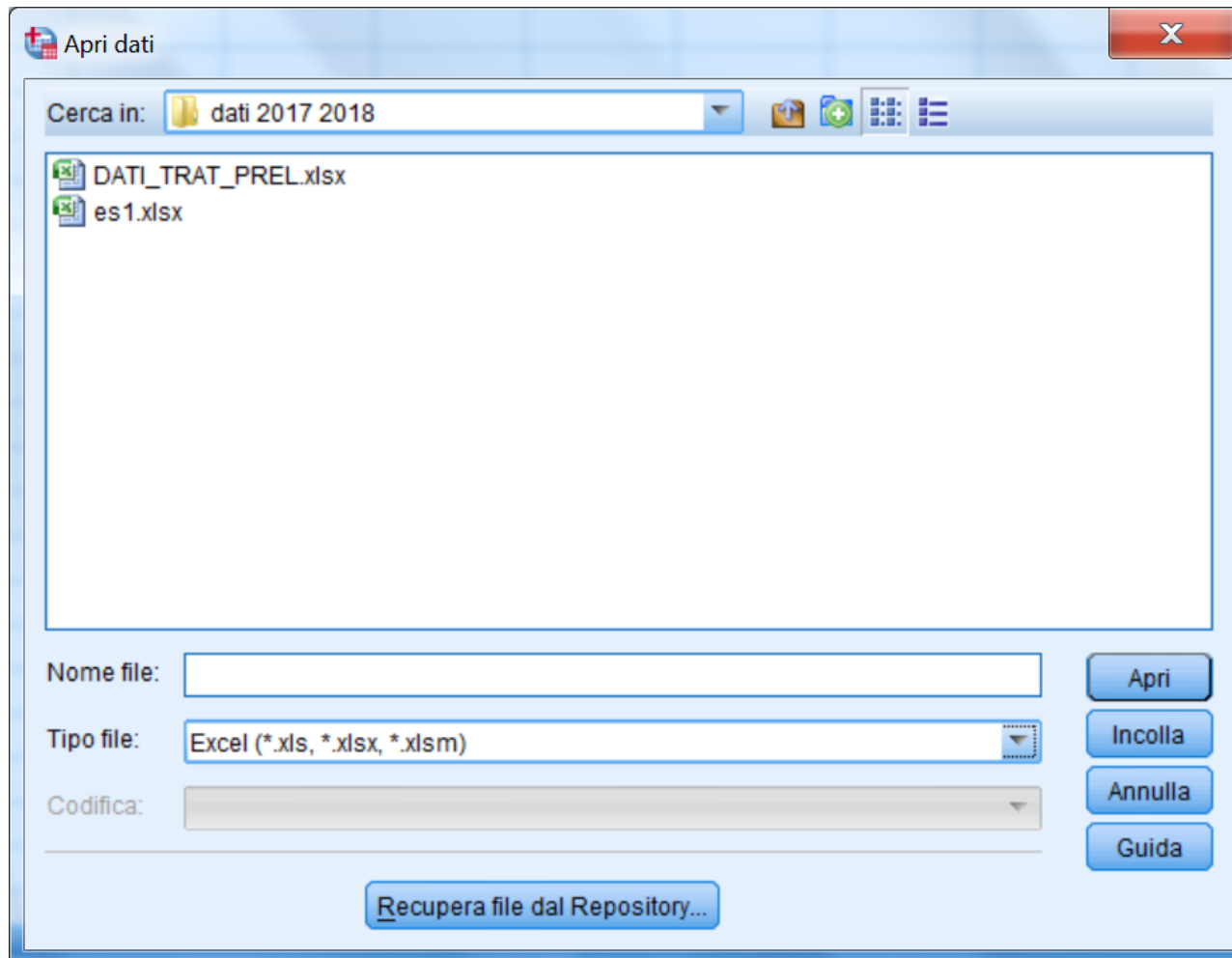


	A	B	C	D	E	F	G	H
1	sex	age	att	ns	contco	compas	int	
2	1	43	16	9	10	2	7	
3	1	30	54	6	3	0	3	
4	1	45	29	4	2	1	4	
5	1	34	30	8	2	0	2	
6	9	99	37	4	2	0	2	
7	2	51	32	2	8	0	6	
8	9	99	31	10	4	2	4	
9	2	28	30	6	2	0	4	
10	2	26	30	8	8	4	2	
11	1	30	42	8	10	0	2	
12	1	51	43	10	8	8	8	
13	1	50	22	8	9	0	4	
14	1	29	34	6	2	0	2	
15	2	32	27	8	10	4	8	
16	1	40	50	4	8	1	6	
17	1	28	28	6	10	0	2	
18	2	26	50	10	8	0	10	
19	1	28	32	6	3	1	4	
20	2	18	42	4	10	0	4	
21	2	25	24	4	4	0	2	

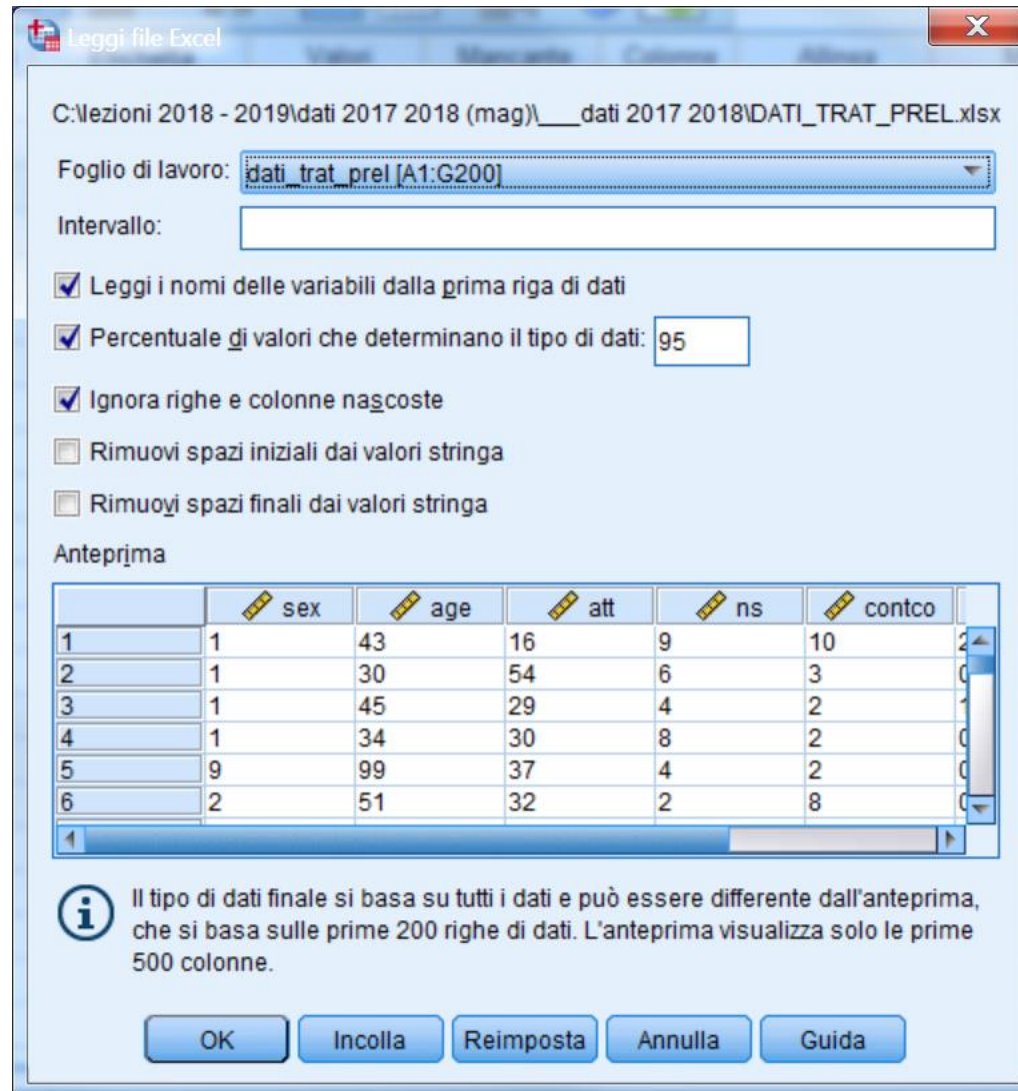
**DATI\_TRAT\_PREL.xlsx**



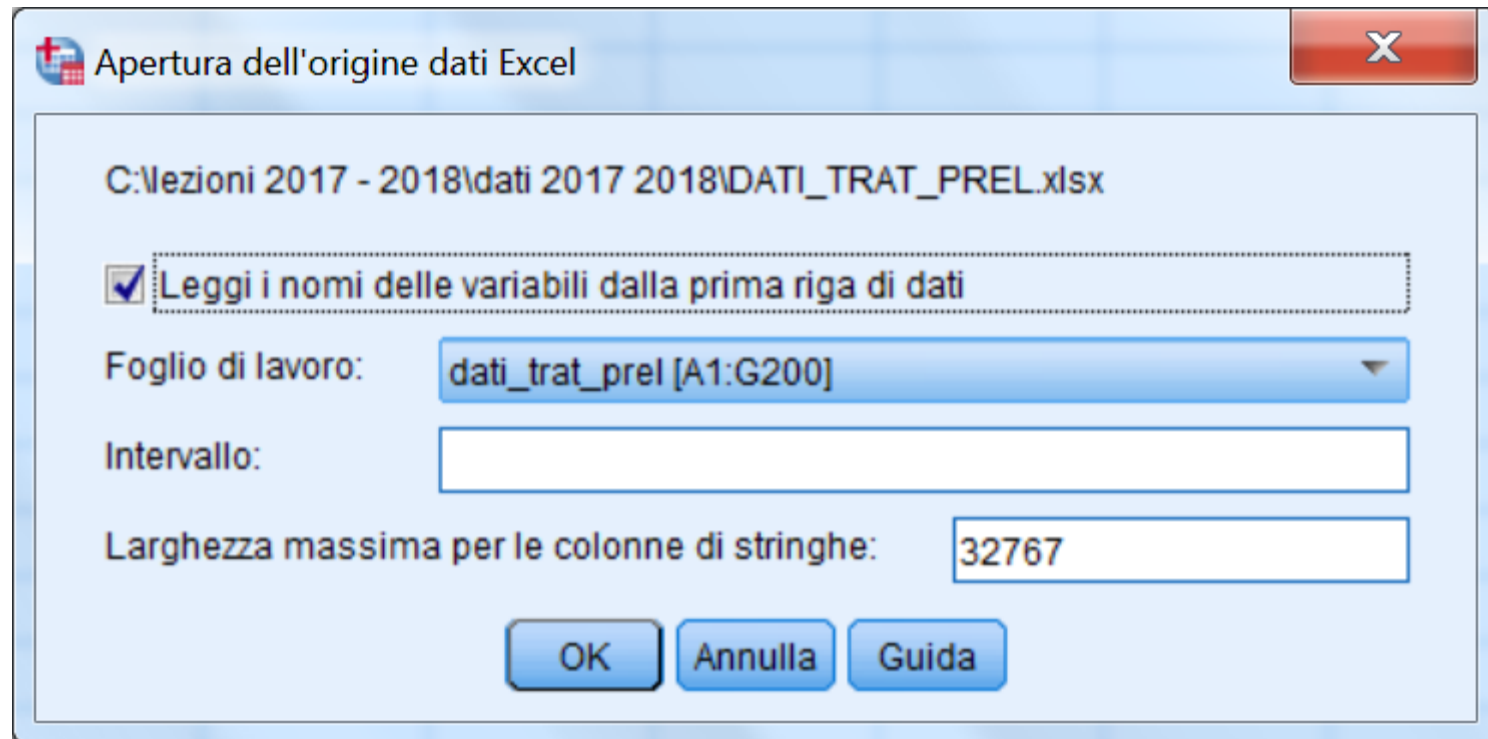
# Aprire i dati



# Leggere i dati (spss 25)



# Leggere i dati (versioni precedenti di spss)



# Aprire i dati

\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Guida

	sex	age	att	ns	contco	compas	int	v
1	1	43	16	9	10	2	7	
2	1	30	54	6	3	0	3	
3	1	45	29	4	2	1	4	
4	1	34	30	8	2	0	2	
5	9	99	37	4	2	0	2	
6	2	51	32	2	8	0	6	
7	9	99	31	10	4	2	4	
8	2	28	30	6	2	0	4	
9	2	26	30	8	8	4	2	
10	1	30	42	8	10	0	2	
11	1	51	43	10	8	8	8	
12	1	50	22	8	9	0	4	
13	1	29	34	6	2	0	2	
14	2	32	27	8	10	4	8	
15	1	40	50	4	8	1	6	
16	1	28	28	6	10	0	2	
17	2	26	50	10	8	0	10	
18	1	28	32	6	3	1	4	
19	2	18	42	4	10	0	4	
20	2	25	24	4	4	0	2	
21	2	33	50	10	10	0	10	

Vista dati Vista Variabile

# SPSS

## Il menu file

### *Salvare ed esportare un file dati*

**Per salvare un file di dati scegliere dal menu: File  
⇒Salva oppure File ⇒Salva con nome.**

**Nel secondo caso si aprirà una finestra di dialogo analoga a quella relativa all'apertura dei file che consente di specificare il percorso per il file da salvare, e di definire il tipo di file che viene salvato.**

**I formati di file definibili sono quelli esaminati nella slide relativa all'apertura dei file.**

# Salvare i dati

\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct Marketing Grafici Programmi d

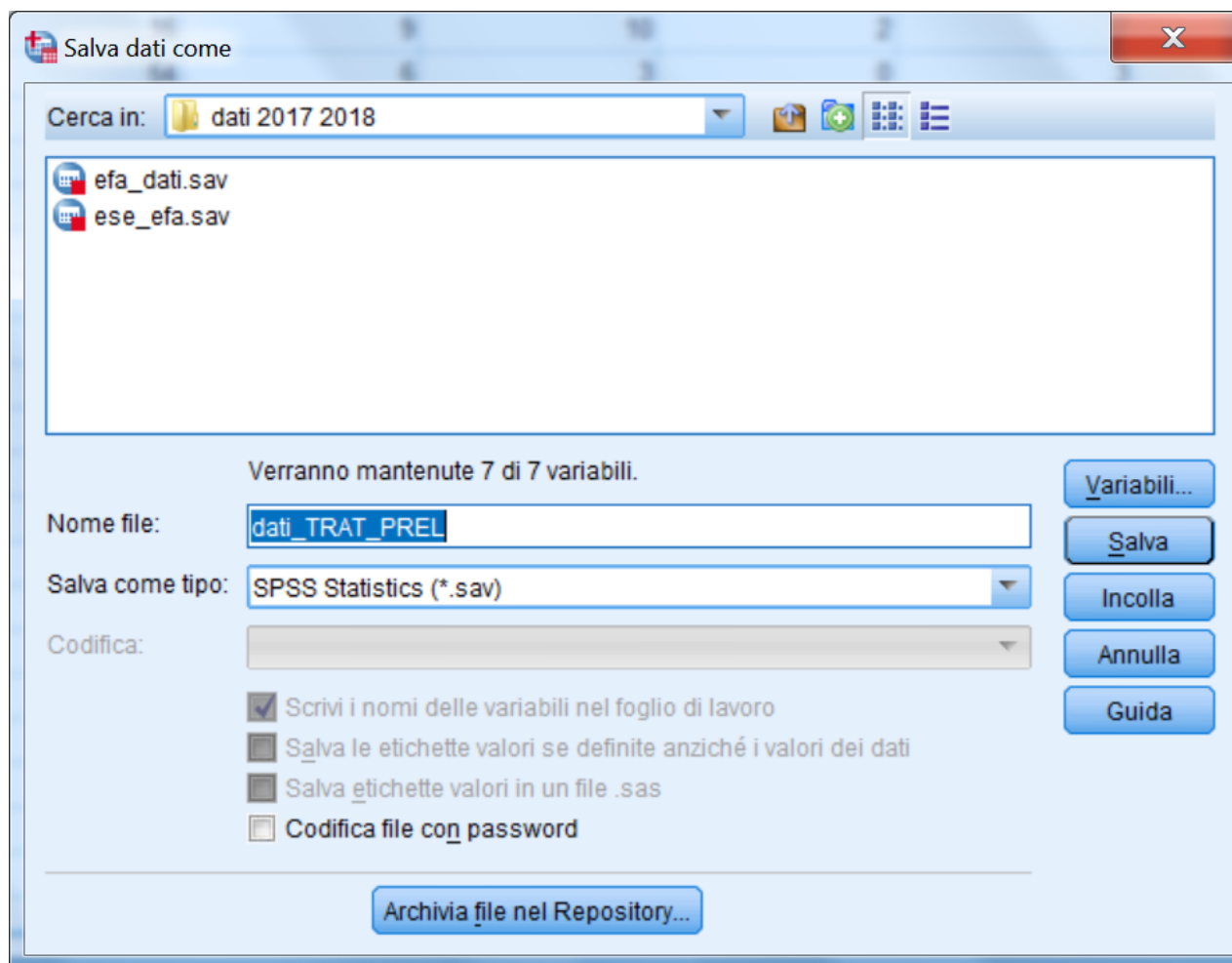
Nuovo ▶  
 Apri ▶  
 Apri database ▶  
 Leggi dati testo...  
 Leggi dati Cognos... ▶  
 Leggi dati triple-S  
 Chiudi Ctrl+F4  
 Salva Ctrl+S  
 Salva con nome...  
 Salva tutti i dati  
 Esporta ▶  
 Contrassegna file come di sola lettura  
 Raccogli informazioni variabili  
 Ridenomina dataset...  
 Visualizza informazioni file di dati ▶  
 Gestisci dataset  
 Memorizza in cache i dati...  
 Arresta processore Ctrl+Punto  
 Imposta opzioni di output Visualizzatore (Sintassi)...  
 Cambia server...  
 Repository ▶  
 Anteprima di stampa

	compas	
10	2	
3	0	
2	1	
2	0	
2	0	
8	0	
4	2	
2	0	
8	4	
10	0	
8	8	
9	0	
2	0	
10	4	
8	1	
10	0	
8	0	
3	1	
10	0	
4	0	



## Salvare i dati

**Salviamo il nostro file importato da excel per usarlo come file .sav nei prossimi esempi (altrimenti alla chiusura del programma andrebbe perso).**



# **SPSS**

## **Il menu file**

**E' possibile escludere variabili dal file che viene salvato cliccando sul pulsante "Variabili" e scegliendo quali variabili eliminare.**

**Nella figura successiva viene mostrata la finestra di dialogo che consente di filtrare le variabili: se viene lasciata l'opzione di default tutte le variabili vengono mantenute nel file che viene salvato.**

**Per eliminare una variabile dal nuovo file è sufficiente effettuare un clic del mouse sul quadrato corrispondente alla variabile nella colonna "Mantieni".**

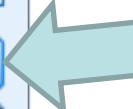
# SPSS

## Il menu file

Verranno mantenute 7 di 7 variabili.

Nome file:

Salva come tipo:



Salva dati come: Variabili

Solo le variabili selezionate verranno salvate sul file di dati specificato.

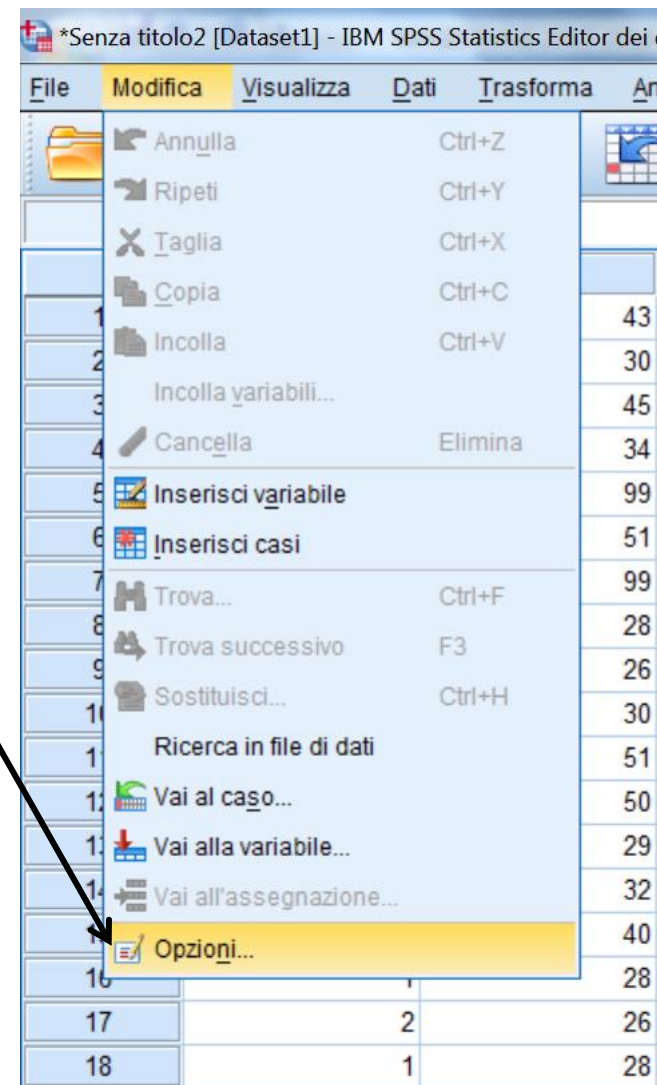
Mantieni	Nome	Etichetta	Ordine
<input checked="" type="checkbox"/>	sex		1
<input checked="" type="checkbox"/>	age		2
<input checked="" type="checkbox"/>	att		3
<input checked="" type="checkbox"/>	ns		4
<input checked="" type="checkbox"/>	contco		5
<input checked="" type="checkbox"/>	compas		6
<input checked="" type="checkbox"/>	int		7

Selezionate: 7 di 7 variabili.

# SPSS

## Il menu modifica

Questo menu consente di copiare, tagliare, incollare e cancellare righe e colonne nell'Editor dei dati, trovare dei valori specifici per una data variabile e definire le **opzioni** di base per il programma (es. definire il tipo di visualizzazione delle variabili negli elenchi e negli output) per le quali si rimanda ai manuali specifici e alle funzioni di aiuto in linea.



# SPSS

## Il menu visualizza

**Questo menù definisce il modo in cui vengono visualizzate la barra di stato, le barre degli strumenti, le variabili, le griglie della tabella dei dati, e definisce i caratteri utilizzati per visualizzare le diverse informazioni. In particolare:**

**- L'opzione *Barra di stato* consente di mostrare o nascondere la barra di stato, ovvero quella zona della parte inferiore di una finestra SPSS nella quale sono visualizzate le informazioni sullo stato di esecuzione dei programmi, sullo stato del filtro e della ponderazione dei casi**

# SPSS

## Il menu visualizza

The screenshot shows the IBM SPSS Statistics Editor interface. The 'Visualizza' menu is open, displaying various options for customizing the data view. The background shows a data table with 10 rows and 4 columns.

1				16
2				54
3				29
4				30
5				37
6				32
7	9	99		31
8	2	28		30
9	2	26		30
10	1	30		42

The 'Visualizza' menu options are:

- Barra di stato
- Barre degli strumenti
- Editor del menu...
- Caratteri...
- Linee della griglia
- Etichette valori
- Contrassegna dati assegnati
- Personalizza vista Variabile...
- Variabili Ctrl+T

1

Vista dati Vista Variabile

IBM SPSS Statistics Il processore è pronto Unicode:ON

# SPSS

## Vista dati



\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

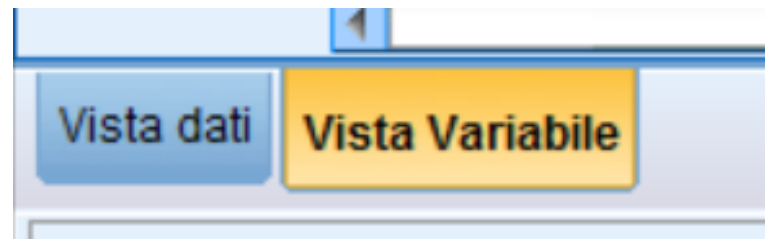
File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi

5:

	sex	age	att	ns	c
1	1	43	16	9	
2	1	30	54	6	
3	1	45	29	4	
4	1	34	30	8	
5	9	99	37	4	
6	2	51	32	2	
7	9	99	31	10	
8	2	28	30	6	
9	2	26	30	8	
10	1	30	42	8	

# SPSS

## Vista Variabile



\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Guida

	Nome	Tipo	Larghezza	Decimali	Etichetta	Valori	Mancante/i	Colonne	Allinea	Misura	Ruolo
1	sex	Numerico	12	0		Nessuno	Nessuno	12	Destra	Nominale	Input
2	age	Numerico	12	0		Nessuno	Nessuno	12	Destra	Scala	Input
3	att	Numerico	12	0		Nessuno	Nessuno	12	Destra	Scala	Input
4	ns	Numerico	12	0		Nessuno	Nessuno	12	Destra	Nominale	Input
5	contco	Numerico	12	0		Nessuno	Nessuno	12	Destra	Nominale	Input
6	compas	Numerico	12	0		Nessuno	Nessuno	12	Destra	Nominale	Input
7	int	Numerico	12	0		Nessuno	Nessuno	12	Destra	Nominale	Input
8											
9											



# SPSS

## Vista Variabile

\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Gu

	Nome	Tipo	Larghezza	Decimali	Etichetta	Valori	Mancante/i	Colonne
1	sex	Numerico	12	0		Nessuno	Nessuno ...	12
2	age	Numerico	12	0		Nessuno	Nessuno	12
3	att	Numerico	12	0		Nessuno	Nessuno	12
4							Nessuno	12
5							Nessuno	12

Valori mancanti

Nessun valore mancante

Valori mancanti discreti

9

Intervallo più un valore mancante discreto facoltativo

Basso:  Alto:

Valore discreto:

OK Annulla Guida

# SPSS

## Vista Variabile

\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Guida

	Nome	Tipo	Larghezza	Decimali	Etichetta	Valori	Mancante/i	Colonne	Allinea	Misura
1	sex	Numerico	12	0		Nessuno	Nessuno	12	Destra	Nominale
2	age	Numerico	12	0		Nessuno	Nessuno	12	Destra	Scala
3	att	Numerico	12	0		Nessuno	Nessuno	12	Destra	Scala
4	ns	Numerico	12	0						
5	contco	Numerico	12	0						
6	compas	Numerico	12	0						
7	int	Numerico	12	0						
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										

Etichette valori

Etichette valori

Valore:

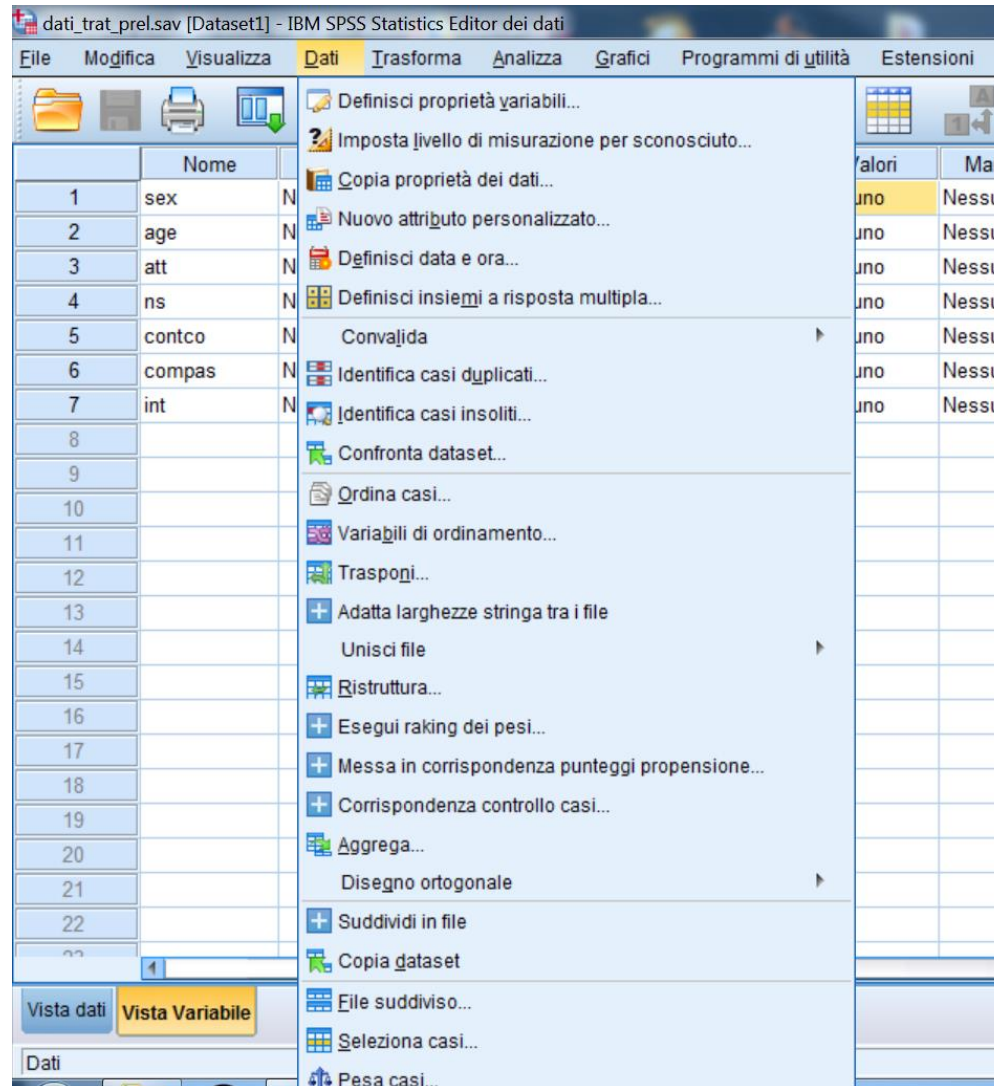
Etichetta:

1 = "MASCHIO"

# SPSS

## Il menu dati

Effettua operazioni sulle variabili e sui casi.



The screenshot shows the SPSS Statistics Editor interface. The 'Dati' menu is open, displaying various options for data manipulation. On the left, a list of variables is shown with their names and types.

	Nome	
1	sex	N
2	age	N
3	att	N
4	ns	N
5	contco	N
6	compas	N
7	int	N
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		

The 'Dati' menu options include:

- Definisci proprietà variabili...
- Imposta livello di misurazione per sconosciuto...
- Copia proprietà dei dati...
- Nuovo attributo personalizzato...
- Definisci data e ora...
- Definisci insiemi a risposta multipla...
- Convalida
- Identifica casi duplicati...
- Identifica casi insoliti...
- Confronta dataset...
- Ordina casi...
- Variabili di ordinamento...
- Trasponi...
- Adatta larghezze stringa tra i file
- Unisci file
- Ristruttura...
- Esegui raking dei pesi...
- Messa in corrispondenza punteggi propensione...
- Corrispondenza controllo casi...
- Aggrega...
- Disegno ortogonale
- Suddividi in file
- Copia dataset
- File suddiviso...
- Seleziona casi...
- Pesa casi...

# SPSS

## Il menu dati

I sotto-menù più utili sono:

**“Copia proprietà dei dati”** consente all’utente di prendere un file dati SPSS esterno ed utilizzarlo come modello per la definizione del file dati corrente. In particolare, sia le proprietà del file (es., etichetta del file, insieme a risposta multipla, ecc.), sia quelle delle variabili (es., etichette dei valori, valori mancanti, etichette delle variabili, ecc.) del file “modello” possono essere utilizzate per definire quelle del file corrente



# SPSS

## Il menu dati

I sotto-menù più utili sono:

“*Unisci file*” consente di unire due file in un unico file e presenta due diverse modalità fondamentali:

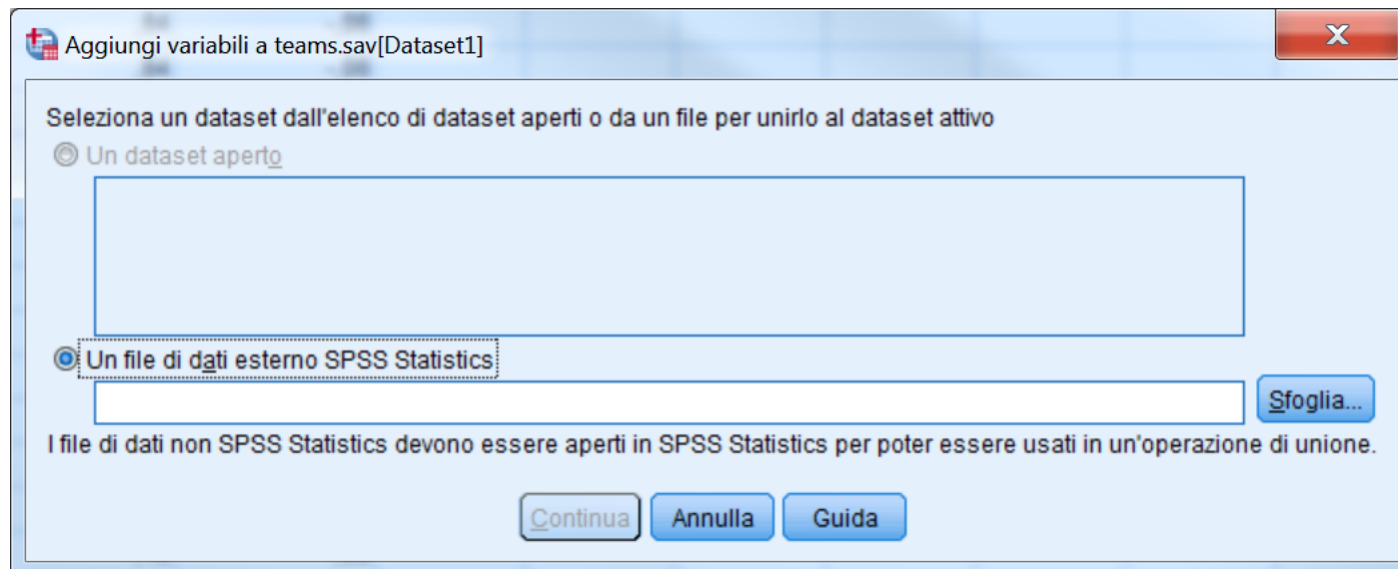
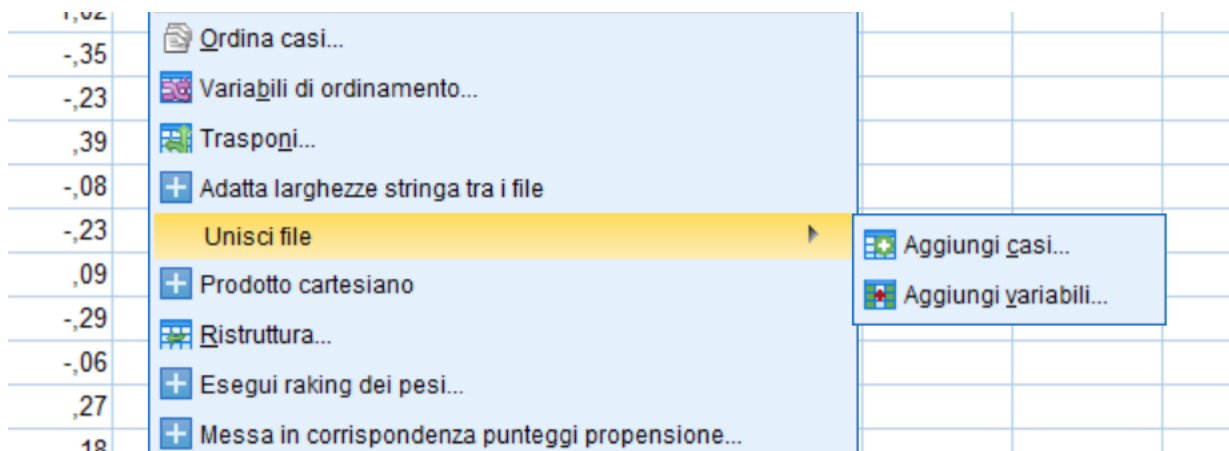
***Aggiungi casi e Aggiungi variabili.***

***Aggiungi casi*** unisce il file attivo con un secondo file che contiene le stesse variabili ma casi differenti.

***Aggiungi variabili*** unisce il file attivo con un file dati esterno che contiene gli stessi casi ma variabili differenti da quelle nel file attivo. I casi devono avere *lo stesso ordine* in entrambi i file. Se si utilizza una “variabile chiave” per appaiare i casi, i due file devono *essere ordinati* in modo crescente rispetto alla variabile chiave.

# SPSS

## Il menu dati



# SPSS

## Il menu dati

I sotto-menù più utili sono:

“***Selezione Casi***” consente di definire sottoinsiemi di casi che vengono selezionati tramite un criterio specificato dall'utente stesso. Per la selezione dei casi l'utente può specificare un'operazione di natura più o meno complessa, oppure avvalersi del generatore di numeri casuali di SPSS. I casi non selezionati possono essere *filtrati* o *cancellati* del tutto dal file. La modalità che prevede che i casi siano filtrati crea una nuova variabile, “filter\_\$”, che serve per indicare lo stato attuale del filtro. Il valore di tale variabile è uguale a 1 per i casi che soddisfano la condizione di selezione, mentre è uguale a 0 per i casi che non soddisfano tale condizione e che quindi vengono esclusi dall'analisi.



# SPSS

## Il menu dati

The image shows two overlapping dialog boxes in SPSS. The background dialog is 'Seleziona casi' (Select Cases), and the foreground dialog is 'Seleziona casi: Se...' (Select Cases: If...).

**Seleziona casi (Background Dialog):**

- Seleziona:**
  - Tutti i casi
  - Se la condizione è soddisfatta
  - Campione random di...
  - Basato su intervallo di...
  - Usa variabile filtro:
- Output:**
  - Escludi casi non selezionati
  - Copia casi selezionati
  - Elimina casi non selezionati

**Seleziona casi: Se... (Foreground Dialog):**

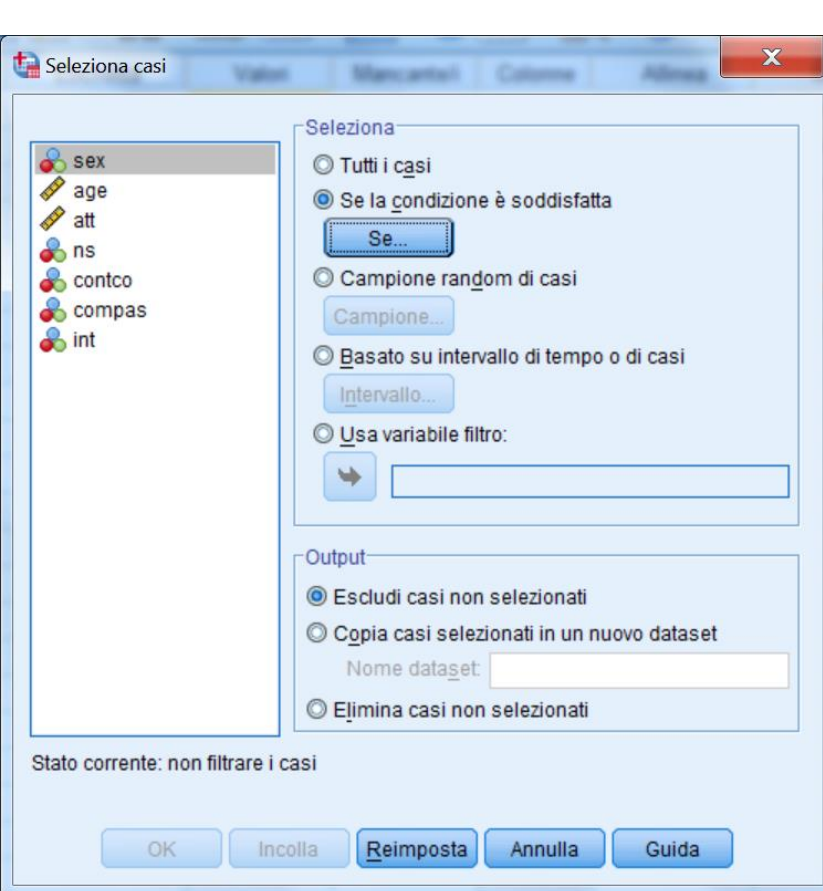
- Condition:** sex = 1
- Group of functions:** Tutto, Aritmetico, CDF e CDF noncentrale, Conversione, Data/Ora corrente, Aritmetica data, Creazione data
- Functions and special variables:** (Empty list)

Buttons: OK, Incolla, Reimposta, Continua, Annulla, Guida



# SPSS

## Il menu dati



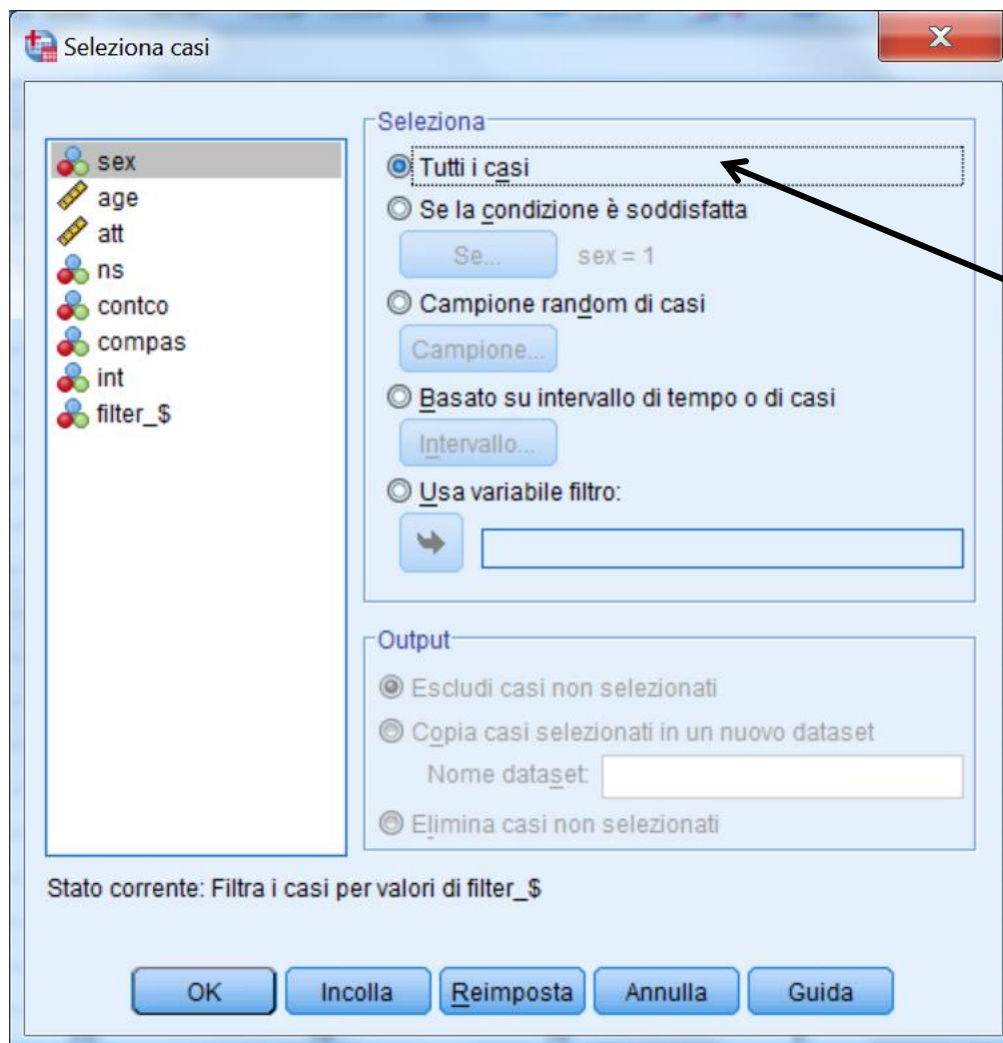
\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Pr

	sex	age	att	ns
1	1	43	16	9
2	1	30	54	6
3	1	45	29	4
4	1	34	30	8
<del>5</del>	9	99	37	4
<del>6</del>	2	51	32	2
<del>7</del>	9	99	31	10
<del>8</del>	2	28	30	6
<del>9</del>	2	26	30	8
10	1	30	42	8
11	1	51	43	10
12	1	50	22	8
13	1	29	34	6
<del>14</del>	2	32	27	8
15	1	40	50	4
16	1	28	28	6
<del>17</del>	2	26	50	10
18	1	28	32	6
<del>19</del>	2	18	42	4
<del>20</del>	2	25	24	4
<del>21</del>	2	22	50	10

# SPSS

## Il menu dati



**Per togliere il filtro cliccare su "Tutti i casi"**

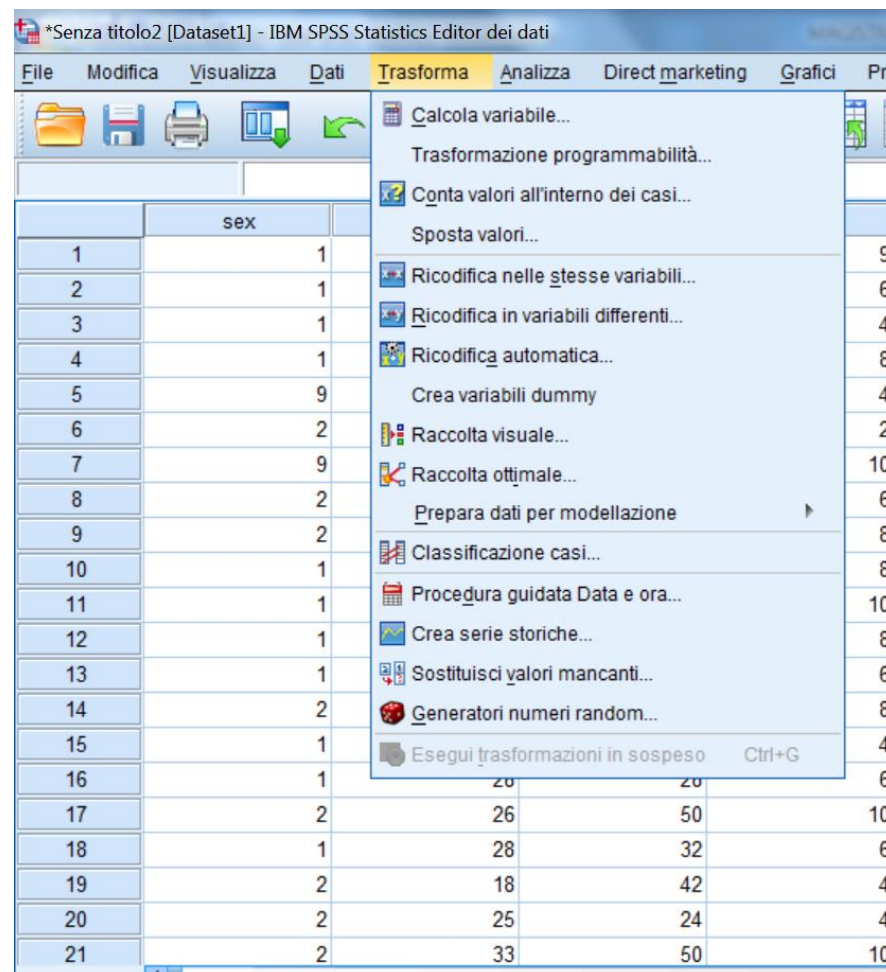
# SPSS

## Il menu Trasforma

Consente di modificare le variabili (o definirne delle nuove) operando trasformazioni su variabili già esistenti

Sono presenti i seguenti comandi:

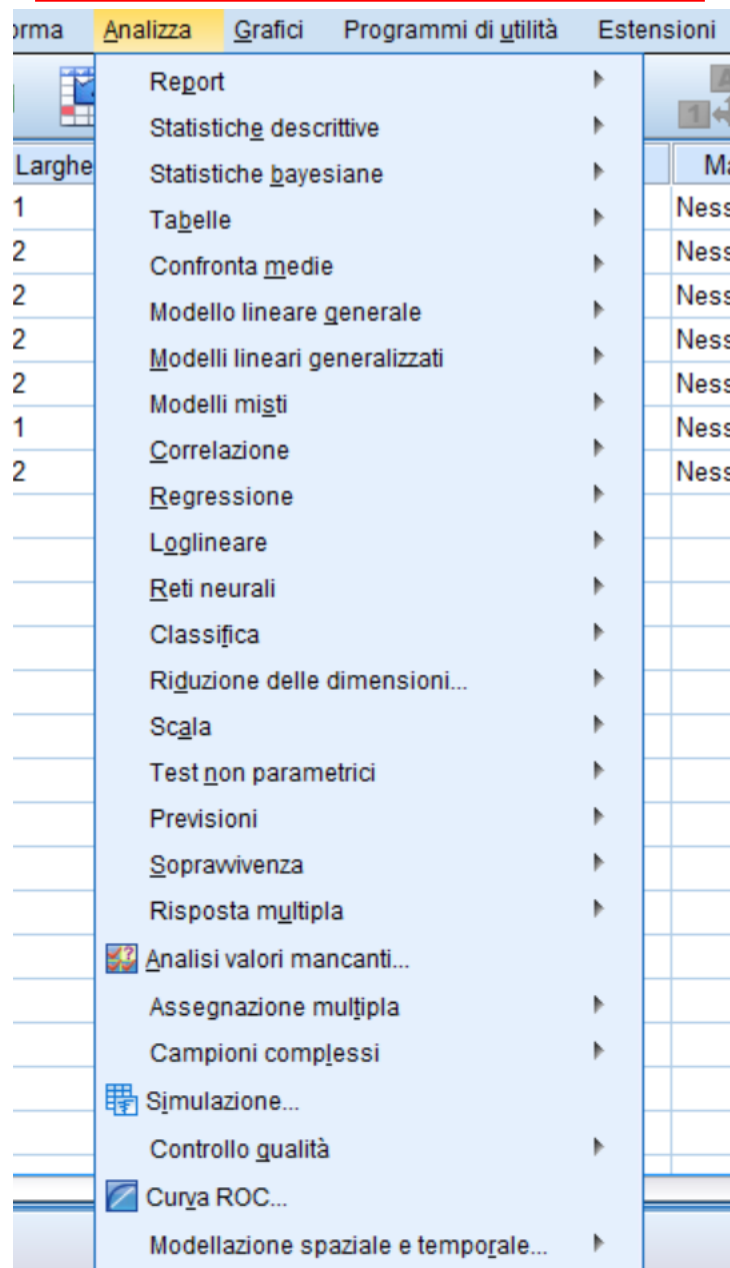
1. **Calcola variabile:** consente di calcolare i valori di una variabile in base alle trasformazioni numeriche di altre variabili.
2. **Ricodifica:** è possibile scegliere tra due opzioni ricodifica nelle stesse variabili e ricodifica in variabili differenti



# SPSS

## Il menu Analizza

È il menu più importante di SPSS, quello che consente di effettuare le analisi statistiche

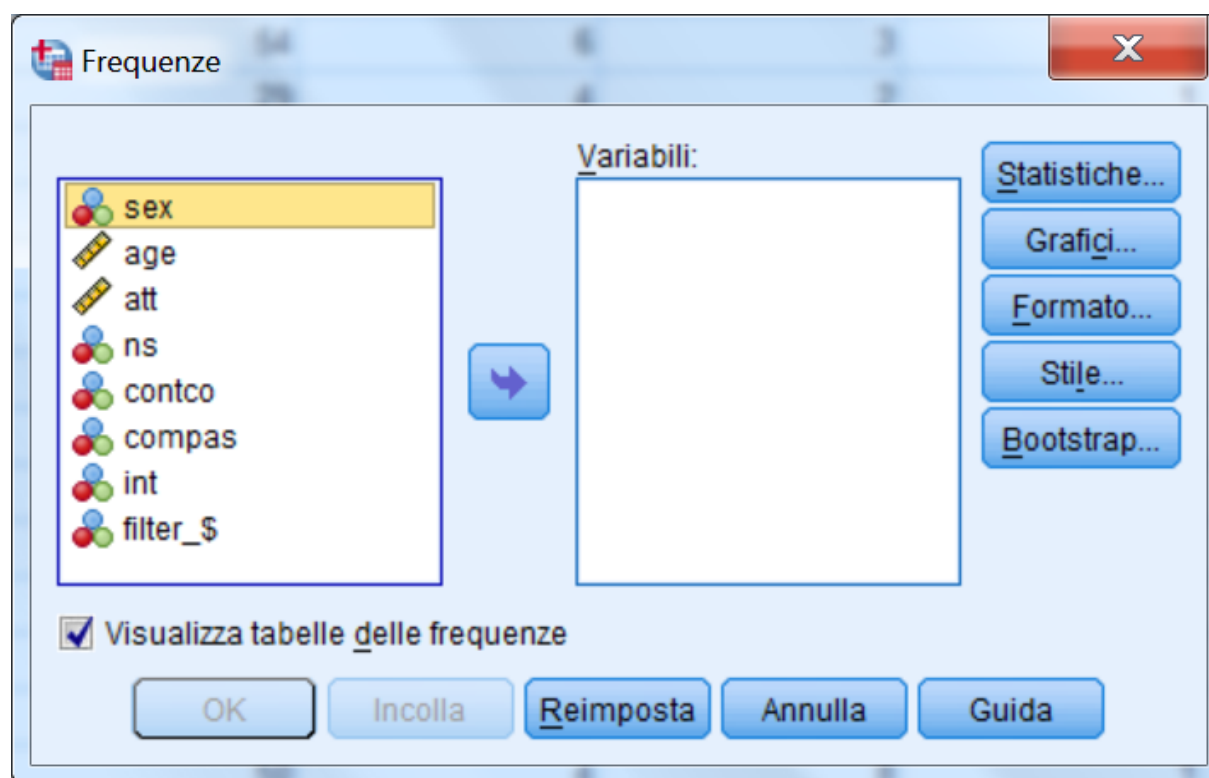


Il menù **ANALIZZA** si trova in tutte le finestre di SPSS

# SPSS

## Le finestre di Dialogo

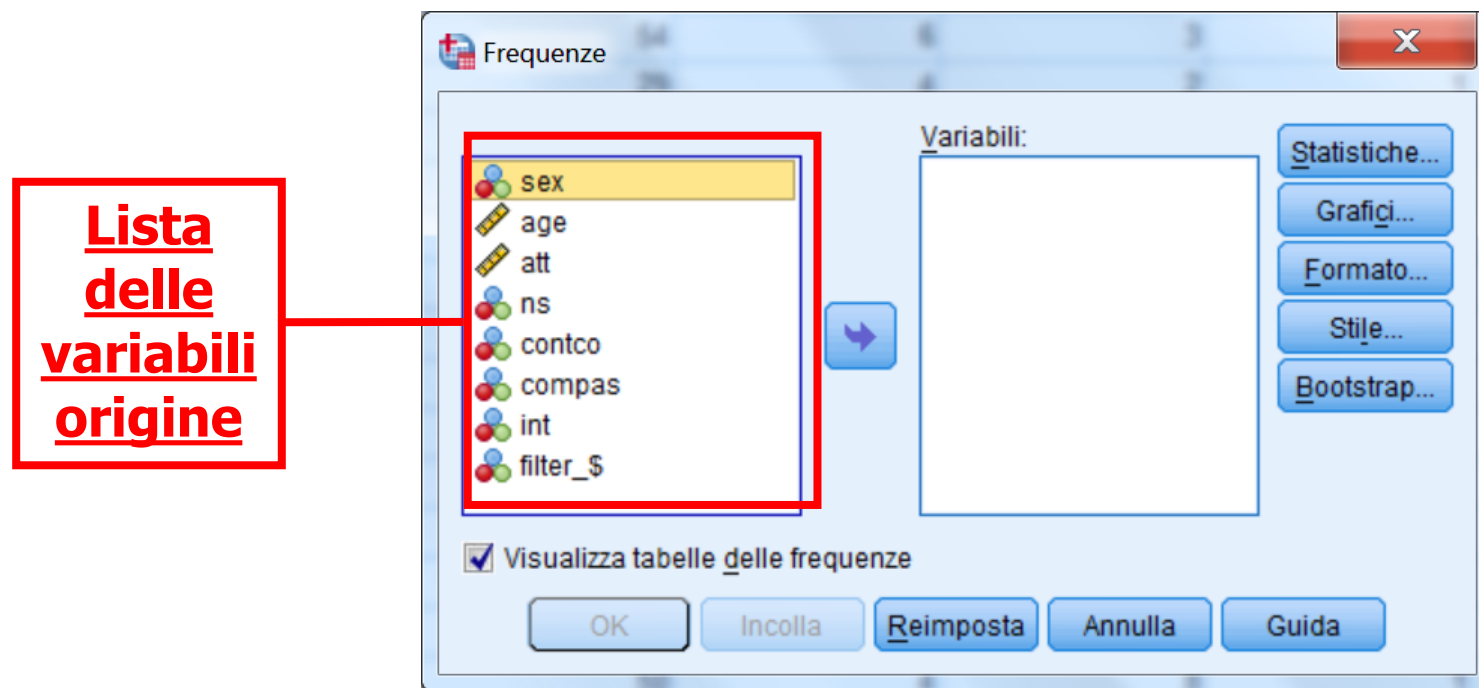
La maggior parte delle opzioni nel menu "Analizza" consentono di aprire delle "finestre di dialogo"



Le finestre di dialogo vengono utilizzate per selezionare le variabili da analizzare (e le diverse opzioni disponibili)

# SPSS

Le finestre di dialogo sono composte da alcuni elementi fondamentali

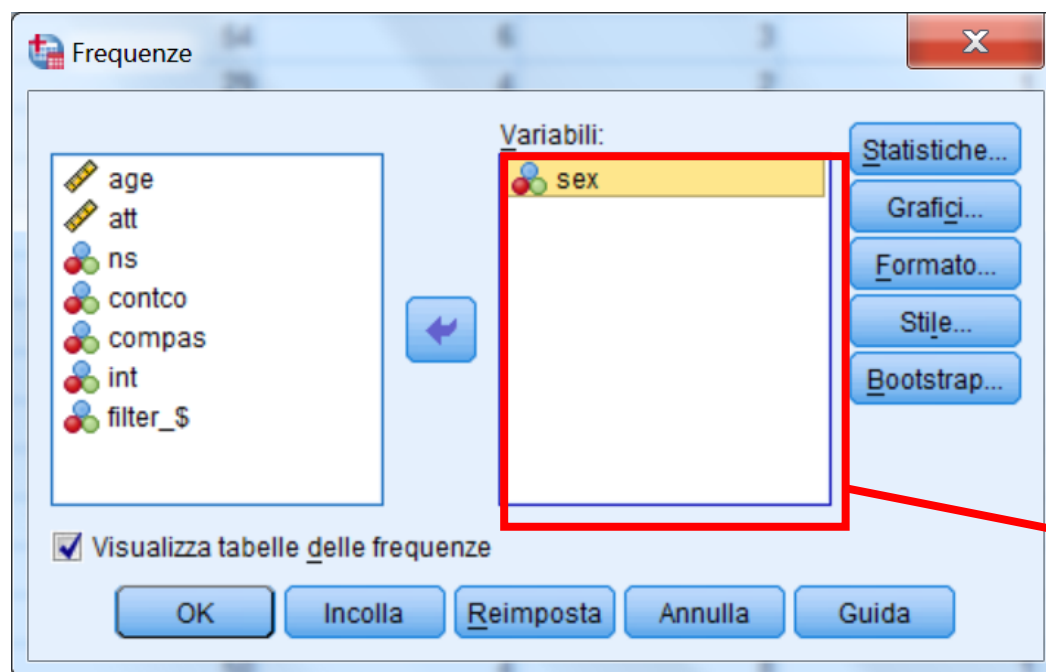


- **Lista delle variabili origine**: è la lista delle variabili contenute nel data file attivo al momento. Non tutte le variabili del file possono comparire in questa lista, ma solo i tipi di variabili consentite dalla procedura selezionata. Ad esempio, una variabile alfanumerica (o "stringa") può apparire soltanto in alcune procedure elementari.

# SPSS

## - Lista delle variabili bersaglio (o variabili attive):

Una o più liste che indicano quali variabili sono state scelte per le analisi. Ad esempio, quali sono le variabili dipendenti e quelle indipendenti



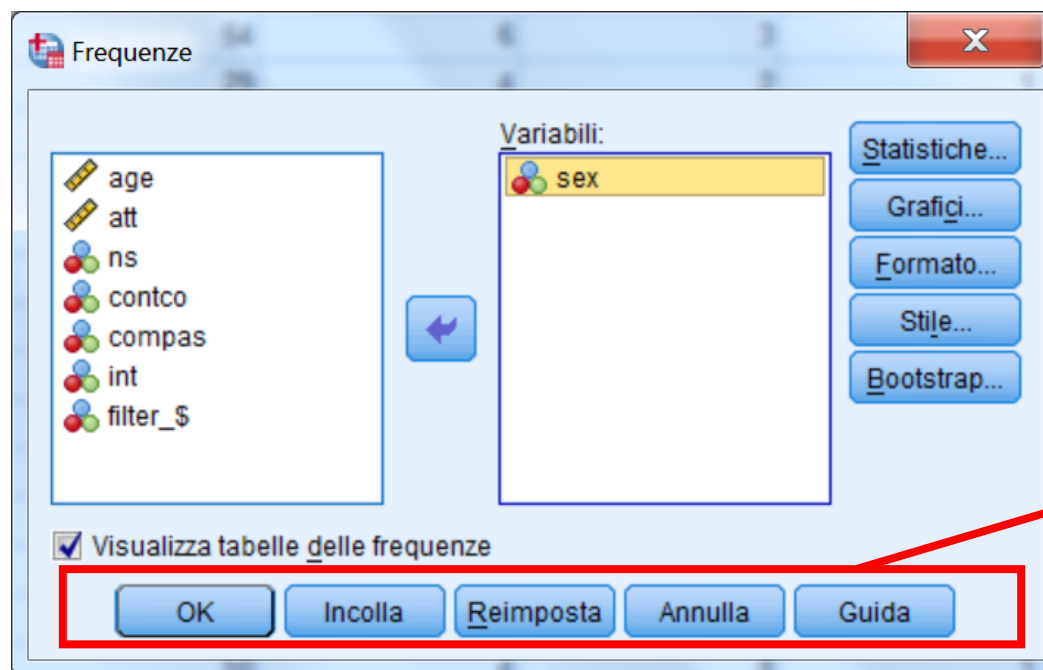
Lista delle  
variabili  
bersaglio

Il comando calcola la distribuzione di frequenza delle variabili incluse nella lista delle variabili bersaglio (a destra)



# SPSS

**-Bottoni dei comandi**: sono i pulsanti che consentono al programma di realizzare un'azione, ad esempio eseguire una procedura di analisi statistica



**I Bottoni dei comandi**

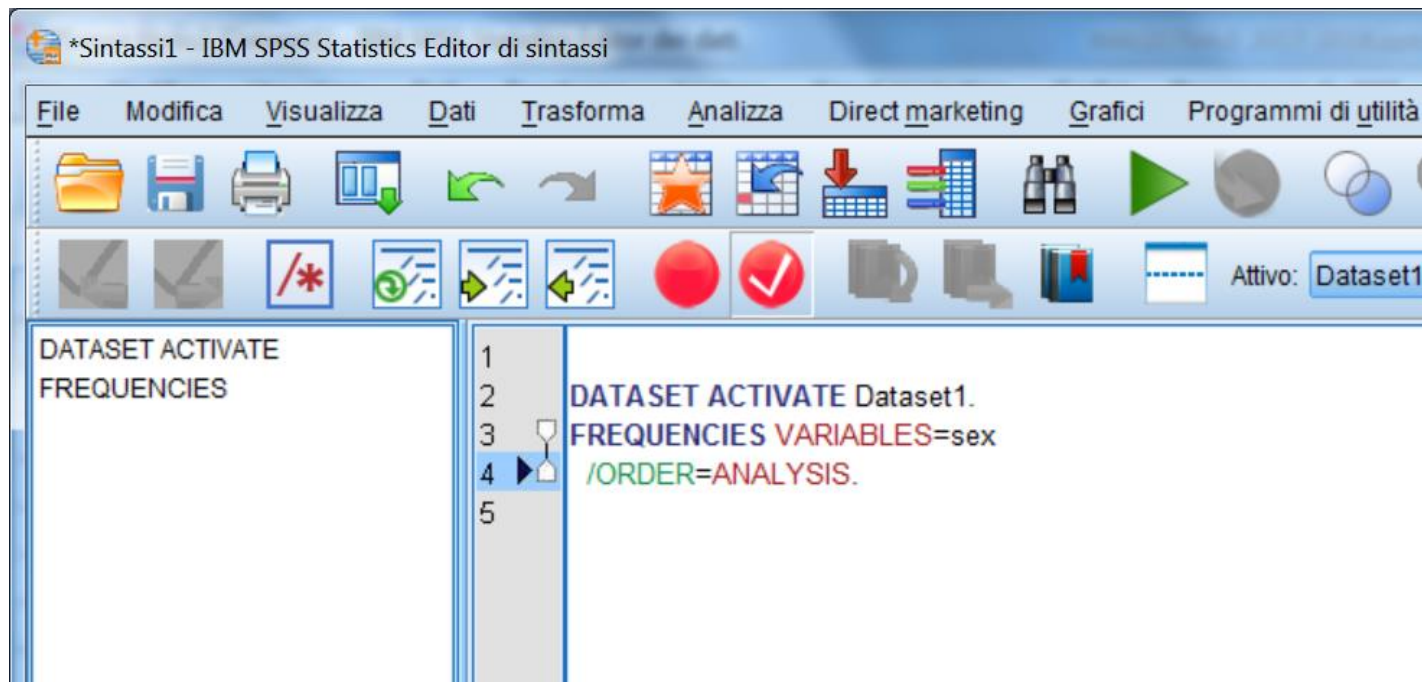
**Premendo sul pulsante OK si eseguono le analisi. REIMPOSTA azzerava tutte le modifiche apportate nella finestra di dialogo. CANCELLA chiude la finestra. AIUTO rappresenta una funzione di aiuto on-line relativa alla finestra di dialogo**



# SPSS

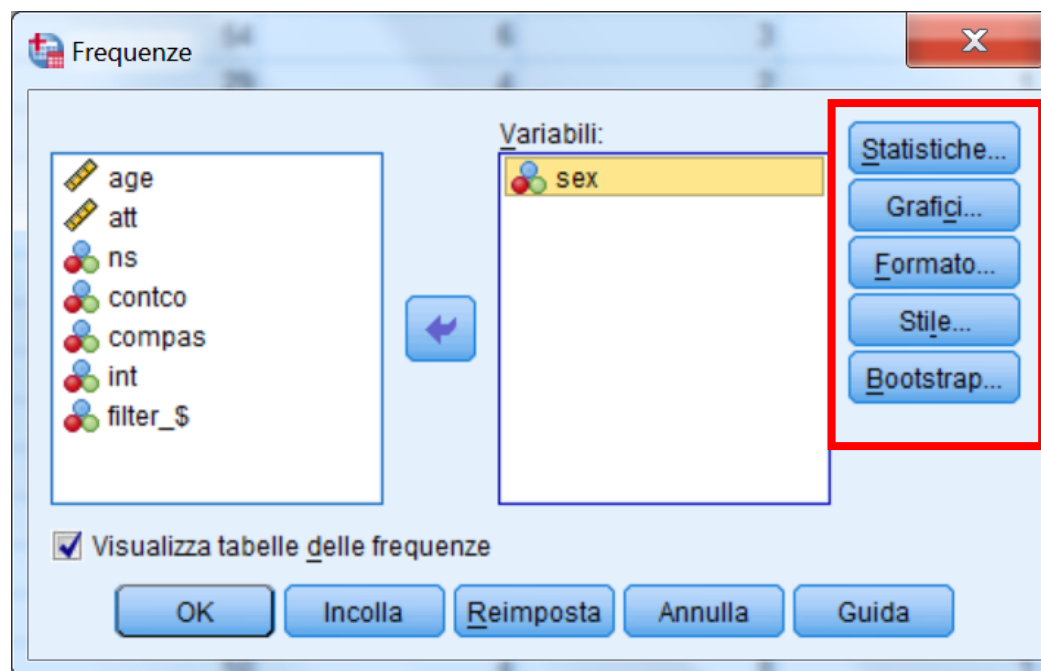
**INCOLLA** consente di “tradurre” i comandi e le opzioni selezionate nella finestra di dialogo nel linguaggio di programmazione di SPSS.

Le linee di sintassi vengono inserite nella finestra Sintassi attiva al momento (se non c'è nessuna finestra Sintassi aperta, ne viene creata una nuova).



# SPSS

I pulsanti posti a destra nella finestra di dialogo consentono di aprire delle ulteriori finestre di dialogo in cui è possibile specificare una serie di opzioni relative alla procedura in corso



Questi pulsanti consentono di aprire ulteriori finestre di dialogo

Questi pulsanti sono diversi per ciascuna finestra di dialogo

# SPSS

## Le Barre degli strumenti

Ogni finestra ha la propria barra degli strumenti, che fornisce un metodo più rapido, grazie all'utilizzo di un unico pulsante, per accedere ad alcuni dei comandi utilizzati più frequentemente



Posizionandosi con il mouse sulle icone, viene fornita una breve descrizione di ciascun comando

# SPSS

## Analisi dei dati con SPSS

**Analisi monovariate:** prendono in esame una sola variabile per volta: indici di tendenza centrale, indici di dispersione (statistiche descrittive -> frequenze e/o descrittive)

**Analisi bivariate:** prendono in esame l'andamento congiunto di due variabili: correlazione (correlazione -> bivariata), regressione (regressione -> lineare), analisi della varianza (modello lineare generalizzato -> univariata)

**Analisi multivariate:** prendono in esame simultaneamente più di due variabili: analisi fattoriale (riduzione dimensione -> fattoriale)

## **Esplorazione dei dati: data screening**

**SPSS consente di calcolare una serie di statistiche che riassumono l'informazione nei dati.**

**L'esplorazione iniziale dei dati è necessaria per esaminare se:**

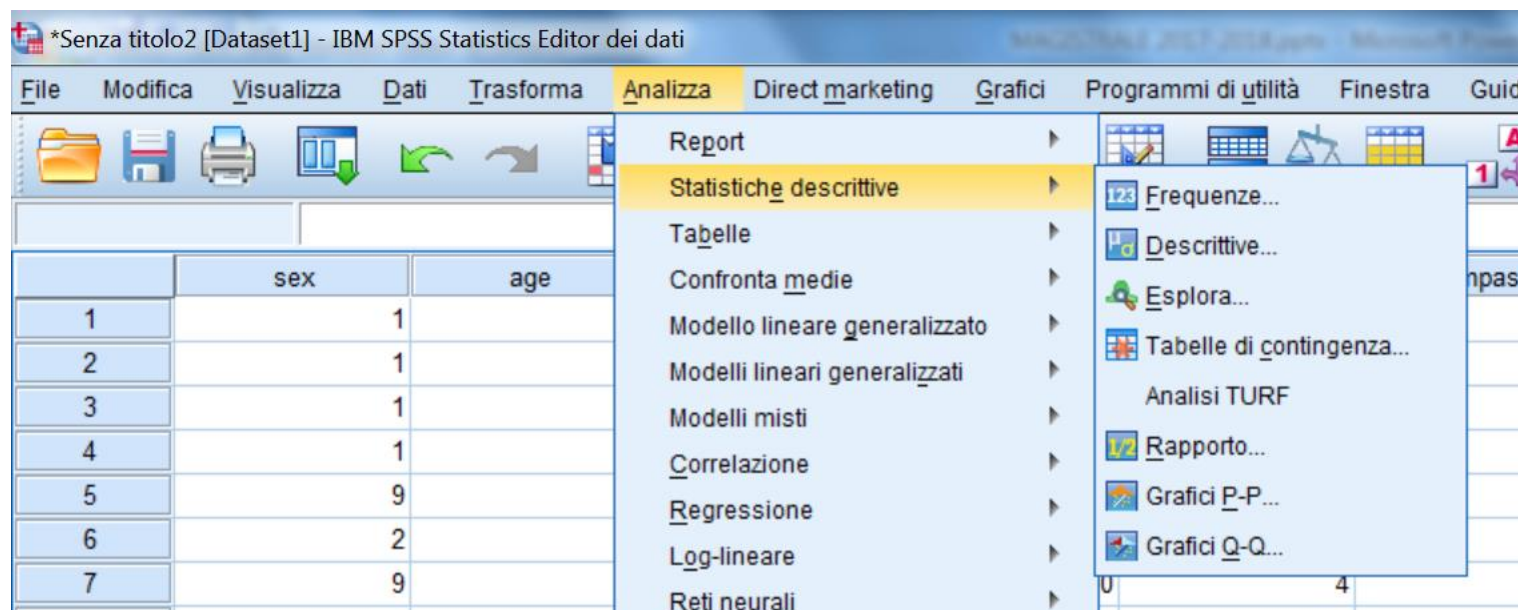
- ci sono errori nei dati, e quindi le variabili assumono valori fuori scala (ad esempio, un item che varia da 1 a 5 ha un punteggio di 8)**
- ci sono "valori anomali" (outliers) ovvero soggetti che presentano valori estremamente elevati in una variabile**
- ci sono casi con valori mancanti**

**L'esplorazione iniziale dei dati è necessaria anche per studiare le caratteristiche distributive delle variabili.**

## Esplorazione dei dati: data screening

**E' possibile esplorare i dati richiedendo:**

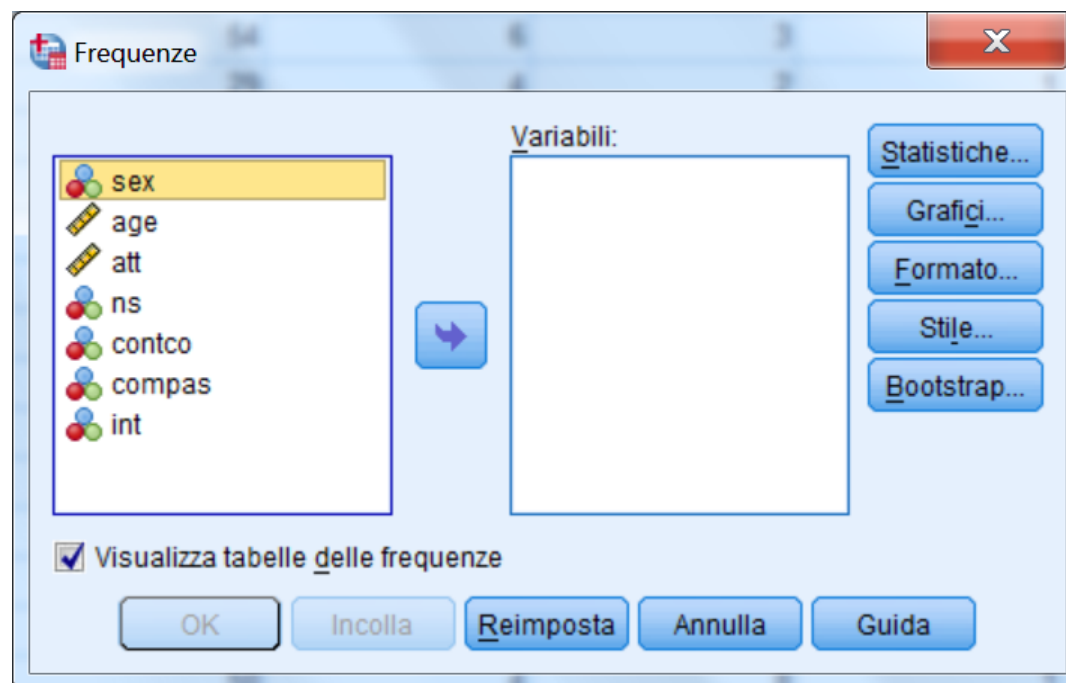
- **distribuzioni di frequenza (procedura Frequenze)**
- **statistiche descrittive come media, deviazione standard, curtosi, asimmetria (procedura Descrittive)**
- **tabelle di esplorazione (procedura Esplora)**
- **tabelle di contingenza**
- **rappresentazioni grafiche**



# SPSS

## La procedura Frequenze

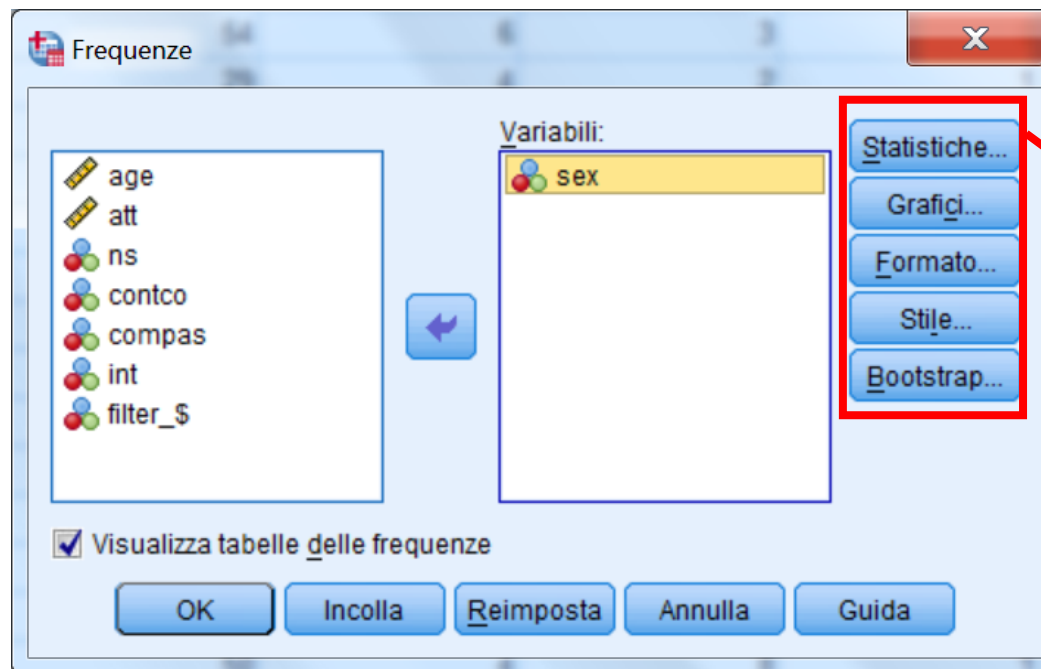
La procedura "Frequenze" consente di effettuare una serie di analisi preliminari, tramite statistiche descrittive e grafici. Selezionando la procedura frequenze si aprirà questa finestra di dialogo:



# SPSS

## La procedura Frequenze

Una volta selezionate la variabili di interesse (es. "Estroversione"), possiamo chiedere diversi tipi di statistiche (tramite il pulsante STATISTICHE) e di grafici (tramite il pulsante GRAFICI).



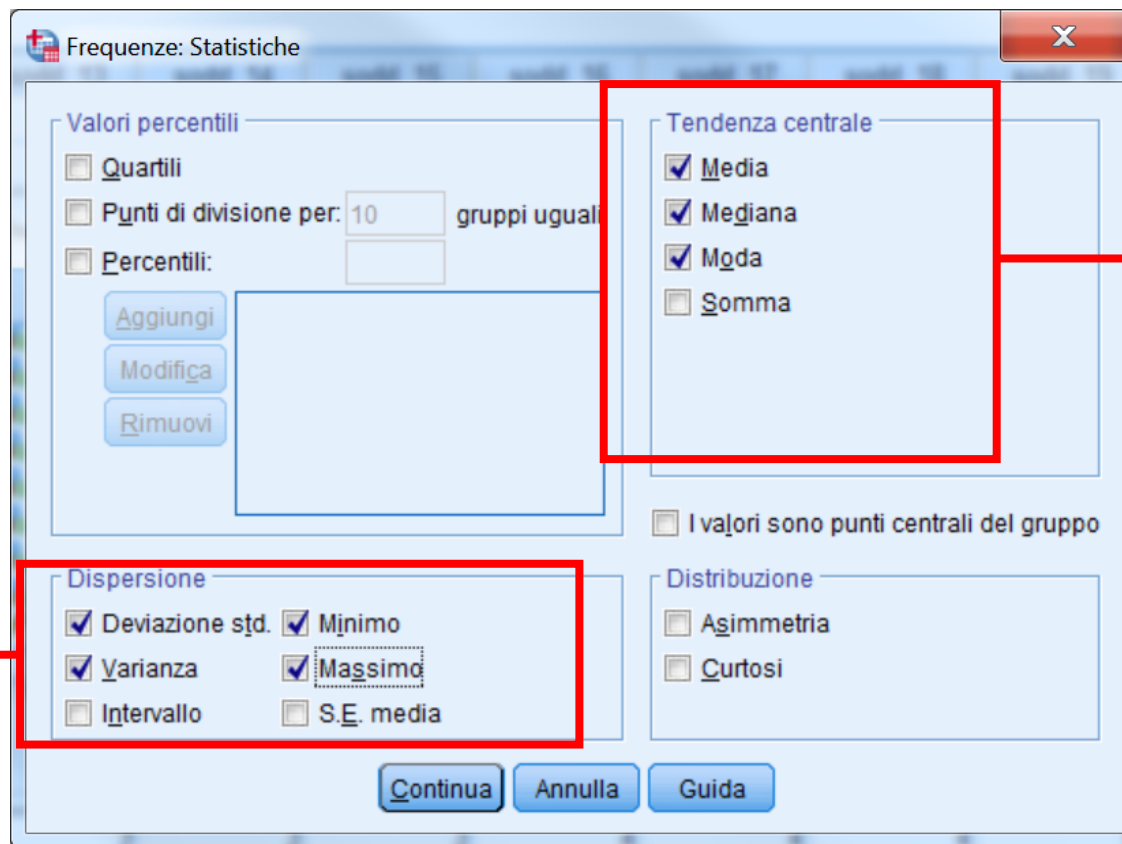
**Il pulsante "Statistiche" consente di richiedere una serie di statistiche descrittive**

**Il pulsante FORMATO (sulla destra) consente di specificare il formato in cui i dati sono presentati nelle tabelle**



# SPSS

Cliccando sul pulsante "Statistiche" si aprirà questa finestra:



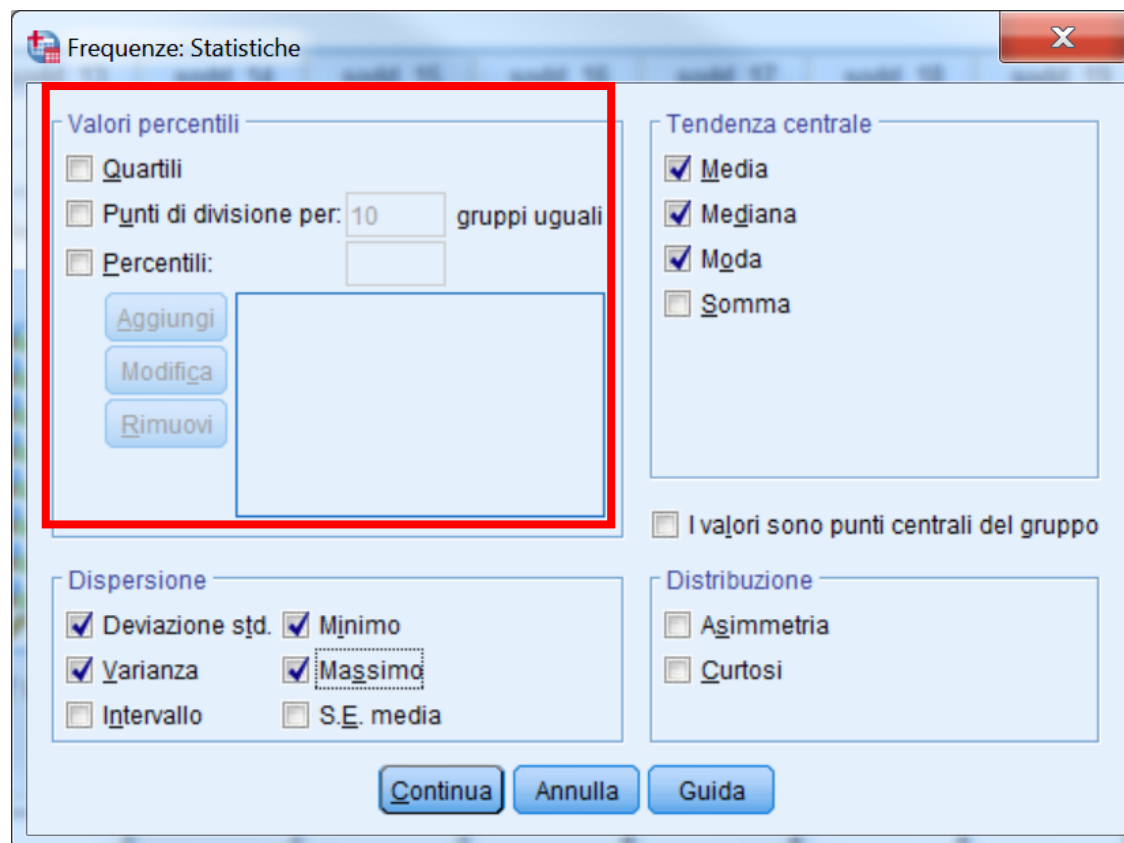
Indici di dispersione

Indici di tendenza centrale

L'opzione "Statistiche" consente calcolare una serie di statistiche, come ad gli indici di tendenza centrale e gli indici di dispersione

# SPSS

È possibile inoltre calcolare i quartili e percentili.

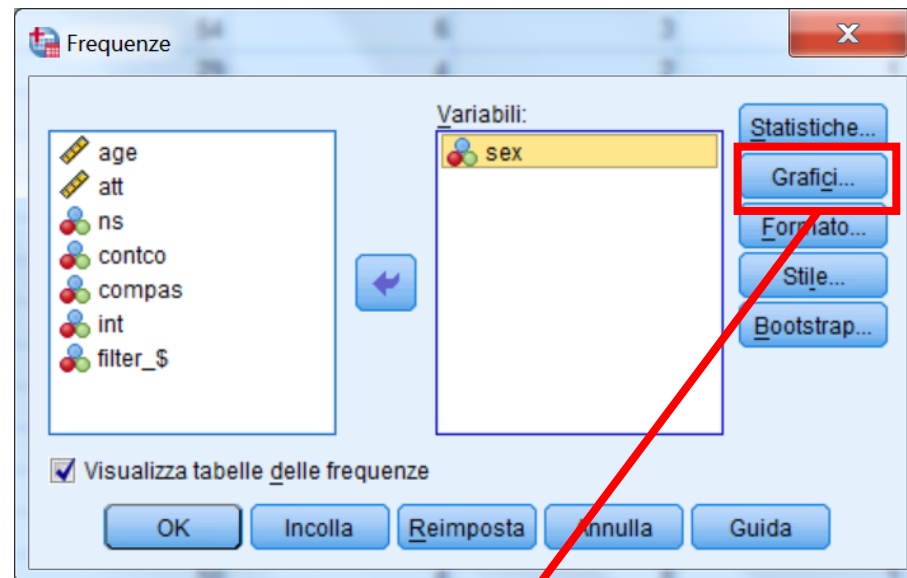


**RICORDA:** i Quartili indicano quei valori che dividono la distribuzione in quattro parti uguali.

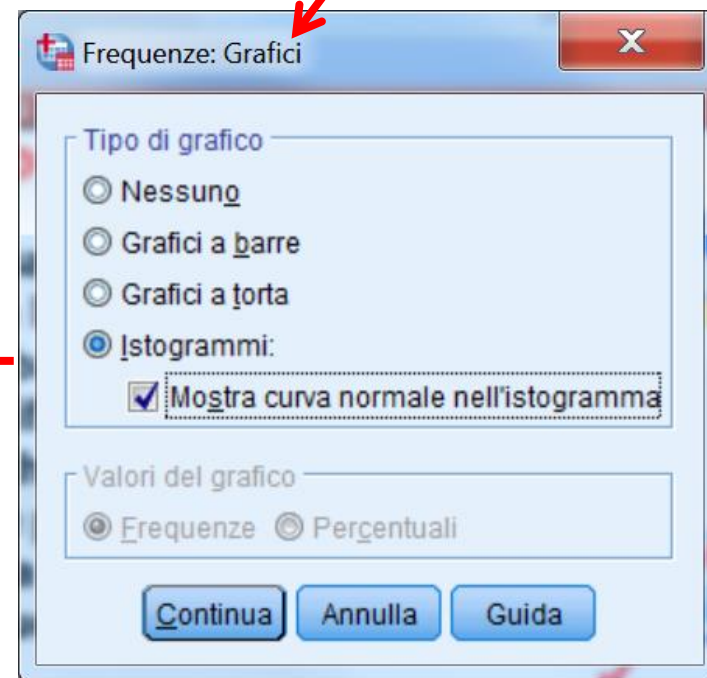
I Percentili indicano quei valori che dividono la distribuzione in 100 parti uguali.

# SPSS

Selezionando il pulsante "Grafici" si aprirà la finestra riportata in basso, che consente di specificare il tipo di grafico che vogliamo utilizzare per rappresentare la distribuzione di frequenze.



Questa finestra consente di creare diversi tipi di grafici: grafici a barre, grafici a torta e Istogrammi

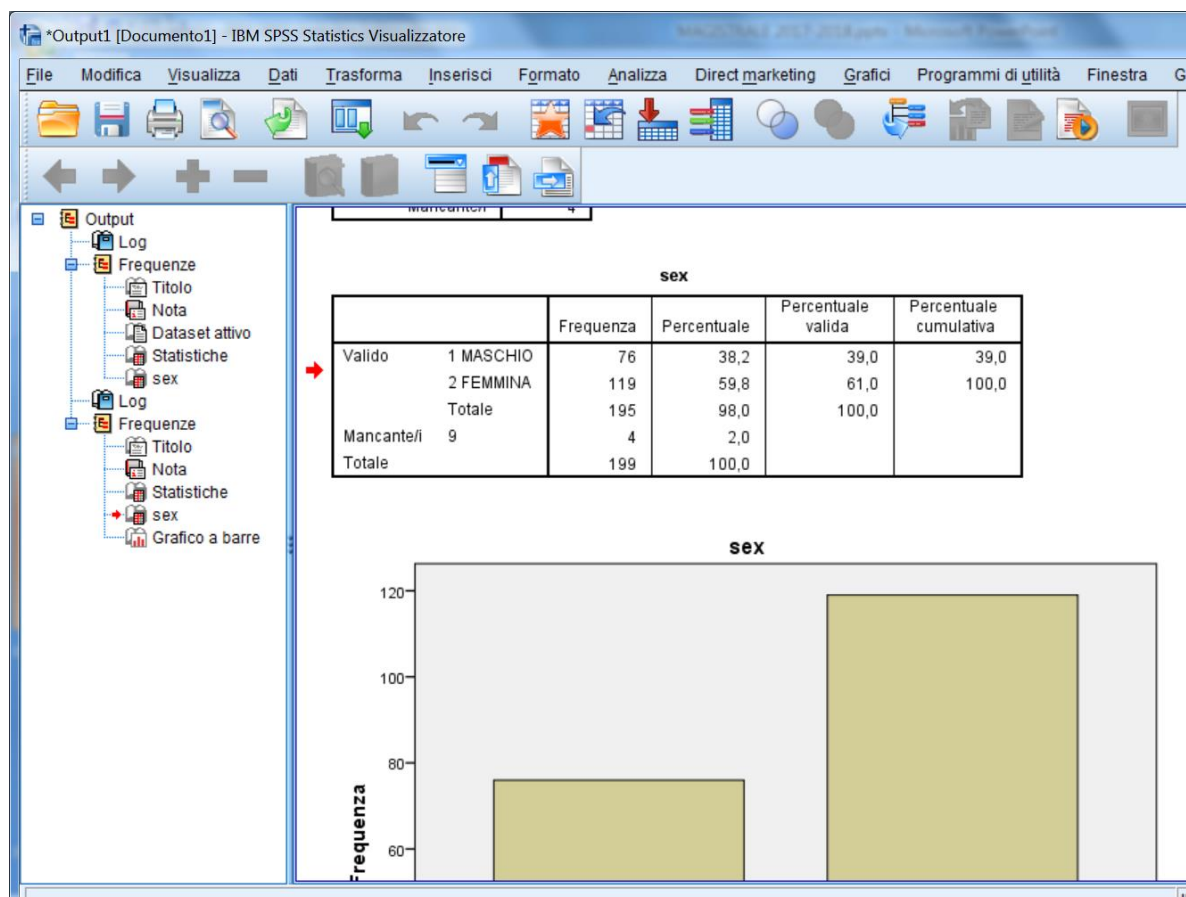


# SPSS

## L'output della procedura Frequenze

Nell'output vengono riportate una serie di tabelle e un grafico

La tabella seguente contiene i valori delle statistiche descrittive che abbiamo richiesto nella finestra "Statistiche"



# SPSS

## L'output della procedura Frequenze

Nell'output vengono riportate una serie di tabelle e un grafico

La tabella seguente contiene i valori delle statistiche descrittive che abbiamo richiesto nella finestra "Statistiche"

sex

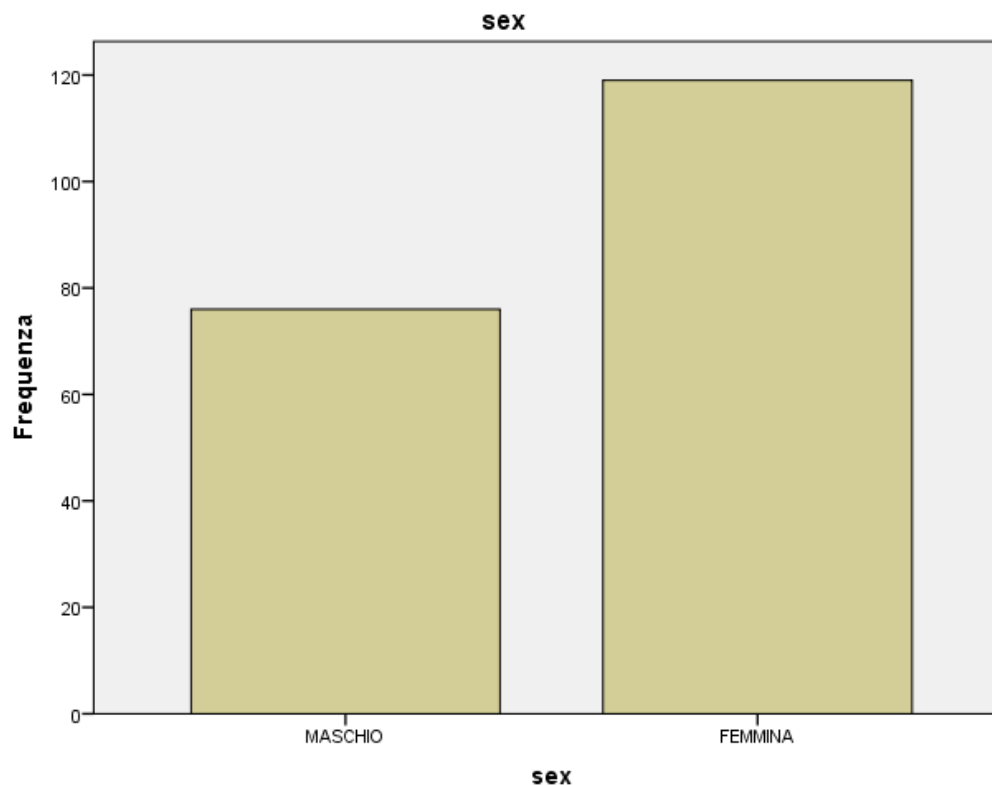
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulativa
Valido	1 MASCHIO	76	38,2	39,0	39,0
	2 FEMMINA	119	59,8	61,0	100,0
	Totale	195	98,0	100,0	
Mancanteli	9	4	2,0		
Totale		199	100,0		

# SPSS

## L'output della procedura Frequenze

Nell'output vengono riportate una serie di tabelle e un grafico

Il grafico seguente contiene il diagramma a barre delle frequenze



# SPSS

## L'output della procedura Frequenze

Nell'output vengono riportate una serie di tabelle e un grafico

La tabella seguente contiene i valori delle statistiche descrittive che abbiamo richiesto nella finestra "Statistiche"

### Statistiche

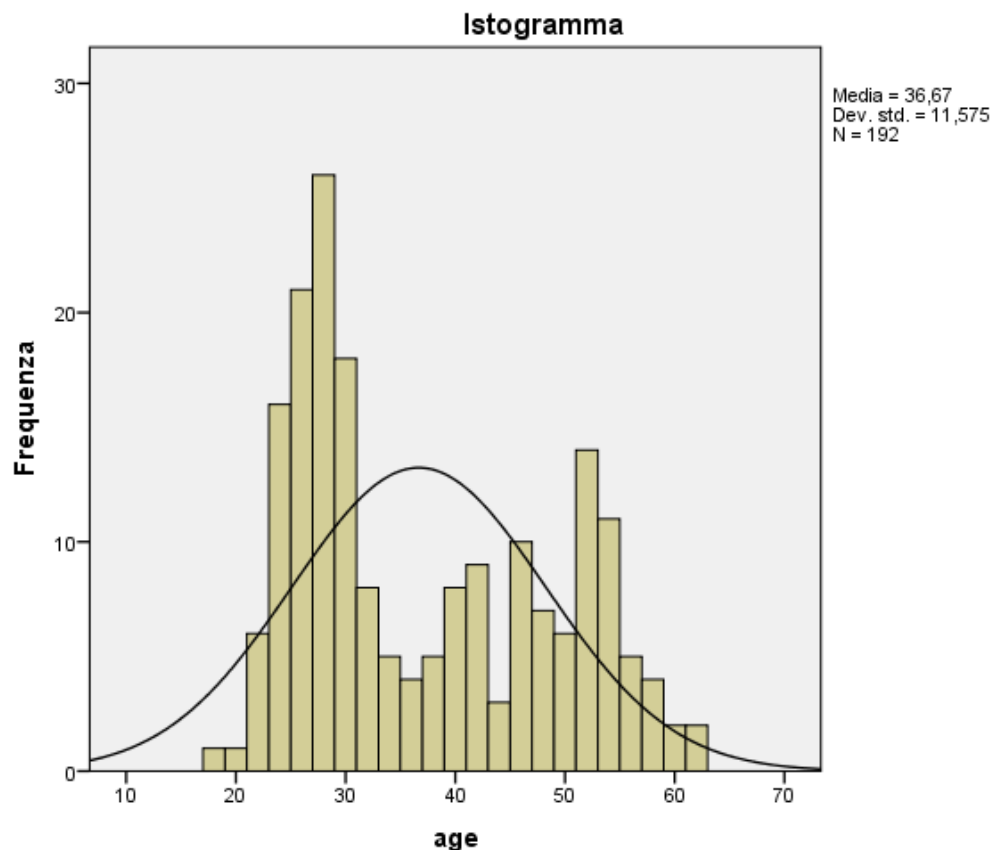
age

N	Valido	192
	Mancante/i	7
Media		36,67
Mediana		32,00
Modalità		28
Deviazione std.		11,575
Varianza		133,983
Asimmetria		,425
Errore standard dell'asimmetria		,175
Curtosi		-1,219
Errore standard della curtosi		,349
Minimo		18
Massimo		62

# SPSS

## L'output della procedura Frequenze

In questa figura viene riportato l'istogramma della variabile



**Una curva normale sovrapposta all'istogramma consente di valutare se i dati sono distribuiti normalmente**



# SPSS

## La procedura Descrittive

La procedura "Descrittive" consente di calcolare alcune statistiche univariate. Questa procedura è consigliabile per lo screening di file che contengono molte variabili

\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma **Analizza** Direct marketing Grafici Programmi di utilità Finestra Guida

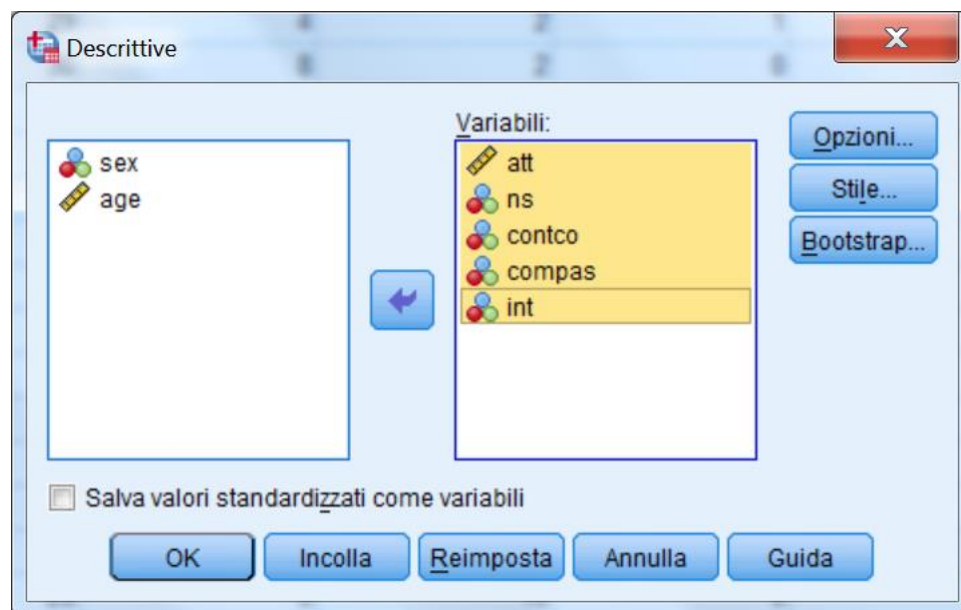
Report  
**Statistiche descrittive**  
 Tabelle  
 Confronta medie  
 Modello lineare generalizzato  
 Modelli lineari generalizzati  
 Modelli misti  
 Correlazione  
 Regressione  
 Log-lineare  
 Reti neurali  
 Classifica  
 Riduzione delle dimensioni  
 Scala

123 Frequenze...  
**Descrittive...**  
 Esplora...  
 Tabelle di contingenza...  
 Analisi TURF  
 Rapporto...  
 Grafici P-P...  
 Grafici Q-Q...

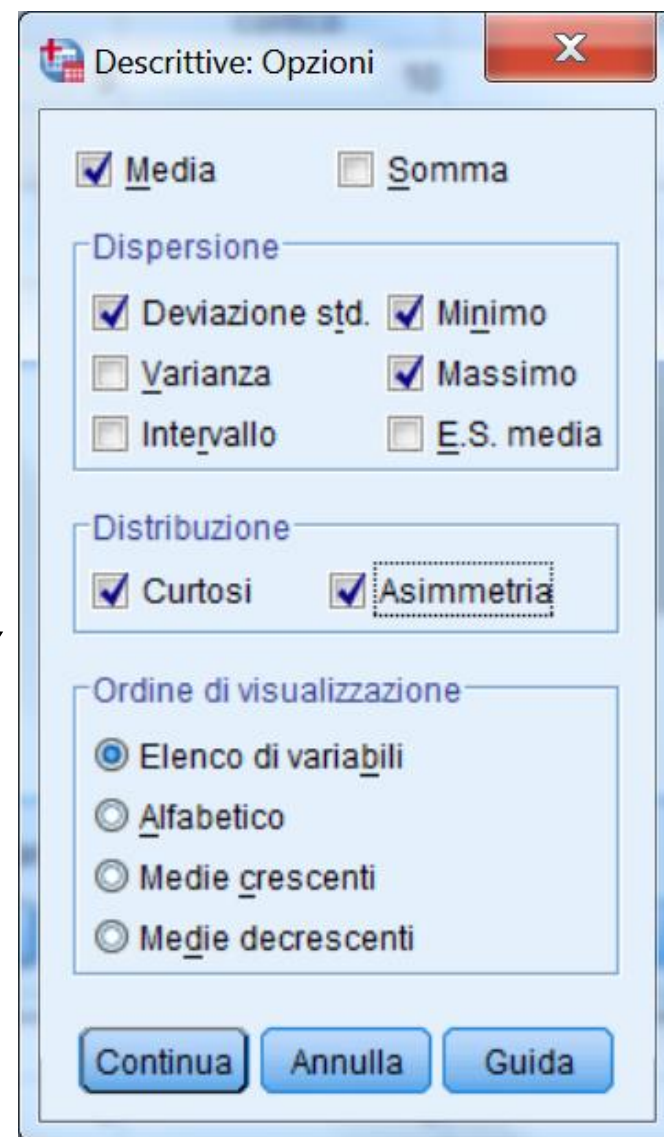
	sex	age
1	1	
2	1	
3	1	
4	1	
5	9	
6	2	
7	9	
8	2	
9	2	
10	1	
11	1	

0	4	2
6	2	0
8	8	4
8	10	0
0	8	8

# Chiedere le statistiche descrittive per vedere se le distribuzioni delle variabili sono normali



Selezionando il pulsante "Opzioni" si apre questa finestra di dialogo:



## Statistiche descrittive

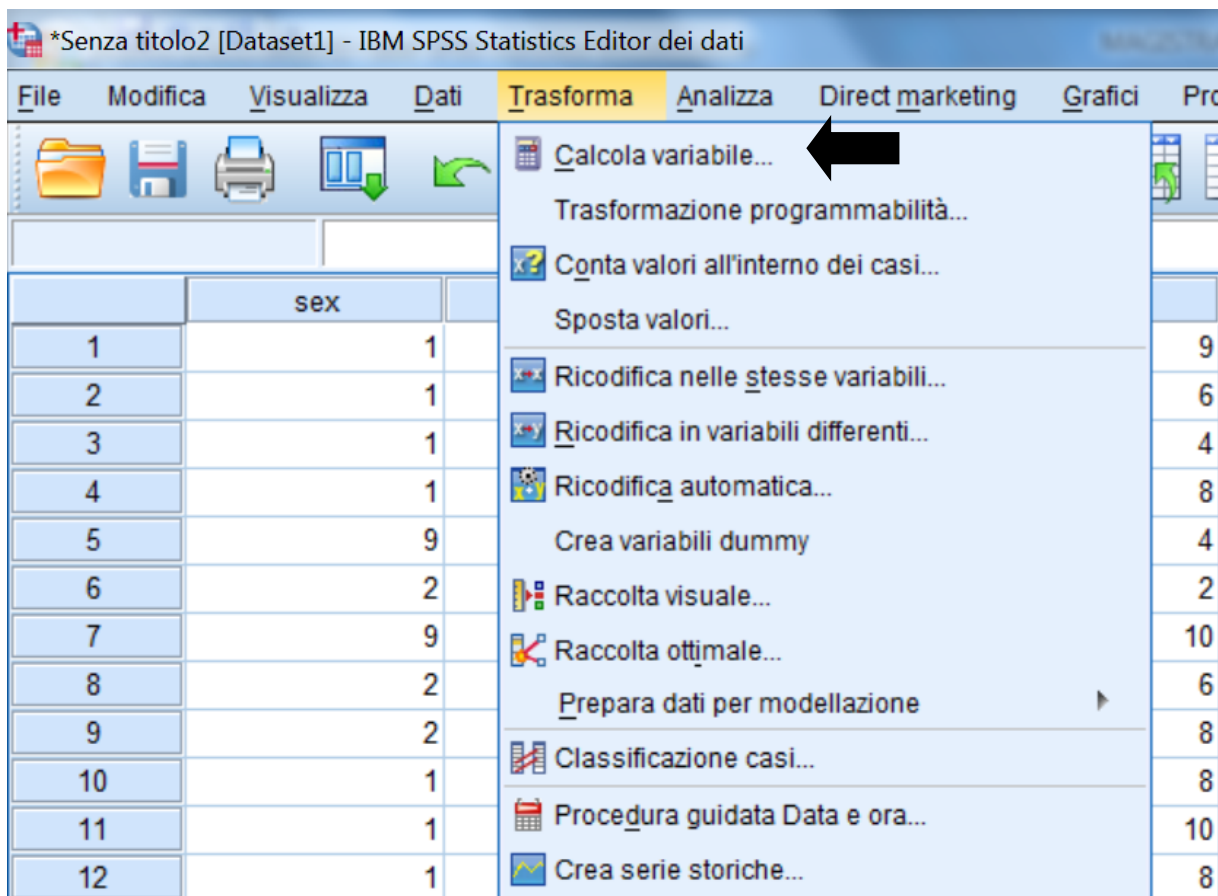
	N	Minimo	Massimo	Media	Deviazione std.	Asimmetria		Curtosi	
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Errore std	Statistica	Errore std
att	199	16	54	42,80	7,311	-,869	,172	,209	,343
ns	199	2	10	7,88	1,801	-,554	,172	-,339	,343
contco	199	2	10	8,68	1,863	-1,850	,172	3,444	,343
compas	199	0	8	2,64	1,969	,391	,172	-,446	,343
int	199	2	10	7,29	2,544	-,680	,172	-,630	,343
Validi (listwise)	199								

**Questa variabile ha una distribuzione fortemente non normale !**

# Trasformazione della variabile "contco"

**Asimmetria Negativa Sostanziale**  
**(valori tra -1 e -2)**

**Logaritmo =  $X^* = \text{Log}_{10}(K-X)$**   
**(qui,  $K = 10+1=11$ )**



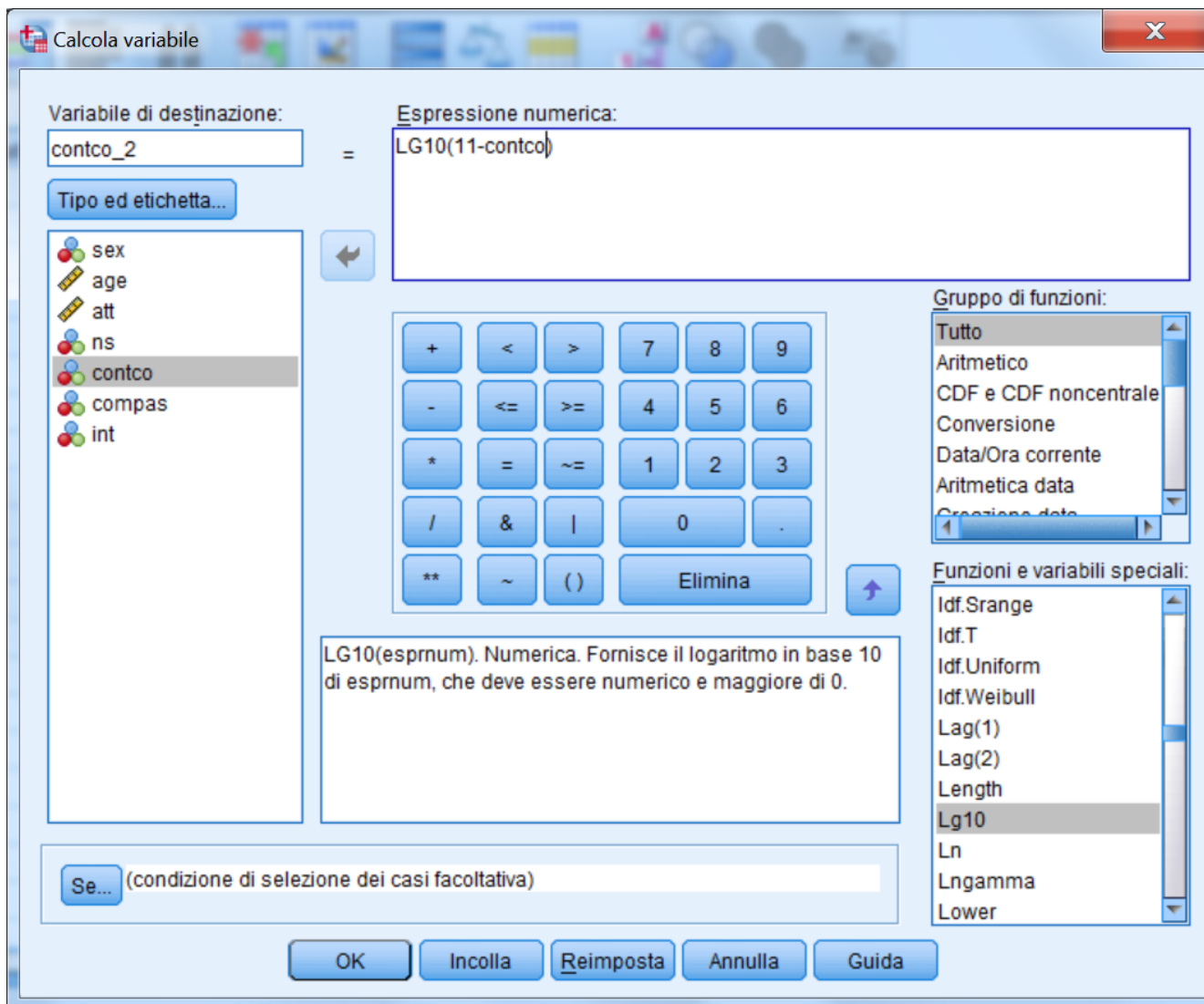
The screenshot shows the IBM SPSS Statistics Editor interface. The 'Trasforma' menu is open, and the 'Calcola variabile...' option is selected, indicated by a black arrow. The background shows a data table with columns 'sex' and 'contco'.

	sex	contco
1	1	9
2	1	6
3	1	4
4	1	8
5	9	4
6	2	2
7	9	10
8	2	6
9	2	8
10	1	8
11	1	10
12	1	8

# Ricodificare la variabile "contco"

**Asimmetria Negativa Sostanziale**  
**(valori tra -1 e -2)**

**Logaritmo =  $X^* = \text{Log}_{10}(K-X)$**   
**( $K = \max(X)+1=10+1=11$ )**



## Ricodificare la variabile "contco"

**Asimmetria Negativa Sostanziale**  
**(valori tra -1 e -2)**

**Logaritmo =  $X^* = \text{Log}_{10}(K-X)$**   
**(qui,  $K = 10+1=11$ )**

**Comando di Sintassi:**

**COMPUTE contco\_2=LG10(11-contco).**

\*Senza titolo2 [Dataset1] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Guida

1 : contco\_2 ,0

	sex	age	att	ns	contco	compas	int	contco_2
1	1	43	16	9	10	2	7	,00
2	1	30	54	6	3	0	3	,90
3	1	45	29	4	2	1	4	,95
4	1	34	30	8	2	0	2	,95
5	9	99	37	4	2	0	2	,95
6	2	51	32	2	8	0	6	,48
7	9	99	31	10	4	2	4	,85
8	2	28	30	6	2	0	4	,95
9	2	26	30	8	8	4	2	,48
10	1	30	42	8	10	0	2	,00
11	1	51	43	10	8	8	8	,48
12	1	50	22	8	9	0	4	,30
13	1	29	34	6	2	0	2	,95
14	2	32	27	8	10	4	8	,00
15	1	40	50	4	8	1	6	,48
16	1	28	28	6	10	0	2	,00
17	2	26	50	10	8	0	10	,48
18	1	28	32	6	3	1	4	,90
19	2	18	42	4	10	0	4	,00
20	2	25	24	4	4	0	2	,85
21	2	33	50	10	10	0	10	,00

# Controlliamo se la normalizzazione è avvenuta chiedendo di nuovo le descrittive

Statistiche descrittive

	N	Minimo	Massimo	Media	Deviazione std.	Asimmetria		Curtosi	
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Errore std	Statistica	Errore std
att	199	16	54	42,80	7,311	-,869	,172	,209	,343
ns	199	2	10	7,88	1,801	-,554	,172	-,339	,343
contco	199	2	10	8,68	1,863	-1,850	,172	3,444	,343
compas	199	0	8	2,64	1,969	,391	,172	-,446	,343
int	199	2	10	7,29	2,544	-,680	,172	-,630	,343
contco_2	199	,00	,95	,2576	,29111	,664	,172	-,767	,343
Validi (listwise)	199								

**Ora i valori sono accettabili !**



## Esplorazione dei dati: gli outliers (valori anomali)

**I valori anomali sono quei valori che risultano differenziarsi particolarmente nella distribuzione dei punteggi.**

**I valori anomali, o outliers, **univariati** sono quei casi che in una variabile presentano valori estremamente elevati o estremamente bassi rispetto al resto della distribuzione.**

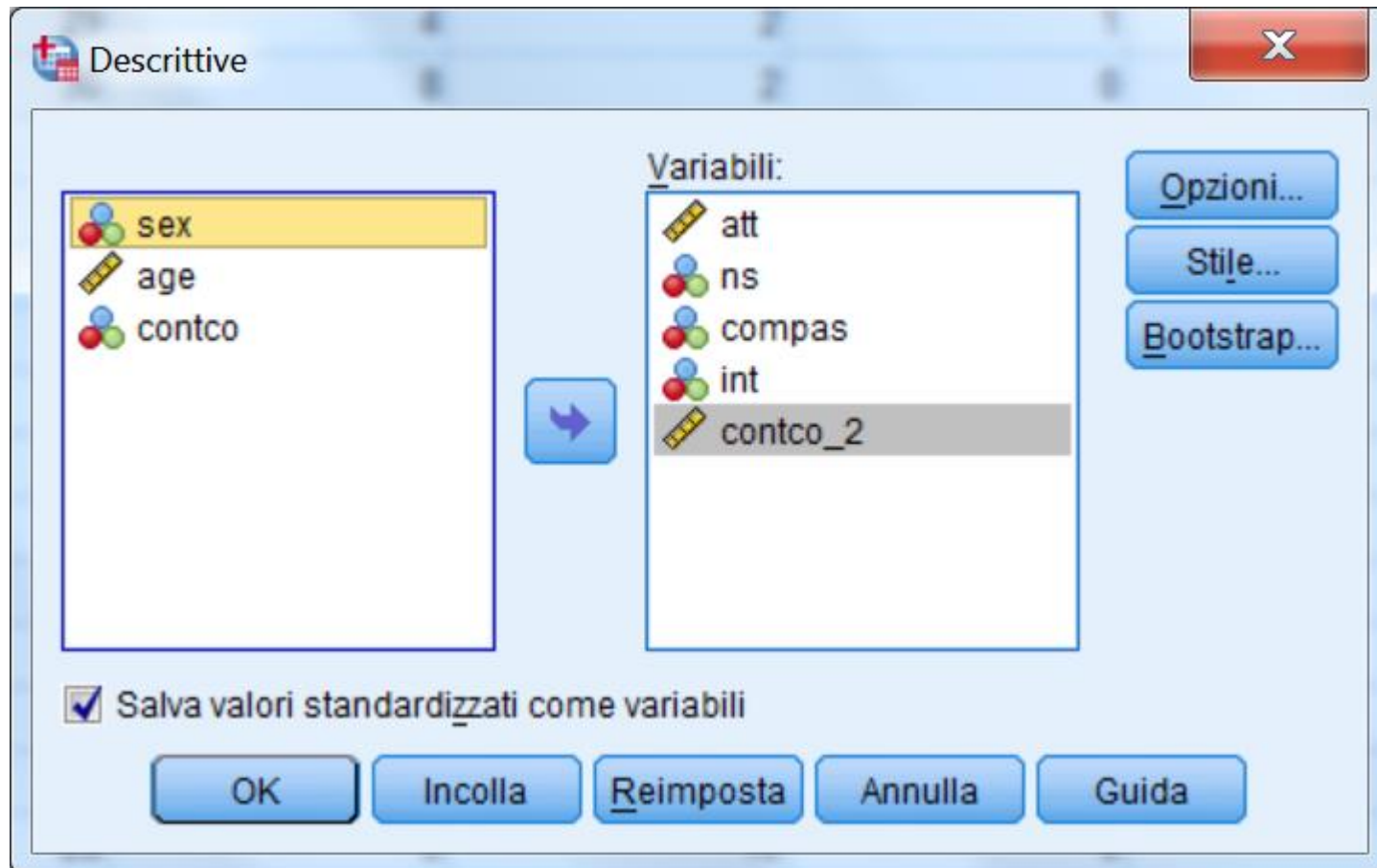
**Per individuare gli outliers univariati è possibile **standardizzare** i punteggi relativi alla variabile in esame e chiedere una distribuzione delle frequenze.**

**Vengono considerati come possibili valori anomali quei punteggi che corrispondono a una **z maggiore di 3 in valore assoluto**.**

**E' necessario considerare la distribuzione nella sua interezza e vedere se i punteggi troppo alti o troppo bassi rappresentano casi isolati dal resto della distribuzione oppure no.**

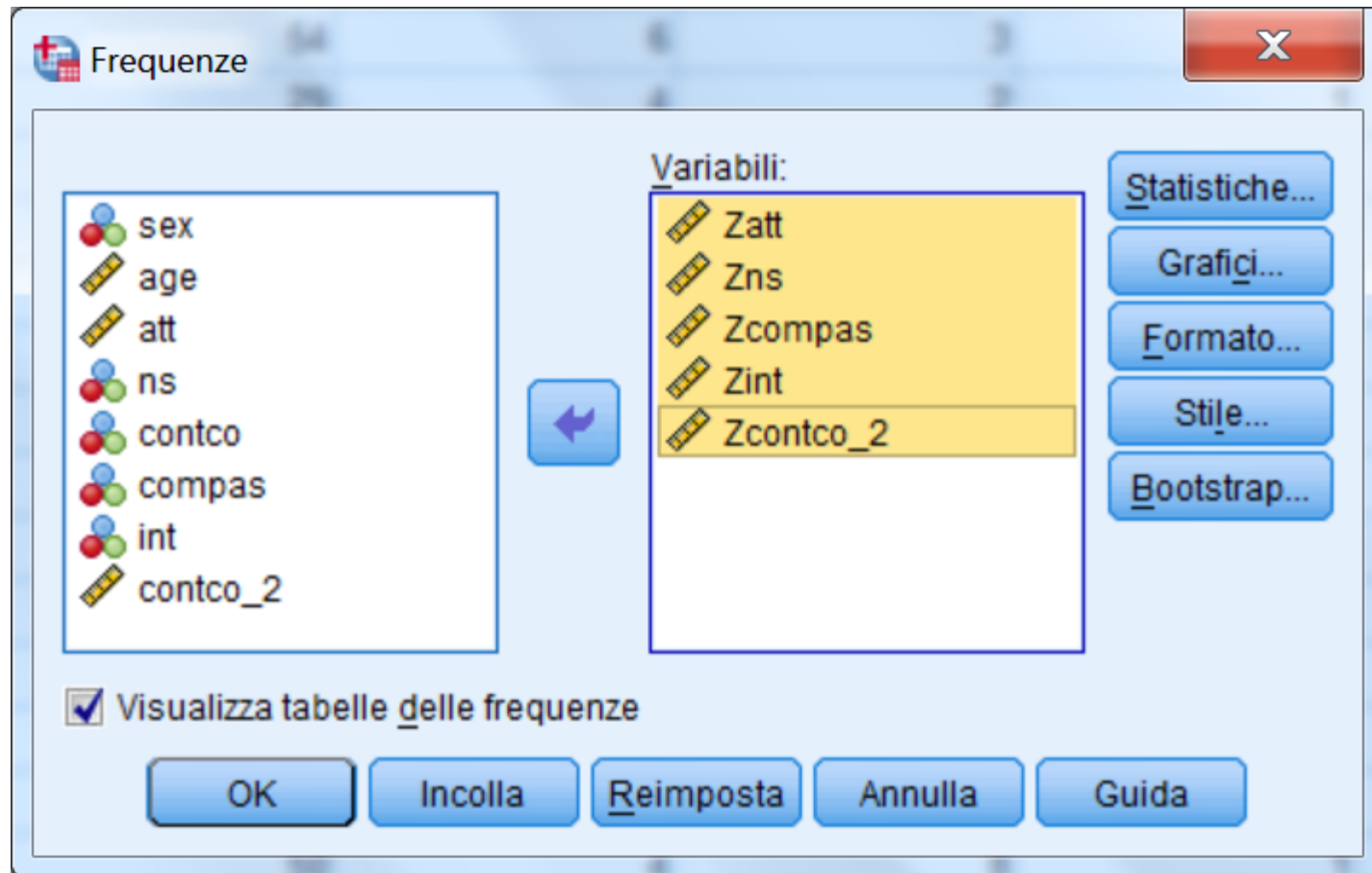
# Esplorazione dei dati: gli outliers (valori anomali)

## Standardizziamo (z) le variabili



# Esplorazione dei dati: gli outliers (valori anomali)

Chiediamo le frequenze delle nuove z



# Esplorazione dei dati: gli outliers (valori anomali)

Punteg(att)

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	-3,66545	1	,5	,5
	-2,84479	1	,5	1,0
	-2,57124	1	,5	1,5
	-2,29769	2	1,0	2,5
	-2,16092	1	,5	3,0
	-2,02414	3	1,5	4,5
	-1,88737	3	1,5	6,0
	-1,75059	5	2,5	8,5
	-1,61381	2	1,0	9,5
	-1,47704	2	1,0	10,6
	-1,34026	1	,5	11,1
	-1,20349	8	4,0	15,1
	-1,06671	1	,5	15,6
	-,92994	5	2,5	18,1
	-,79316	8	4,0	22,1
	-,65639	10	5,0	27,1
	-,51961	10	5,0	32,2
	-,38283	6	3,0	35,2
	-,24606	8	4,0	39,2
	-,10928	12	6,0	45,2
	,02749	7	3,5	48,7
	,16427	7	3,5	52,3
	,30104	2	1,0	53,3
	,43782	6	3,0	56,3
	,57459	12	6,0	62,3
	,71137	7	3,5	65,8
	,84815	8	4,0	69,8
	,98492	59	29,6	99,5
	1,53202	1	,5	100,0
Totale	199	100,0	100,0	

Punteg(ns)

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	-3,26703	1	,5	,5
	-2,15663	9	4,5	5,0
	-1,60143	10	5,0	10,1
	-1,04623	30	15,1	25,1
	-,49103	17	8,5	33,7
	,06417	62	31,2	64,8
	,61937	14	7,0	71,9
	1,17457	56	28,1	100,0
Totale	199	100,0	100,0	

**Chi sono questi due  
soggetti ?**

# Esplorazione dei dati: gli outliers (valori anomali)

DATI\_TRAT\_PREL.sav [Dataset3] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Guida

1 : Zatt -3,66544756828786

	int	contco_2	Zatt	Zns	Zcompas	Zint	Zcontco_2
1	7	,00	-3,66545	,61937	-,32667	-,11458	-,88476
2	3	,90	1,53202	-1,04623	-1,34239	-1,68712	2,21750
3	4	,95	-1,88737	-2,15663	-,83453	-1,29399	2,39322
4	2	,95	-1,75059	,06417	-1,34239	-2,08026	2,39322
5	2	,95	-,79316	-2,15663	-1,34239	-2,08026	2,39322
6	6	,48	-1,47704	-3,26703	-1,34239	-,50772	,75423
7	4	,85	-1,61381	1,17457	-,32667	-1,29399	2,01829
8	4	,95	-1,75059	-1,04623	-1,34239	-1,29399	2,39322
9	2	,48	-1,75059	,06417	,68906	-2,08026	,75423
10	2	,00	-,10928	,06417	-1,34239	-2,08026	-,88476
11	8	,48	,02749	1,17457	2,72051	,27855	,75423
12	4	,30	-2,84479	,06417	-1,34239	-1,29399	,14933
13	2	,95	-1,20349	-1,04623	-1,34239	-2,08026	2,39322
14	8	,00	-2,16092	,06417	,68906	,27855	-,88476
15	6	,48	,98492	-2,15663	-,83453	-,50772	,75423
16	2	,00	-2,02414	-1,04623	-1,34239	-2,08026	-,88476
17	10	,48	,98492	1,17457	-1,34239	1,06482	,75423
18	4	,90	-1,47704	-1,04623	-,83453	-1,29399	2,21750

DATI\_TRAT\_PREL.sav [Dataset3] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza **Dati** Trasforma Analizza Direct marketing Grafici Pr

1 : Zatt -3,665

	int
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	

Vista dati Vista Variabile

Seleziona casi...

- Definisci proprietà variabili...
- Imposta livello di misurazione per sconosciuto...
- Copia proprietà dei dati...
- Nuovo attributo personalizzato...
- Definisci date...
- Definisci insiem*i* a risposta multipla...
- Convalida
- Identifica casi duplicati...
- Identifica casi insoliti...
- Confronta dataset...
- Ordina casi...
- Ordina le variabili...
- Trasponi...
- Unisci file
- Ristruttura...
- Esegui raking dei pesi...
- Messa in corrispondenza punteggi propensione...
- Corrispondenza caso-controllo...
- Aggrega...
- Suddividi in file
- Progettazione ortogonale
- Copia dataset
- File di suddivisione...
- Seleziona casi...**
- Pesa casi...

**Filtrare i soggetti  
escludendo i due  
outliers**



# Filtrare i soggetti escludendo i due outliers

The image shows the SPSS 'Seleziona casi' (Select Cases) dialog box. The 'Seleziona' section is set to 'Se la condizione è soddisfatta' (If the condition is satisfied). The filter expression 'Zatt > -3 & Zns > -3' is entered in the text box. A calculator interface is visible, showing the expression being typed. The variable list on the left includes 'sex', 'age', 'att', 'ns', 'contco', 'compas', 'int', 'contco\_2', 'Zatt', 'Zns', 'Zcompas', 'Zint', and 'Zcontco\_2'. The 'Zns' variable is highlighted in the list. The 'Gruppo di funzioni' (Function Group) dropdown is set to 'Tutto' (All). The 'Funzioni e variabili speciali' (Special Functions and Variables) section is empty. The 'Stato cc' (Case Status) section is also empty. The 'Continua' (Continue), 'Annulla' (Cancel), and 'Guida' (Help) buttons are visible at the bottom.

Seleziona casi

Seleziona

Tutti i casi

Se la condizione è soddisfatta

Seleziona casi: Se

Zatt > -3 & Zns > -3

Gruppo di funzioni:

Tutto

Aritmetico

CDF e CDF noncentrale

Conversione

Data/Ora corrente

Aritmetica data

Creazione data

Funzioni e variabili speciali:

Stato cc

Zatt > -3 & Zns > -3

Continua Annulla Guida

# Filtrare i soggetti escludendo i due outliers

\*DATI\_TRAT\_PREL.sav [Dataset3] - IBM SPSS Statistics Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Direct marketing Grafici Programmi di utilità Finestra Guida

1 : sex 1

	Zatt	Zns	Zcompas	Zint	Zcontco_2	filter_\$
<del>1</del>	-3,66545	,61937	-,32667	-,11458	-,88476	0
2	1,53202	-1,04623	-1,34239	-1,68712	2,21750	1
3	-1,88737	-2,15663	-,83453	-1,29399	2,39322	1
4	-1,75059	,06417	-1,34239	-2,08026	2,39322	1
5	-,79316	-2,15663	-1,34239	-2,08026	2,39322	1
<del>6</del>	-1,47704	-3,26703	-1,34239	-,50772	,75423	0
7	-1,61381	1,17457	-,32667	-1,29399	2,01829	1
8	-1,75059	-1,04623	-1,34239	-1,29399	2,39322	1
9	-1,75059	,06417	,68906	-2,08026	,75423	1
10	-,10928	,06417	-1,34239	-2,08026	-,88476	1
11	,02749	1,17457	2,72051	,27855	,75423	1

IBM SPSS Statistics Il processore è pronto

Unicode: ON

Filtro attivo



# Filtrare i soggetti escludendo i due outliers

## Statistiche descrittive

	N	Minimo	Massimo	Media	Deviazione std.	Asimmetria		Curtosi	
	Statistica	Statistica	Statistica	Statistica	Statistica	Statistica	Errore std	Statistica	Errore std
att	197	22	54	42,99	7,050	-,760	,173	-,290	,345
ns	197	4	10	7,91	1,759	-,451	,173	-,674	,345
compas	197	0	8	2,66	1,969	,382	,173	-,452	,345
int	197	2	10	7,30	2,555	-,687	,173	-,639	,345
contco_2	197	,00	,95	,2578	,29159	,667	,173	-,765	,345
Validi (listwise)	197								

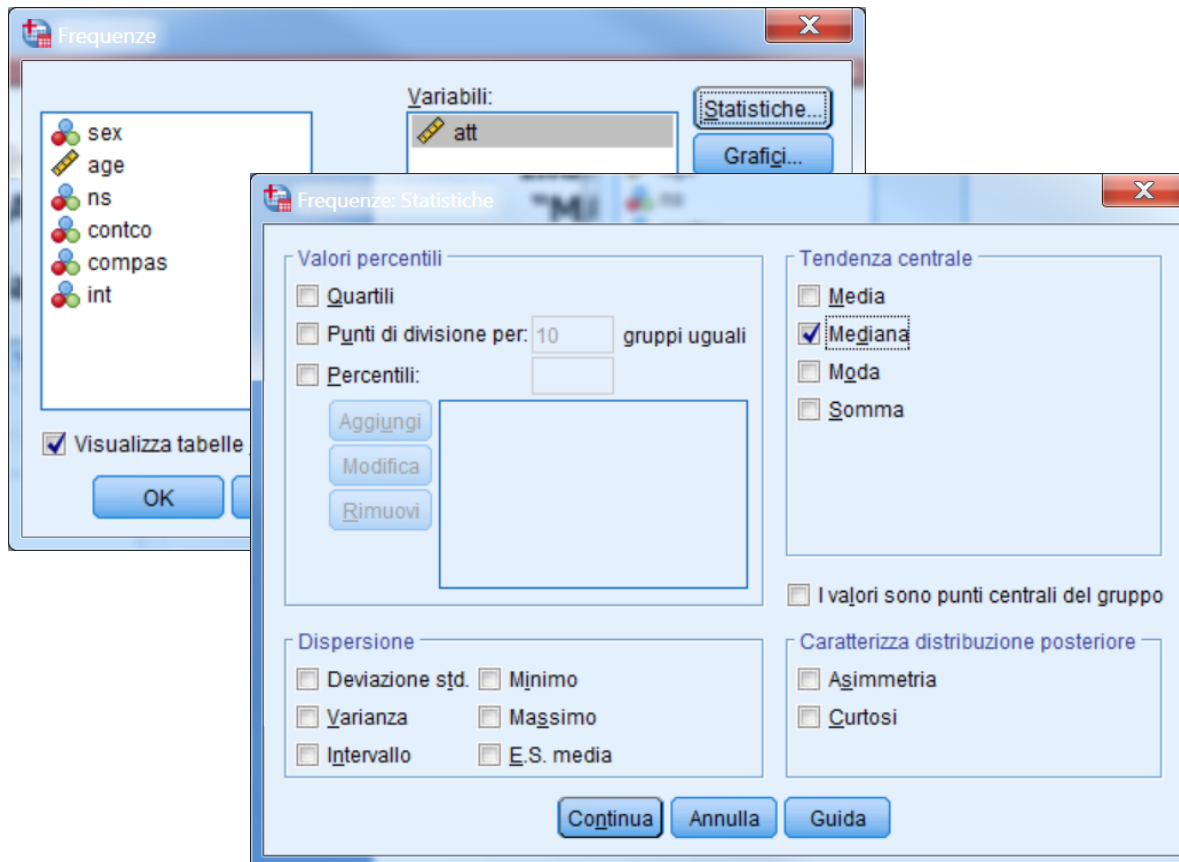
**Le distribuzioni migliorano !**

## Individuazione degli outliers tramite l'indice "MAD" (Valore Assoluto Mediano) on SPSS

- La procedura per calcolare l'indice MAD è semplice:
- (a) Calcolare la mediana tramite la procedura "Frequenze";
  - (b) Sottrarre la mediana dal punteggio nella variabile per ogni soggetto tramite "Trasforma/Calcola variabile" **in valore assoluto,  $|X_i - Mdn|$**  ;
  - (c) Calcolare la mediana della nuova variabile ("Frequenze"): questo è il "MAD"
  - (d) Per ogni soggetto calcolare la formula seguente con "Calcola variabile":  **$MAD_i = |X_i - Mdn| / (1.483 * MAD)$**
  - (e) Sono da considerare outliers quei soggetti il cui valore è maggiore di 3 o di 2.5

# Individuazione degli outliers tramite l'indice "MAD" (Valore Assoluto Mediano) on SPSS

Consideriamo il calcolo dell'indice MAD per ATT.

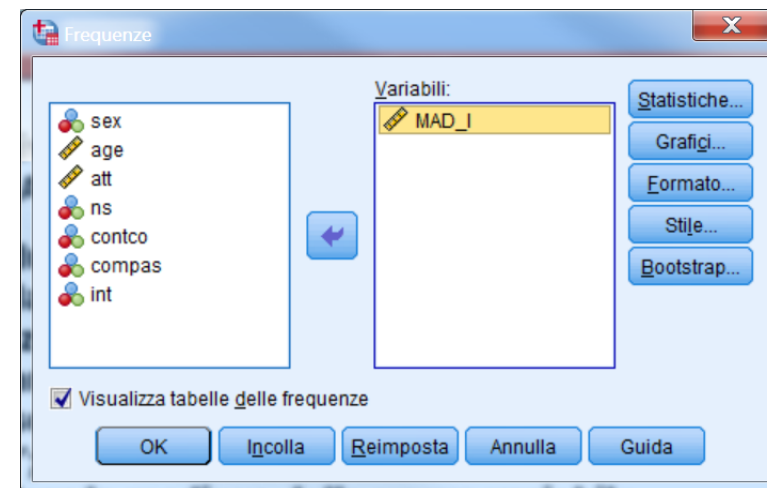
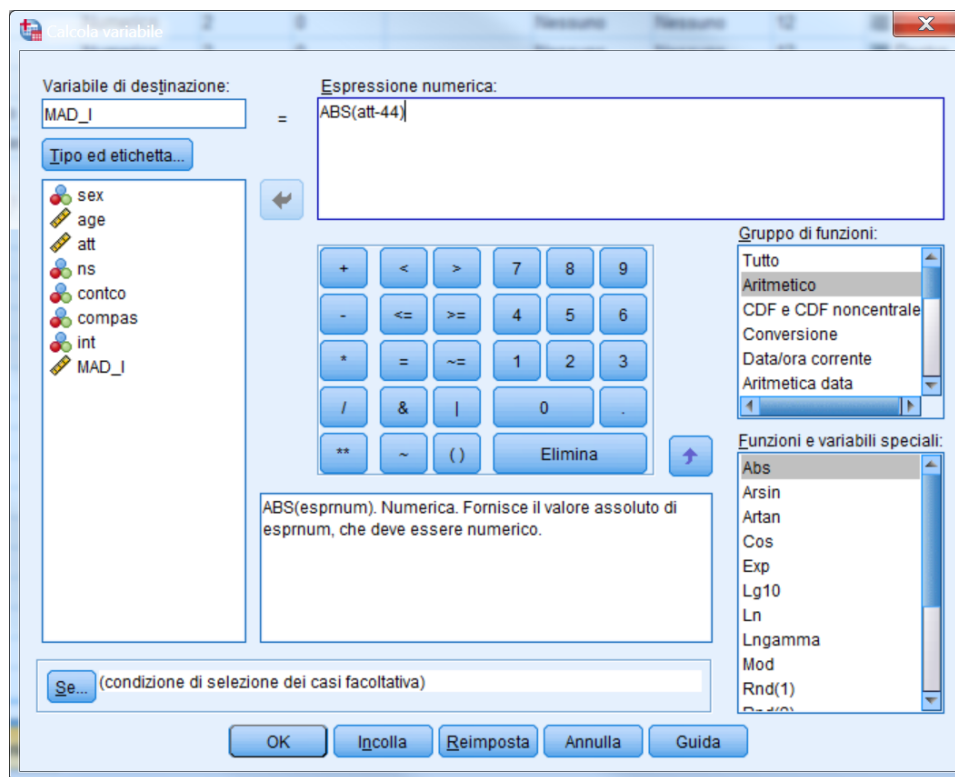


**Statistiche**

Mediana 44,00

# Individuazione degli outliers tramite l'indice "MAD" (Valore Assoluto Mediano) on SPSS

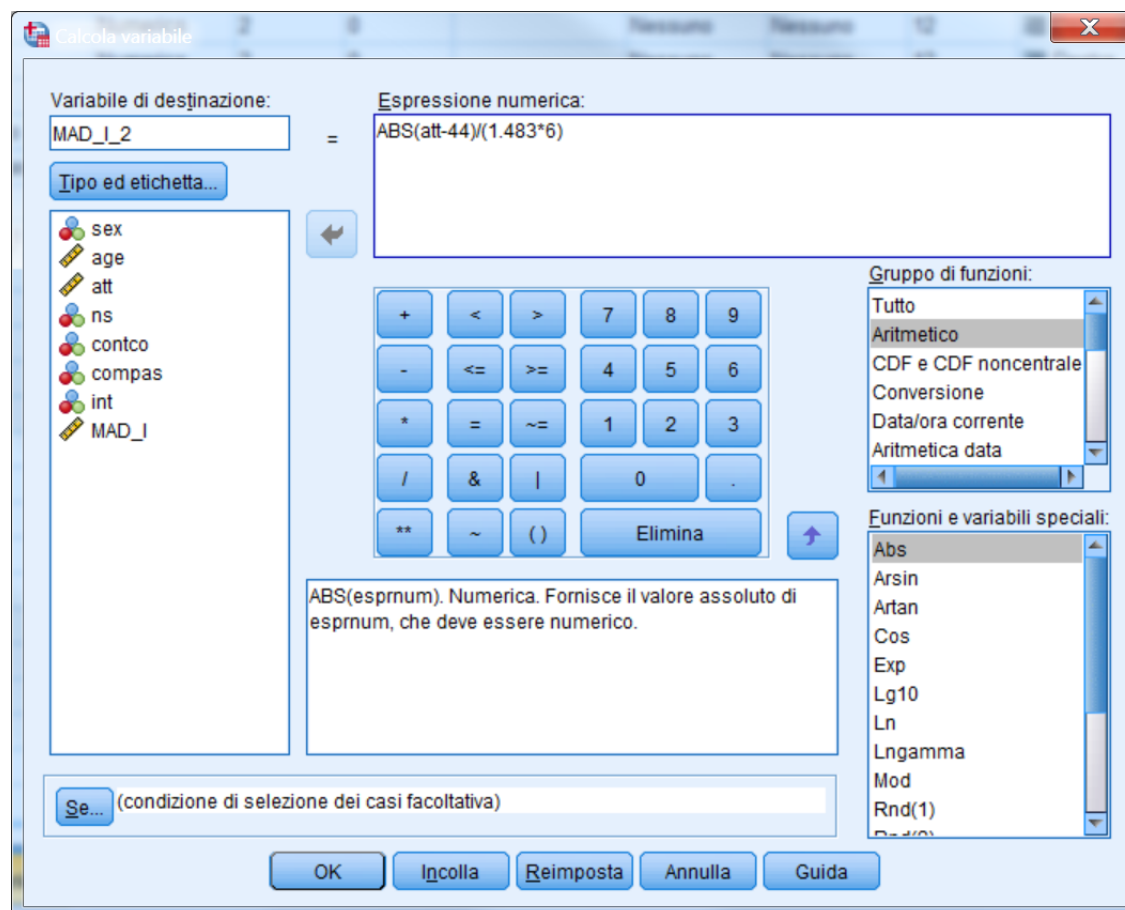
Consideriamo il calcolo dell'indice MAD per ATT.



**Statistiche MAD\_I**  
Mediana 6,0000

# Individuazione degli outliers tramite l'indice "MAD" (Valore Assoluto Mediano) on SPSS

Consideriamo il calcolo dell'indice MAD per ATT.



# Individuazione degli outliers tramite l'indice "MAD" (Valore Assoluto Mediano) on SPSS

In questo caso i due indici z e "MAD" danno risultati analoghi

	MAD_I_2	Zatt	var
0	3,15	-3,66545	
0	2,47	-2,84479	
0	2,25	-2,57124	
0	2,02	-2,29769	
0	2,02	-2,29769	
0	1,91	-2,16092	
0	1,80	-2,02414	
0	1,80	-2,02414	
0	1,80	-2,02414	
0	1,69	-1,88737	
0	1,60	-1,80727	

## Esplorazione dei dati: la normalità multivariata

**Per esaminare l'ipotesi di normalità multivariata Mardia ha sviluppato dei coefficienti di curtosi e di asimmetria multivariata. Se la distribuzione delle  $p$  variabili è normale multivariata, il coefficiente di curtosi multivariata di Mardia dovrebbe essere uguale a  $p(p+2)$  [ $p$ =numero di variabili].**

**Per valutare la normalità multivariata è possibile utilizzare un test grafico che si basa sull'utilizzo dei quantili della distribuzione del chi quadrato e sulla distanza generalizzata o distanza di Mahalanobis.**

**In SPSS la distanza di Mahalanobis è calcolabile utilizzando la procedura della regressione lineare multipla.**

## **Esplorazione dei dati: la normalità multivariata e outliers multivariati**

**\* Calcoliamo preliminarmente una nuova variabile (nord) alla quale vengono assegnati i valori della variabile di sistema \$casenum: questa variabile fornisce il numero d'ordine del soggetto nel file (es., il primo soggetto nel file avrà \$casenum = 1, e così via).**

**\* Questa nuova variabile verrà utilizzata come variabile dipendente in una regressione multipla che ha il solo scopo di calcolare per ogni soggetto la distanza di Mahalanobis, la quale viene salvata nel file come una nuova variabile con il nome mah\_1.**

**I comandi tramite le finestre di dialogo dei menù sono descritti di seguito.**



# Calcolo della variabile "nord"

Calcola variabile

Variabile di destinazione: nord

Espressione numerica: \$CASENUM

Tipo ed etichetta...

- sex
- age
- att
- ns
- contco
- compas
- int
- contco\_2
- Zatt
- Zns
- Zcompas
- Zint
- Zcontco\_2
- filter\_\$

Numero di sequenza del caso corrente. Per ogni caso, \$CASENUM rappresenta il numero di casi letti incluso il caso corrente. Il formato è F8.0. Il valore di \$CASENUM non è necessariamente il numero di riga in una finestra dell'Editor dati (disponibile negli ambienti a finestre) e il valore cambia se il file viene ordinato o se vengono inseriti nuovi casi prima della fine del file.

Gruppo di funzioni:

- Tutto
- Aritmetico
- CDF e CDF noncentrale
- Conversione
- Data/Ora corrente
- Aritmetica data
- Crescimento date

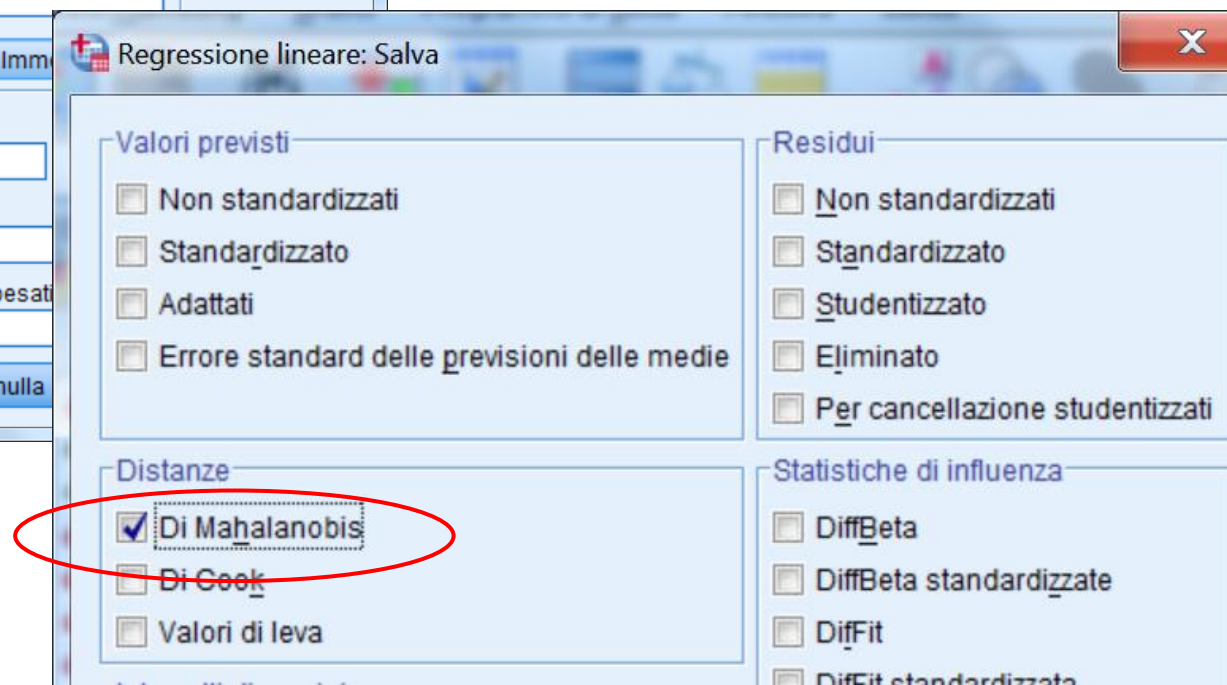
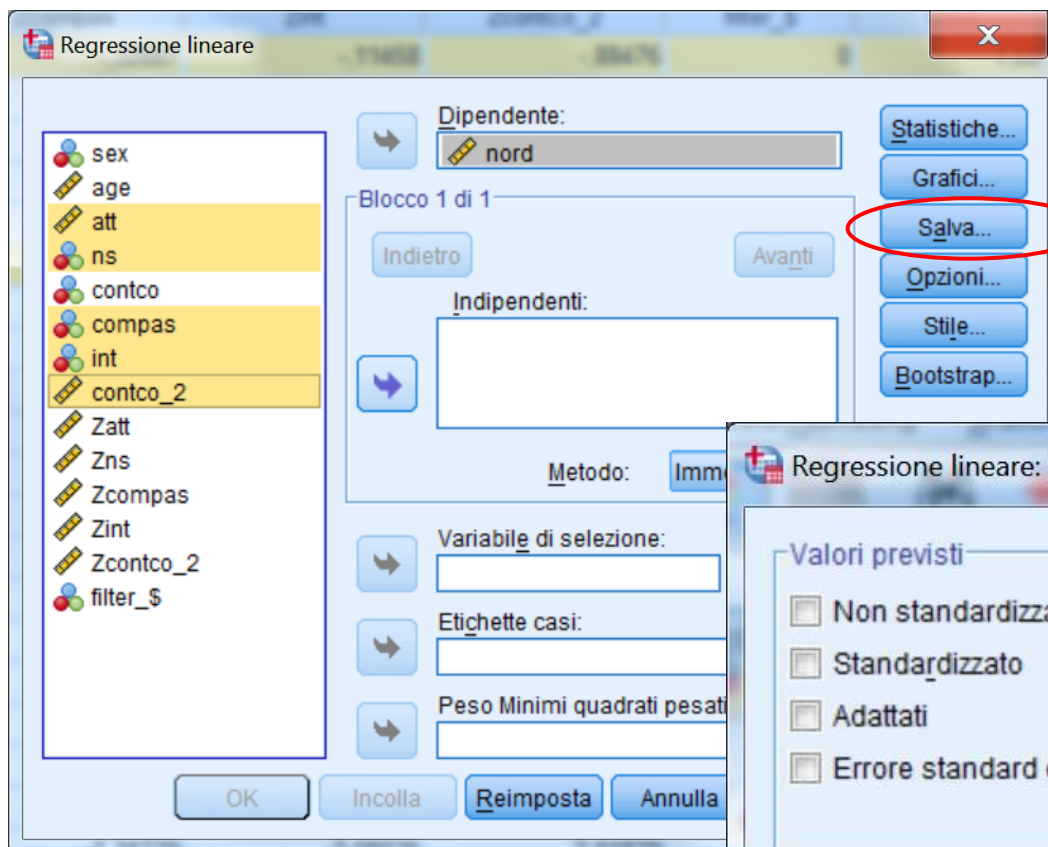
Funzioni e variabili speciali:

- \$Casenum
- \$Date
- \$Date11
- \$JDate
- \$Systemis
- \$Time
- Abs
- Any
- Applymodel
- Arsin
- Artan

Se... (condizione di selezione dei casi facoltativa)

OK Incolla Reimposta Annulla Guida

# Calcolo della distanza di Mahalanobis tramite regressione



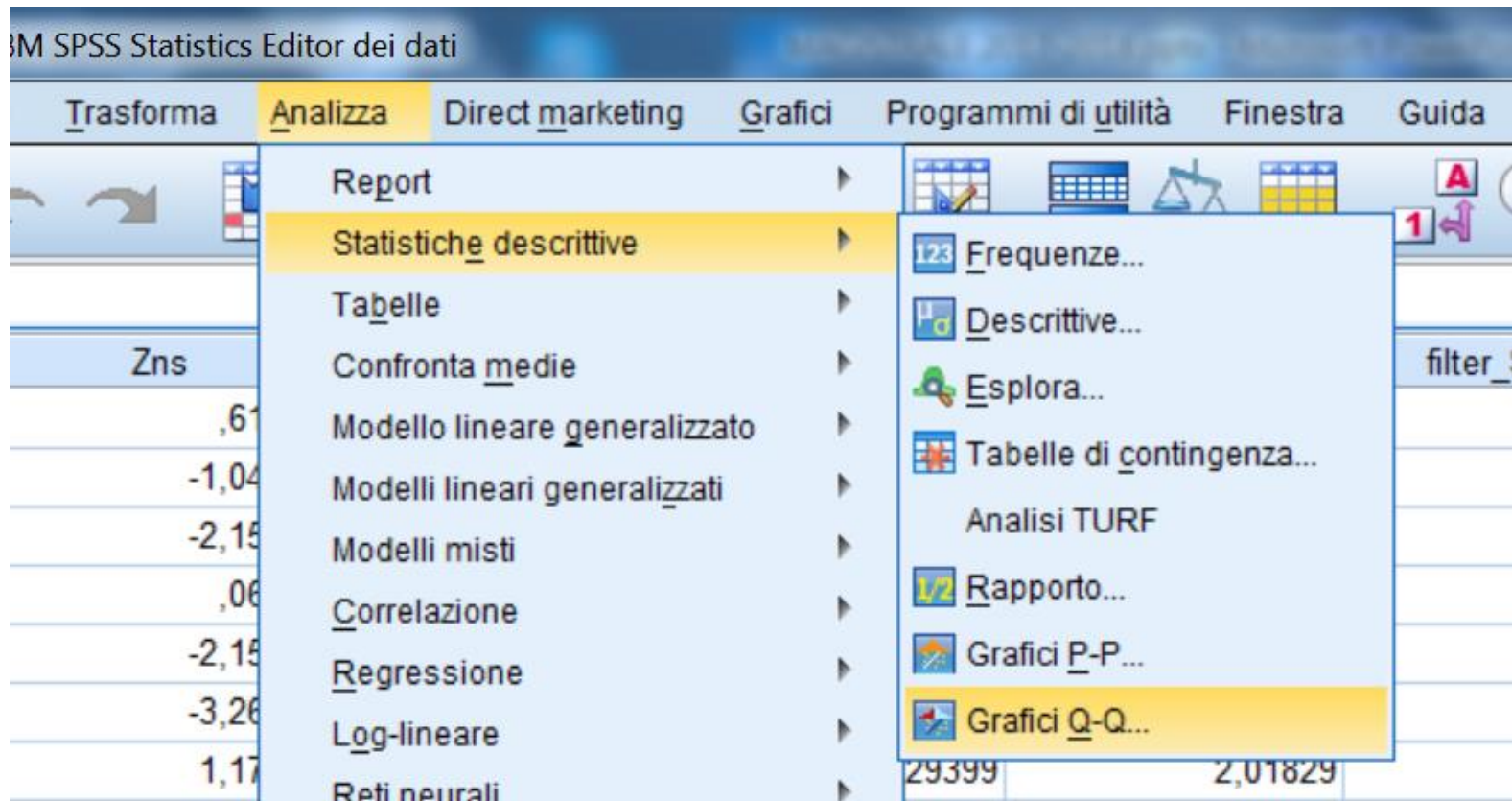
# Esplorazione dei dati: la normalità multivariata e



Zcompas	Zint	Zcontco_2	filter_\$	nord	MAH_1
-,32667	-,11458	-,88476	0	1,00	.
-1,34239	-1,68712	2,21750	1	2,00	21,76511
-,83453	-1,29399	2,39322	1	3,00	10,37619
-1,34239	-2,08026	2,39322	1	4,00	9,71748
-1,34239	-2,08026	2,39322	1	5,00	9,69798
-1,34239	-,50772	,75423	0	6,00	.
-,32667	-1,29399	2,01829	1	7,00	12,99378
-1,34239	-1,29399	2,39322	1	8,00	7,21699
,68906	-2,08026	,75423	1	9,00	14,27052
-1,34239	-2,08026	-,88476	1	10,00	15,05194
2,72051	,27855	,75423	1	11,00	14,89854
-1,34239	-1,29399	,14933	1	12,00	14,29484
-1,34239	-2,08026	2,39322	1	13,00	6,67944
,68906	,27855	-,88476	1	14,00	14,09949
-,83453	-,50772	,75423	1	15,00	12,84113
-1,34239	-2,08026	-,88476	1	16,00	11,75779
-1,34239	1,06482	,75423	1	17,00	10,98545
-,83453	-1,29399	2,21750	1	18,00	5,74019
-1,34239	-1,29399	-,88476	1	19,00	11,17339
-1,34239	-2,08026	2,01829	1	20,00	9,14158

**Viene creata la variabile MAH\_1 nel datafile**

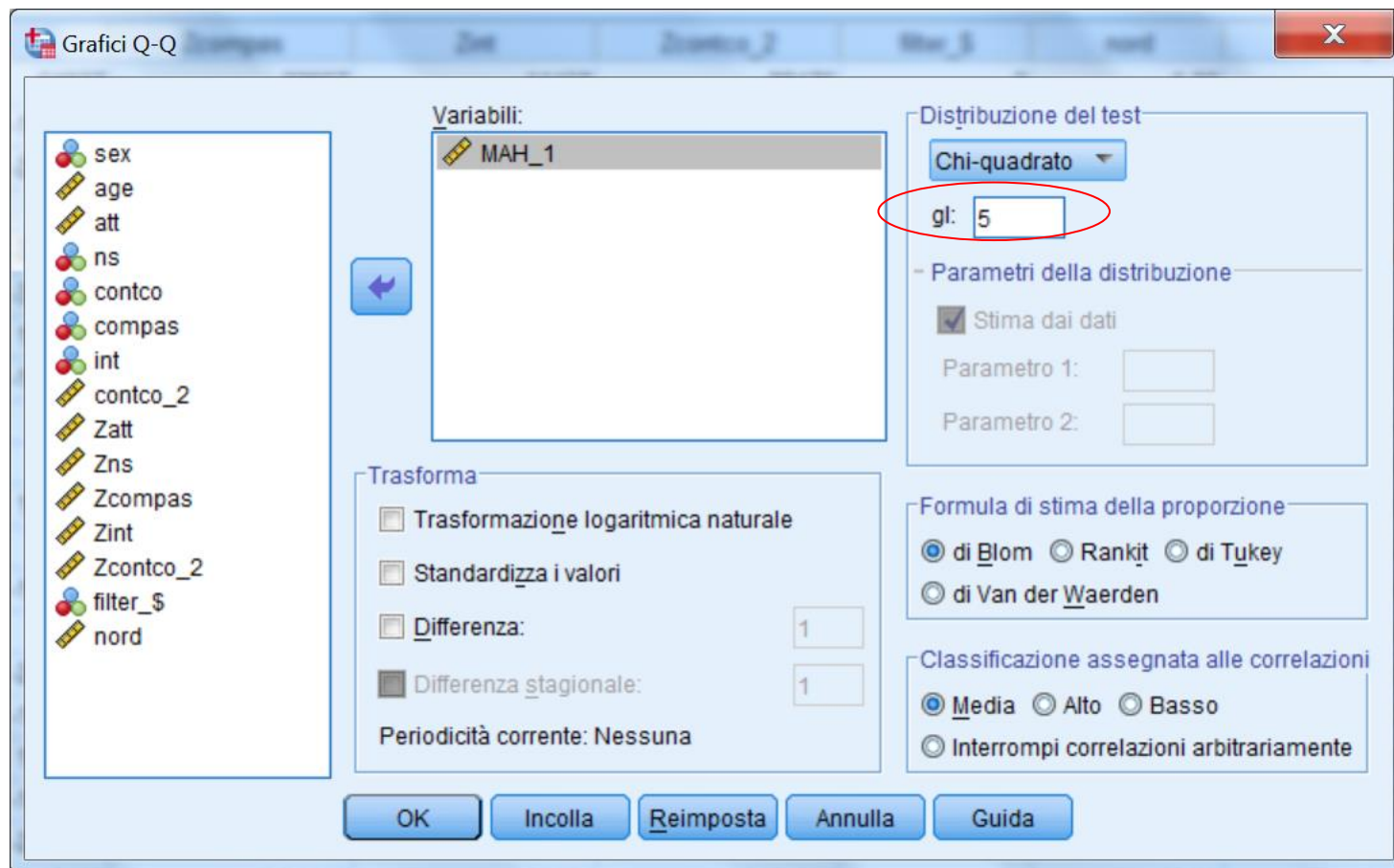
# Test grafico Q-Q Plot



**Si chiede tramite Statistiche descrittive...**

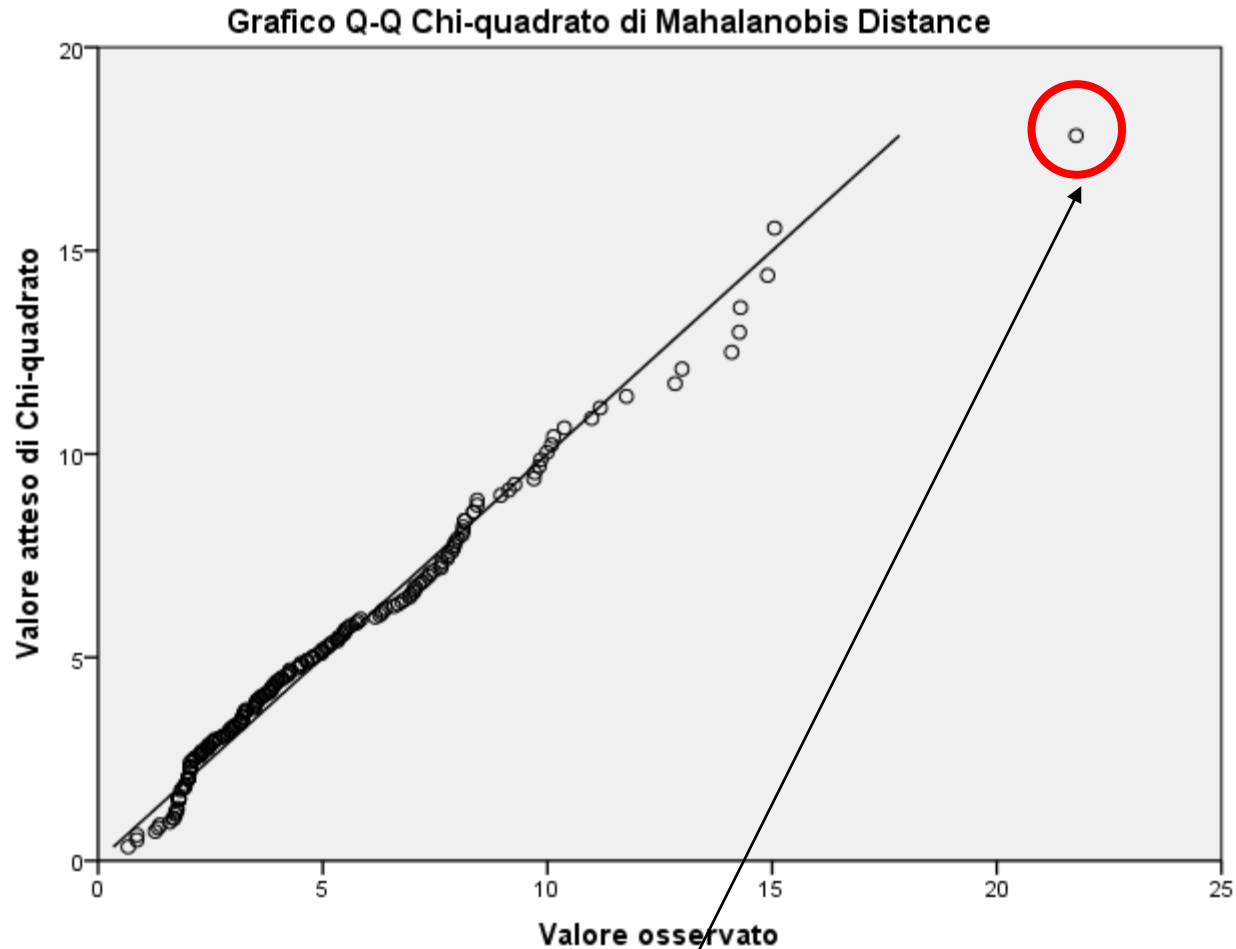


# Test grafico Q-Q Plot



**Specificare la distribuzione chi-quadrato con 5 gradi di libertà (ci sono 5 variabili)**

# Test grafico Q-Q Plot



**E' il possibile outlier multivariato**

# Calcolo del coefficiente di curtosi multivariata

**Formula:**

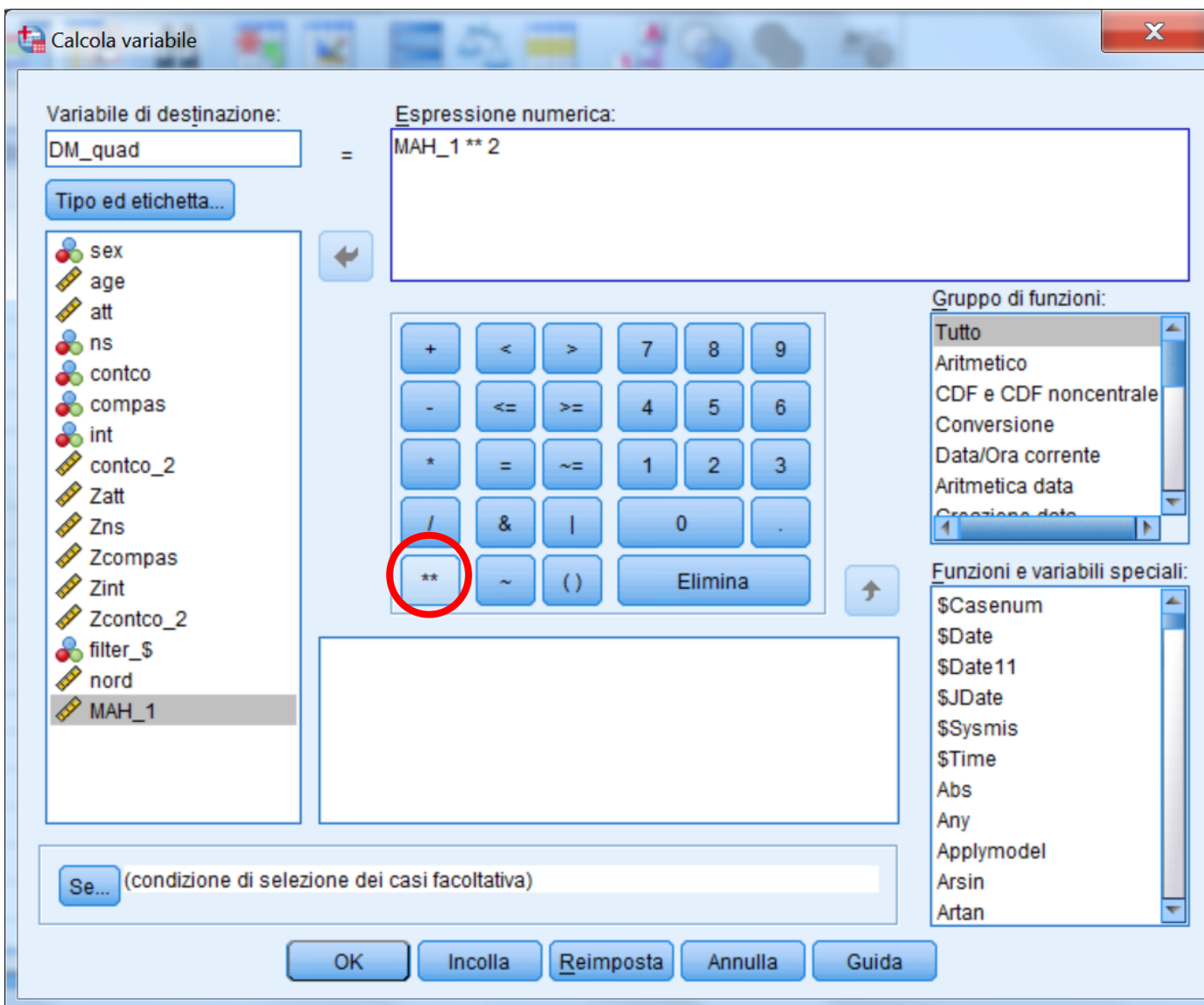
$$\sum_{i=1}^N (D_i^2)^2 / N$$

**Calcoliamo il denominatore**

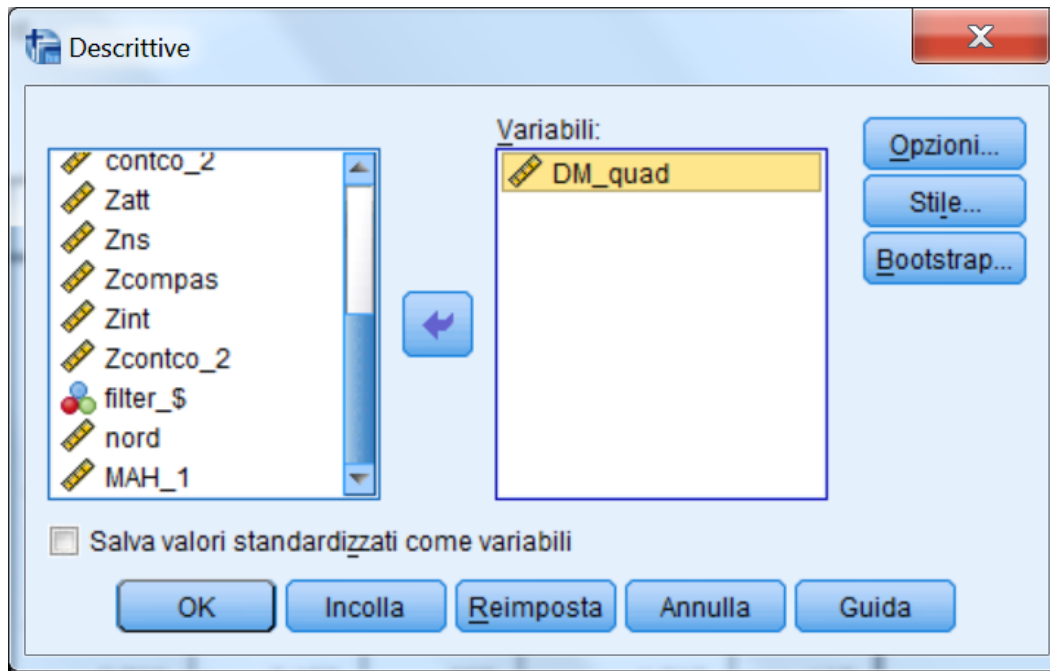
$$D_i^2 = MAH_1$$

**Per cui:**

$$(D_i^2)^2 = (MAH_1)^2$$



## Calcolo del coefficiente di curtosi multivariata



**Calcolando la media della variabile DM\_quad si ottiene il coefficiente di Mardia**

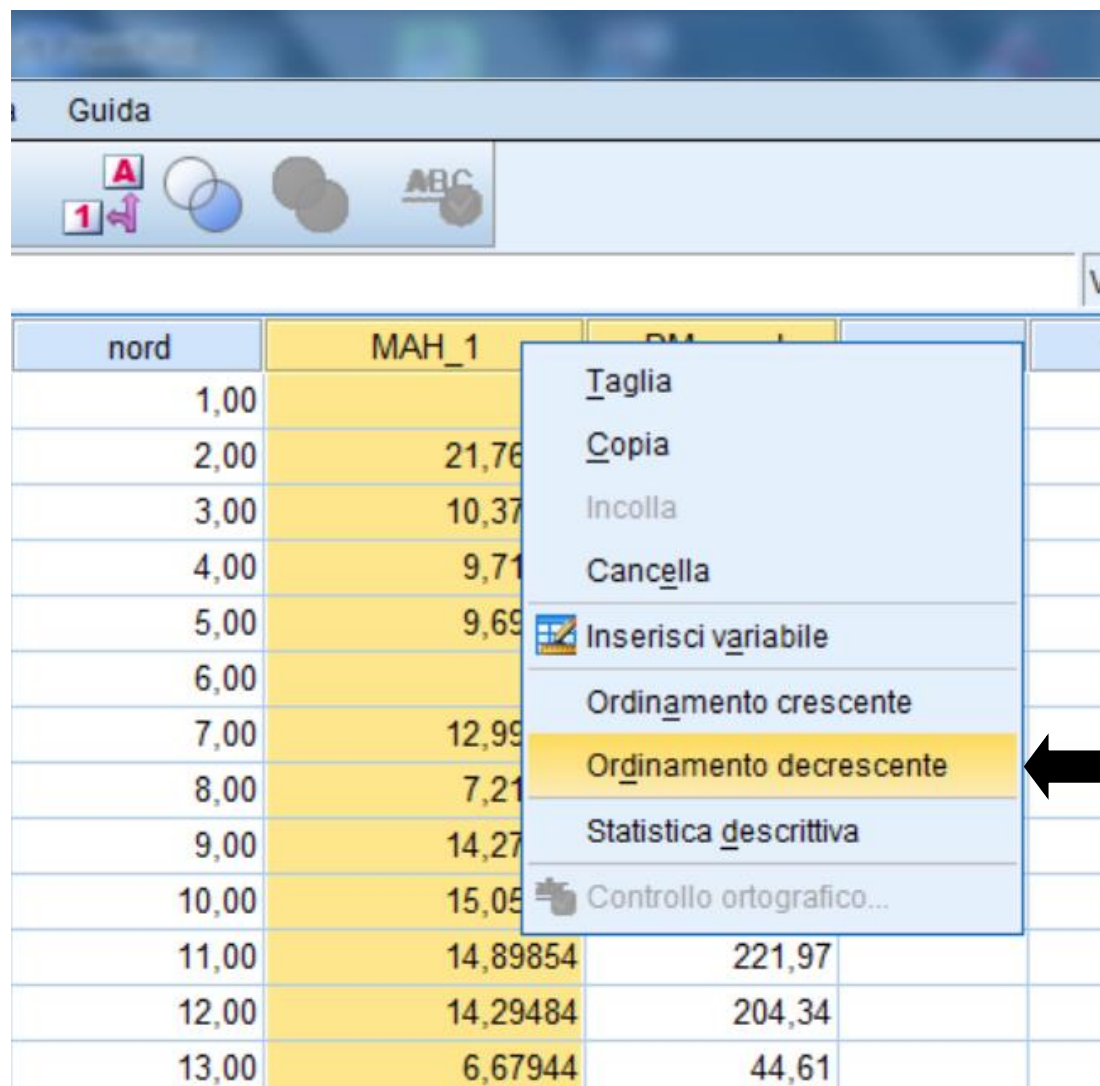
Statistiche descrittive

	N	Minimo	Massimo	Media	Deviazione std.
DM_quad	197	,45	473,72	35,7249	53,18476
Validi (listwise)	197				

**Il coefficiente è 35.72, di poco superiore a 35 (=5\*7), il valore critico con 5 variabili.**



# Esplorazione dei dati: individuare gli outlier multivariati



Guida

1 A

nord	MAH_1	PM
1,00		
2,00	21,76	
3,00	10,37	
4,00	9,71	
5,00	9,69	
6,00		
7,00	12,99	
8,00	7,21	
9,00	14,27	
10,00	15,05	
11,00	14,89854	221,97
12,00	14,29484	204,34
13,00	6,67944	44,61

- Taglia
- Copia
- Incolla
- Cancella
- Inserisci variabile
- Ordinamento crescente
- Ordinamento decrescente
- Statistica descrittiva
- Controllo ortografico...

# Esplorazione dei dati: la normalità multivariata e outliers multivariati



int	Zcontco_2	filter_\$	nord	MAH_1	DM_quad	var
-1,68712	2,21750	1	2,00	21,76511	473,72	
-2,08026	-,88476	1	10,00	15,05194	226,56	
,27855	,75423	1	11,00	14,89854	221,97	
-1,29399	,14933	1	12,00	14,29484	204,34	
-2,08026	,75423	1	9,00	14,27052	203,65	
,27855	-,88476	1	14,00	14,09949	198,80	
-1,29399	2,01829	1	7,00	12,99378	168,84	
-,50772	,75423	1	15,00	12,84113	164,89	

Vanno considerati come outliers multivariati i casi il cui valore risulta significativo al livello  $p < .001$ , considerando come distribuzione di riferimento quella del chi-quadrato con  $p$  gradi di libertà (dove  $p =$  numero di variabili). Con  $p = 5$  (abbiamo infatti 5 variabili) il livello di significatività del  $\chi^2$  è **20.51**, quindi c'è un possibile outlier multivariato.

# Filtrare i soggetti escludendo i due outliers uni- e l'outlier multi-variato

Seleziona casi: Se

Zatt > -3 & Zns > -3 & MAH\_1 < 20

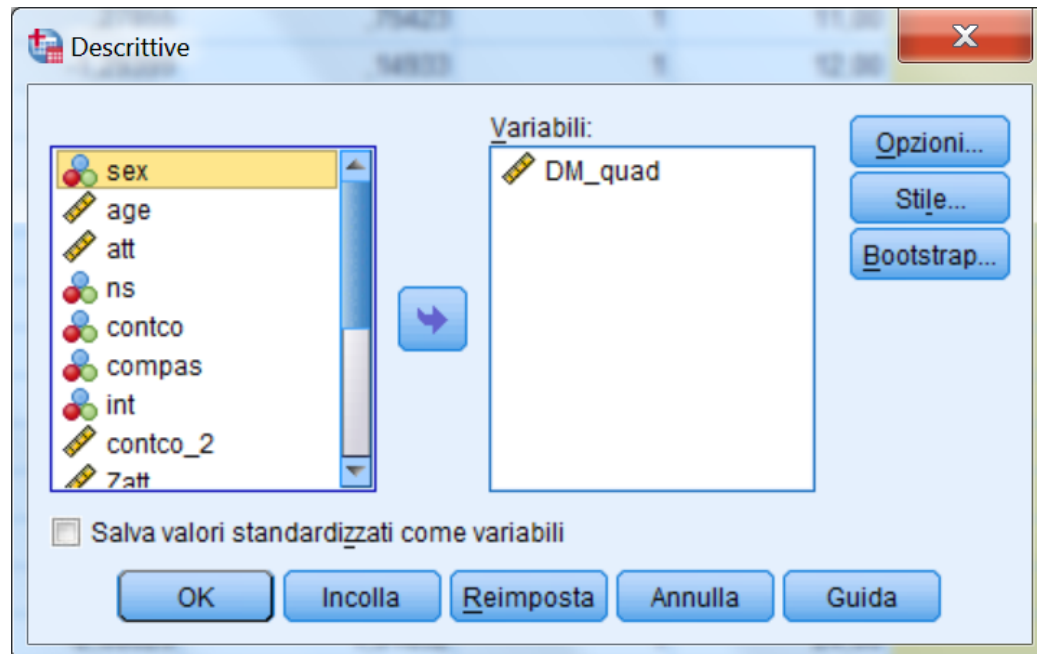
Gruppo di funzioni:

- Tutto
- Aritmetico
- CDF e CDF noncentrale
- Conversione
- Data/Ora corrente
- Aritmetica data
- Creazione data

Funzioni e variabili speciali:

Continua Annulla Guida

# Calcolo del coefficiente di curtosi multivariata



Statistiche descrittive

	N	Minimo	Massimo	Media	Deviazione std.
DM_quad	196	,45	226,56	33,4903	43,06171
Validi (listwise)	196				

**Il coefficiente è 33.49, ora inferiore a 35 ( $=5*7$ ).  
Ora i dati sono pronti per le analisi !!**

## **ESERCIZIO 1: TRATTAMENTI PRELIMINARI CON SPSS**

**Utilizzare i dati in formato testo nel file es1.xlsx**

**VARIABILI:**

**ATTEGGIAMENTO, NORME SOGGETTIVE, SENSO DI CONTROLLO,  
COMPORAMENTO PASSATO, INTENZIONE.**

**LA VARIABILE DIPENDENTE E' "INTENZIONE"**

**Verificare le caratteristiche distributive delle variabili, l'eventuale presenza di outlier, ed eventualmente trasformare le variabili non normali.**

**Salvare il file in formato .sav**