

Elementi di Psicometria

6-Altre tecniche per descrivere insiemi di dati
vers. 1.0c (5 dicembre 2011)
versione per stampa

Germano Rossi¹

`germano.rossi@unimib.it`

¹Dipartimento di Psicologia, Università di Milano-Bicocca

2011-2012

Descrivere dati

■ Riassunti numerici

■ Percentili

I vari percentili possono dare informazioni sull'andamento della distribuzione. In particolare i valori corrispondenti ai quartili e i percentili 5, 10, 90 e 95.

■ Riassunto a 5 numeri

■ Rappresentazioni grafiche

■ Diagramma a scatola

■ Mean-on-spoke

■ Scatterplot bi- e multi- variati

Riassunto a 5 numeri

- Sono 5 valori numerici: Mediana, Q1 e Q3, punteggio minimo e massimo di una distribuzione
- Vengono riportati in diversi modi

Esempio

Math Quiz			
Mediana	30		
Q1 e Q3	22		35
Min Max	9		49

Esempio

Math Quiz				
Min	Q1	Mdn	Q3	Max
9	22	30	35	49

- È più comoda la sua versione grafica: il diagramma a scatola

Diagramma a scatola e baffi

- Il diagramma a scatola e baffi (box-and-whiskers) è stato ideato da Tukey nell'ambito della EDA (*Exploratory data analysis*). È più spesso chiamato “box-plot”
- La scatola è formata dai valori corrispondenti al primo e al terzo quartile
- La linea spessa dentro la scatola corrisponde alla mediana
- I baffi rappresentano cose diverse in base ai software: come primo approccio useremo i valori minimo e massimo

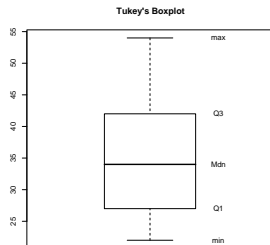
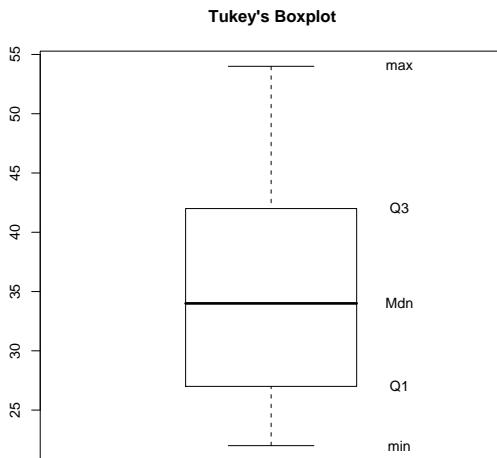


Grafico a scatola [Box-plot] (I/R)



variabile A ordinata: **22**

22 24 24 24 24 24 27

27 **27** 27 30 30 31 31

32 33 33 33 33 **34** 35

35 36 36 36 37 39 39

41 **42** 42 42 44 45 45

45 46 46 54 **54**

N=41; min=22;

Q1(10)=27;

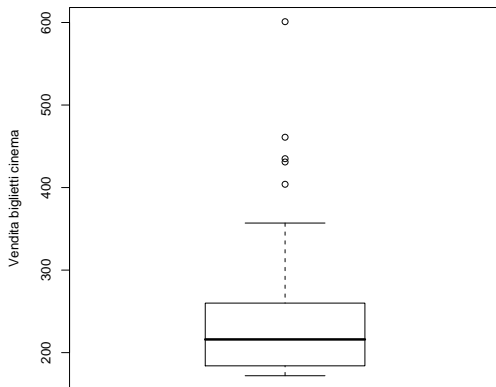
Mdn(21)=34;

Q3(31)=42; max=54

Grafico a scatola

- In realtà i box-plot di SPSS ed R non visualizzano i valori minimo e massimo
- Al loro posto viene usata la semi-differenza interquartilica moltiplicata per 1,5; in una distribuzione normale, questo valore è quasi sempre oltre il massimo e oltre il minimo (quindi si visualizzano max e min)
- inoltre visualizzano i singoli valori anomali (i valori oltre i baffi), evidenziando quindi le code asimmetriche
- L'utilità dei box-plot è più evidente se si incrociano con una variabile categoriale, perché si possono fare confronti sulle distribuzioni dei sotto-campioni

Grafico a scatola (I/R)

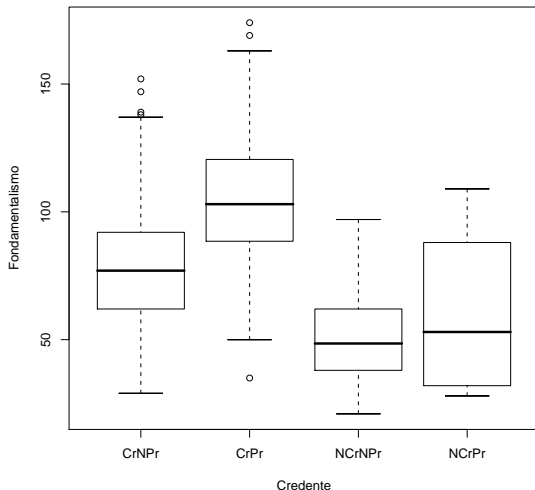


variabile Gross ordinata: **172**

173 173 174 175 176 176 177
177 178 180 180 181 181 181
181 183 183 183 **184** 184 184
185 188 190 191 192 194 197
199 199 202 202 205 206 207
211 213 215 **216** 217 218 219
226 228 229 233 234 235 239
241 242 242 246 250 251 256
257 260 **260** 262 268 286 290
294 306 309 311 313 318 329
330 339 357 404 431 435 461
601

SIQR: $(260 - 184)/2 * 1.5 = 57$
baffi: 127 e 317

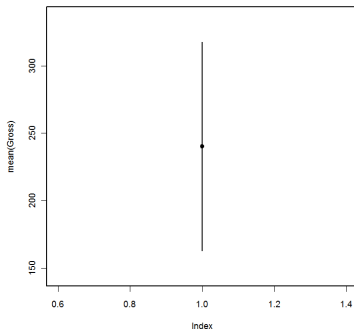
Grafico a scatola (I/R) suddiviso (N/O)



- Per ogni valore della variabile di raggruppamento, viene prodotto un box-plot
- In questo modo si possono vedere le differenze di distribuzione

Mean-on-spoke

- Questo tipo di grafico (chiamato anche medie con errori) rappresenta le medie e le deviazioni standard in modo grafico
- La media è rappresentata da un punto
- Le dev. standard dai “raggi” che partono dal punto (la media)
- sul grafico si possono aggiungere altri valori (valori corrispondenti alla media, ...)



Mean-on-spoke

- Anche in questo caso, è più utile per confrontare fra loro più variabili

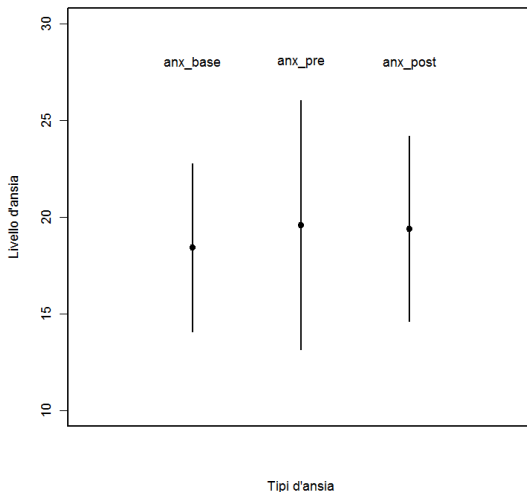
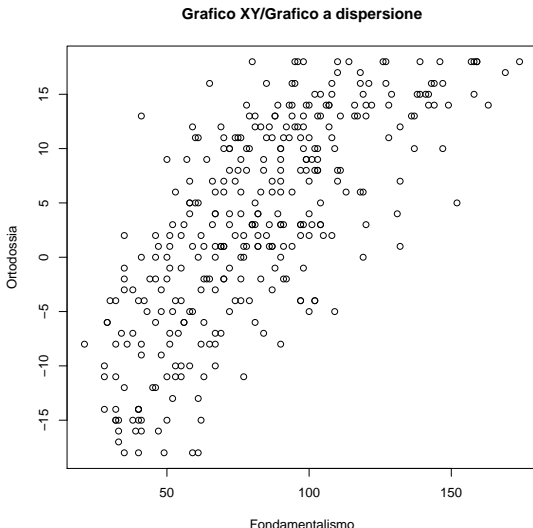
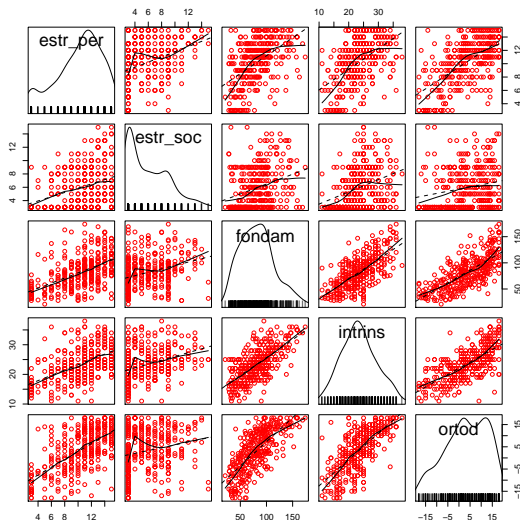


Grafico bivariato (I/R)



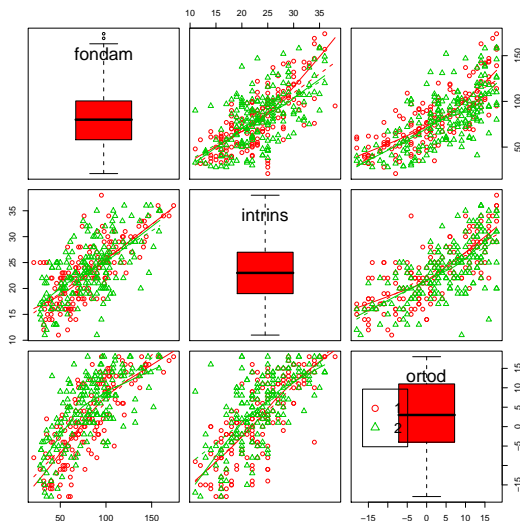
- Utilizza il principio del piano cartesiano
- Ogni punto rappresenta la contingenza di un valore su una variabile e il valore sull'altra per un determinato soggetto
- Può aiutare a capire se due variabile hanno un andamento lineare simile o fra loro proporzionale

Grafico multivariato (I/R)



- Lo stesso di prima, ma con più variabili alla volta
- Ogni colonna utilizza la variabile indicata come asse x
- Ogni riga utilizza la variabile indicata come asse y
- Lungo la diagonale, informazioni riassuntive

Grafico multivariato (I/R) suddiviso (N/O)



- Lo stesso di prima, ma con una variabile di raggruppamento (in questo caso il genere), rappresentata dai colori
- Lungo la diagonale, altre informazioni riassuntive (in questo caso, il box-plot della variabile)

Spss: Riassunto a 5 numeri (Frequenze)

- Spss non produce un riassunto a 5 numeri in modo diretto
- Dovete usare **Analizza | Statistiche descrittive | Frequenze...** e dal pulsante **Statistiche...** scegliere i quartili e il massimo e il minimo

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.
- ☐ Varianza
- ☐ Intervallo
- ☒ Minimo
- ☒ Massimo
- ☐ Errore standard della media

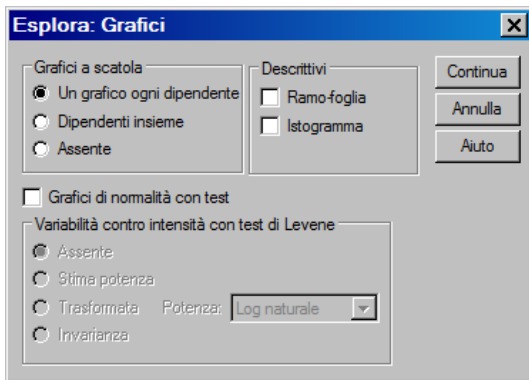
Distribuzione

- ☐ Asimmetria
- ☐ Curtosi

Continua
Annulla
Aiuto

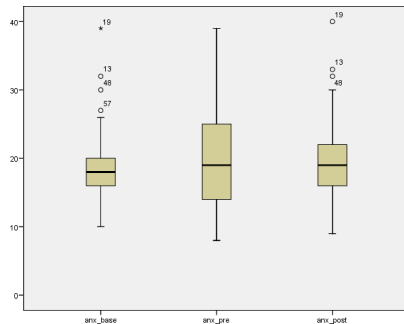
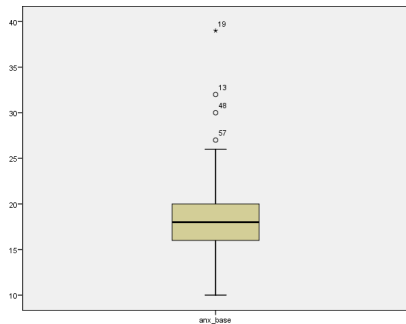
Spss: Box-plot (Esplora)

- Spss produce i box-plot tramite `Analizza | Statistiche descrittive | Esplora...` e dal pulsante `Grafici...` assicuratevi di aver attivato una delle prime due opzioni di **Grafici a scatola**



- Se avete selezionato più variabili, “Un grafico ogni dipendente” produce grafici separati
- “Dipendenti insieme” produce un unico grafico

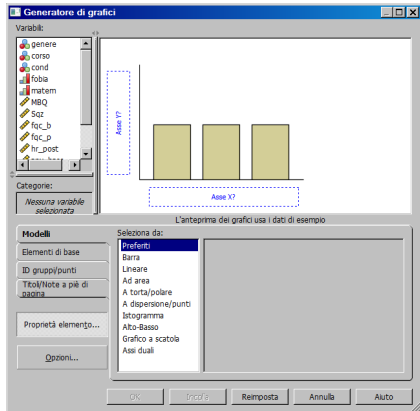
Spss: Box-plot



■ Esempio di box-plot di Spss

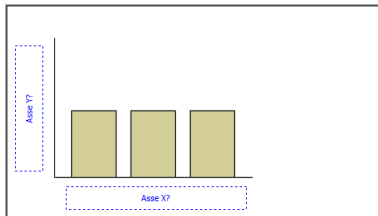
Spss: Mean-on-spoke

- Spss non fa direttamente questi grafici e bisogna usare una procedura complicata
- Clicchiamo su **Grafici | Generatore di grafici...**
- Compare una finestra di avviso. Se avete già assegnato la scala alle variabili premete **OK**
- Altrimenti scegliete **Definisci proprietà variabili...**
- Arriverete ad una nuova schermata dove costruiremo il grafico che ci serve




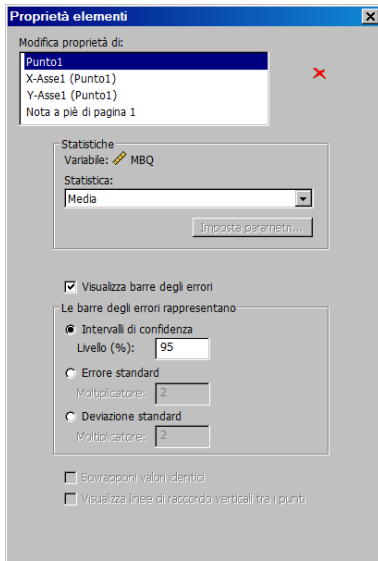
Spss: Mean-on-spoke

- In **Modelli**, fate Click su **Barra**
- e quindi sul terzo disegno della seconda riga
- trascinatelo nell'anteprima
- scegliete una variabile intervallo e trascinatela sopra a *AsseY?*
- Mettete una variabile categoriale in *AsseX?*
- La categoriale dev'essere una costante se non volete un grafico suddiviso in sottocampioni

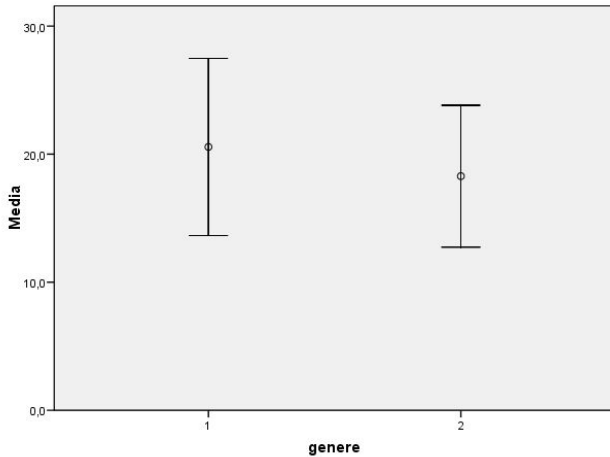


Spss: Mean-on-spoke

- Nel pannello laterale che si è aperto, cercate l'area Le barre degli errori rappresentano
- **selezionate** Deviazione standard
- e impostate il loro valore a 1
- Infine cliccate su 



Spss: Mean-on-spoke



Barre degli errori: +/- 1 SD