Tecniche statistiche di analisi del cambiamento

08-Analisi non parametrica
(vers. 1.2, 18 dicembre 2016)

Germano Rossi[1]
germano.rossi@unimib.it

[1]Dipartimento di Psicologia, Università di Milano-Bicocca

2016-17

Introduzione

- I test non parametrici si usano quando:
- le variabili sono qualitative (in particolare frequenze)
- oppure ci sono ragioni per ritenere che (in una variabile quantitativa) gli assunti di normalità siano violati
  - quando il campione è piccolo (N<30)
  - quando non si distribuisce normalmente e non si riesce/può normalizzare
- I test non parametrici sono “liberi da assunti di distribuzione”
- o perché la distribuzione è costruita in modo diretto
- o la tecnica non assume una distribuzione particolare
- o perché i dati vengono trasformati in ranghi

1 campione/1 variabile

Gruppo	Variabile
Caso A	X_A
B	X_B
C	X_C
D	X_D

- Test di chi-quadro (χ^2)[1]: `Analizza | Test non parametrici | Finestre legacy | Chi quadrato`
- Test logaritmico di chi-quadro ($\mathbf{G}^2$)[1,2]
- Test binomiale[1]: `Analizza | Test non parametrici | Finestre legacy | Binomiale`

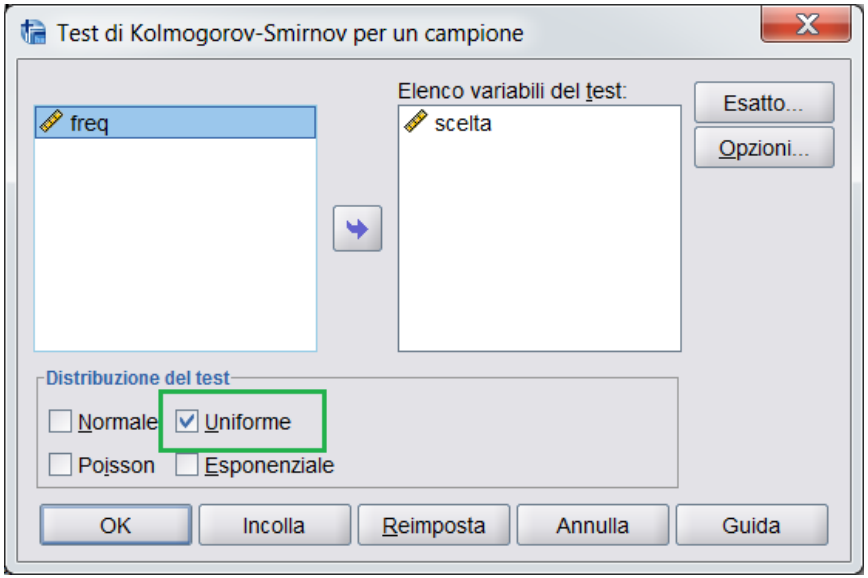
- Test delle successioni (*runs*) per variabili dicotomiche: `Analizza | Test non parametrici | Finestre legacy | Successioni` anche `Esecuzioni`
- Test di Kolmogorov-Smirnov per un campione: `Analizza | Test non parametrici | Finestre legacy | K-S per 1 campione`

[1] Già fatto

[2] Non si può calcolare in SPSS

Test di Kolmogorov-Smirnov per un campione

Tempo scelto per consegnare i compiti						
Scelta	settimane					
studenti	10	8	6	4	2	
freq	0	2	10	15	3	n=30



- Nel test per la normalità, assume distribuzione normale

- il test confronta la distribuzione cumulata osservata con la distribuzione cumulata teorica
- usando la distribuzione Uniforme o Normale o Poisson o Esponenziale
- Viene calcolato un punto z sugli scarti
- Almeno al livello ordinale

Test di Kolmogorov-Smirnov per un campione

		scelta
N		30
Parametri uniformi[a,b]	Minimo	2
	Massimo	5
Differenze più estreme	Assoluta	,267
	Positiva	,233
	Negativa	-,267
Z di Kolmogorov-Smirnov		1,461
Sign. asintotica (a due code)		,028

a. La distribuzione di test è uniforme.

b. Calcolato dai dati.

Test delle successioni

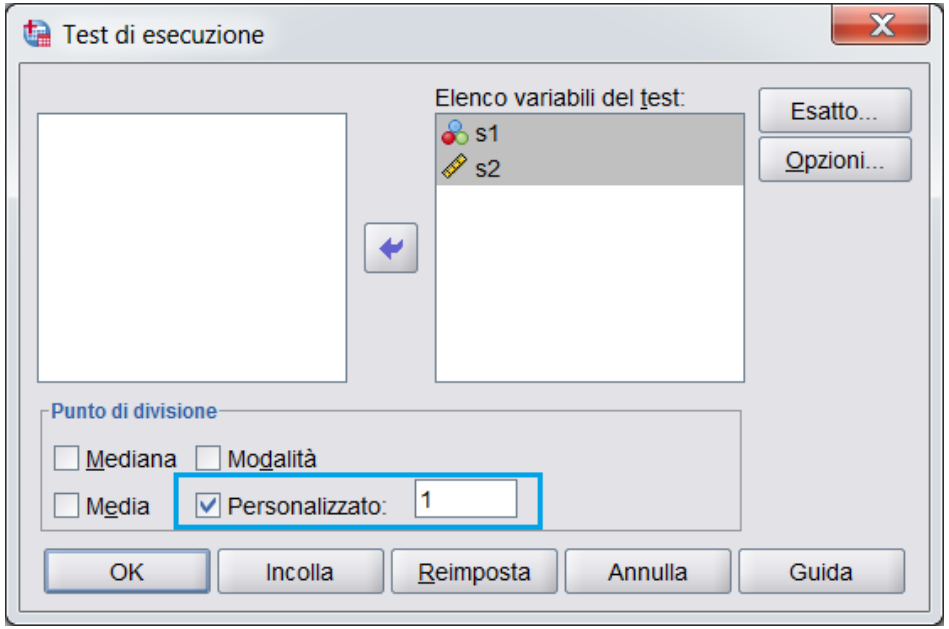
serie A	0 **1 1 1** 0 0 **1** 0 0 **1 1 1** 0 **1** 0 0 0 **1** 0 **1**
serie B	0 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 0 0 0

- si basa sul numero di **successioni** (o comparse) delle due categorie
- una “successione” è composta da simboli uguali
- con N=20, 2 successioni indicherebbero 2 raggruppamenti (tutti 0 poi tutti 1)
- 10 successioni che ogni 0 è seguito da un 1 (o viceversa)
- Il risultato è un punto z

- Nella Serie A abbiamo 6 successioni per 0 e 6 per 1
- Nella Serie B, 3 successioni per 0 e 2 per 1
- In SPSS bisogna sempre indicare un punto di divisione:
  - Può essere Media, Mediana , Modalità (Moda) o personalizzato

Test delle successioni

serie A	0 1 1 1 0 0 1 0 0 1 1 1 0 1 0 0 0 1 0 1
serie B	0 0 0 0 1 1 1 1 1 1 0 0 0 1 1 1 1 0 0 0



Test di esecuzione

	s1	s2
Valore di test[a]	1,00	1,00
Numero totale casi	20	20
Numero di esecuzioni	12	5
Z	,230	-2,527
Sign. asintotica (a due code)	,818	,012

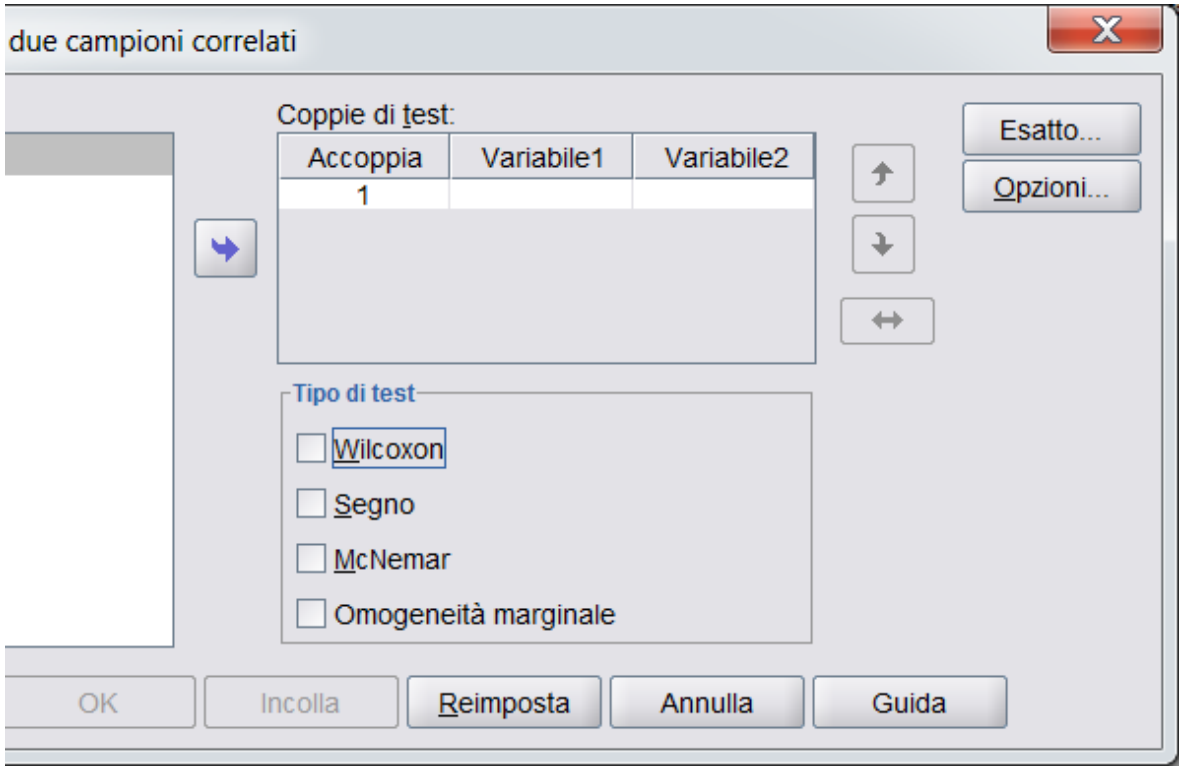
a. Specificato dall'utente.

- "`esecuzioni`" = successioni
- La serie A (s1) è casuale
- La serie B (s2), no

2 campioni dipendenti/appaiati

`Analizza | Test non parametrici | Finestre legacy | 2 campioni correlati`

Gruppo	Condizione 1	Condizione 2
Caso A	X_{A1}	X_{A2}
B	X_{B1}	X_{B2}
C	X_{C1}	X_{C2}
D	X_{D1}	X_{D2}



- Test del Segno
- Test di Wilcoxon
- Test di McNemar
- Test dell'Omogeneità marginale

Test del segno

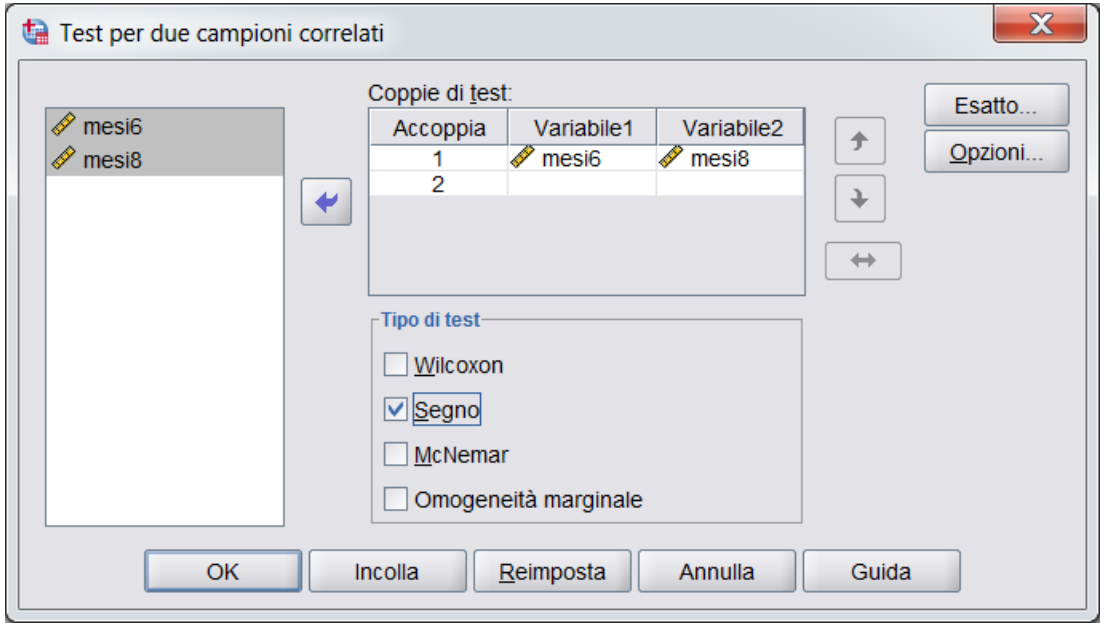
- È un test che si applica usando variabili a livello minimale (ordinale, nominale o intervallo non normale)
- In pratica si confrontano i due valori ($X_1 - X_2$) assunti dalle due variabili per un dato caso statistico
- Se i valori sono uguali ($X_1 = X_2$), si ignorano
- Si contano gli aumenti ($X_1 < X_2$) o le diminuzioni ($X_1 > X_2$)

Guarda la madre (1 min)			
Caso	6 mesi	9 mesi	Diff
A	3	7	+
B	5	6	+
C	5	3	-
D	4	8	+
E	3	5	+
F	7	9	+
G	8	7	-
H	7	9	+

- Si applica il test binomiale su aumenti o diminuzioni
- SPSS confronta i singoli valori numerici
- Ma poi crea una dicotomizzazione, considerando **solo i segni**
- La domanda è se 6 aumenti su 8 casi sono “casuali” o no!

Test del Segno: SPSS

Guarda la madre (1 min)		
Caso	6 mesi	9 mesi
A	3	7
B	5	6
C	5	3
D	4	8
E	3	5
F	7	9
G	8	7
H	7	9



Statistiche del test[a]

	mesi8 - mesi6
Sign. esatta (a due code)	,289[b]

a. Test dei segni

b. È utilizzata la distribuzione binomiale.

Frequenze

		N
mesi8 - mesi6	Differenze negative[a]	2
	Differenze positive[b]	6
	Correlazioni[c]	0
	Totale	8

a. mesi8 < mesi6

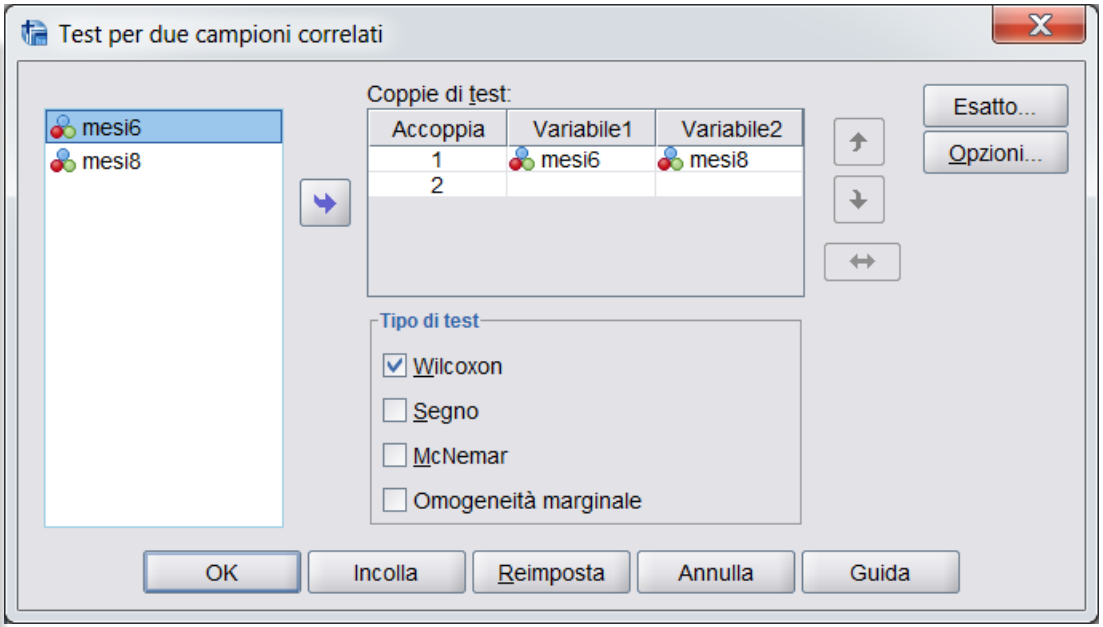
b. mesi8 > mesi6

c. mesi8 = mesi6

Test di Wilcoxon

- È l'alternativa classica al T-test per campioni dipendenti, appaiati, misure ripetute
- Il procedimento è simile al T-Test, ma si lavora sui ranghi dei punteggi
- Il risultato è espresso come punto z
- Si differenzia dal test del segno perché con i ranghi si tiene in considerazione l'ampiezza del cambiamento

Guarda la madre (1 min)			
Caso	6 mesi	9 mesi	Diff
A	3	7	+4
B	5	6	+1
C	5	3	-2
D	4	8	+4
E	3	5	+2
F	7	9	+2
G	8	7	-1
H	7	9	+2



Test di Wilcoxon: SPSS

Preparazione			Calcolo ranghi			Assegnazione		Calcolo			
Diff		Assol	Ord	Pos	Rango	Diff	Rango		Rango	SUM	Media
4	+	4	1	1	1,5	4	7,5	-	4,5	6	3
1	+	1	1	2	1,5	1	1,5		1,5		
-2	-	2	2	3	4,5	-2	4,5	+	7,5	30	5
4	+	4	2	4	4,5	4	7,5		1,5		
2	+	2	2	5	4,5	2	4,5		7,5		
2	+	2	2	6	4,5	2	4,5		4,5		
-1	-	1	4	7	7,5	-1	1,5		4,5		
2	+	2	4	8	7,5	2	4,5		4,5		

Ranghi

		N	Rango della media	Somma dei ranghi
mesi9 - mesi6	Ranghi negativi	2[a]	3,00	6,00
	Ranghi positivi	6[b]	5,00	30,00
	Correlazioni	0[c]		
	Totale	8		

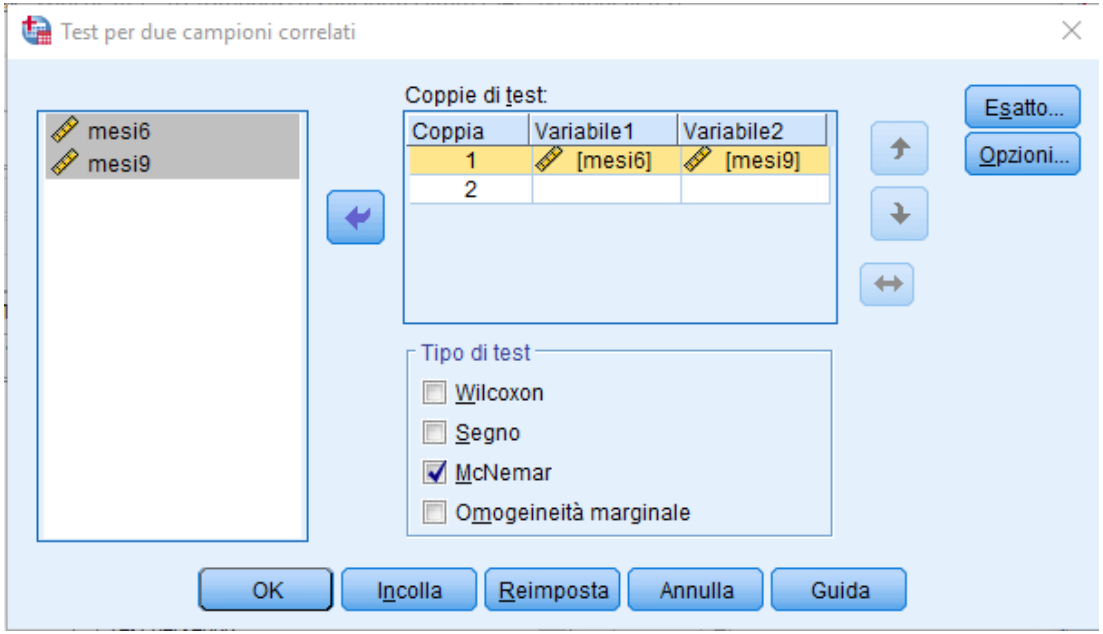
Statistiche del test[a]

	mesi9 - mesi6
Z	-1,706[b]
Sign. asint. (a due code)	,088

a. Test dei ranghi con segno di Wilcoxon

b. Basato su ranghi negativi.

Test di McNemar



mesi6x & mesi9x

mesi6x	mesi9x ,00	mesi9x 1,00
,00	2	3
1,00	0	3

Statistiche del test[a]

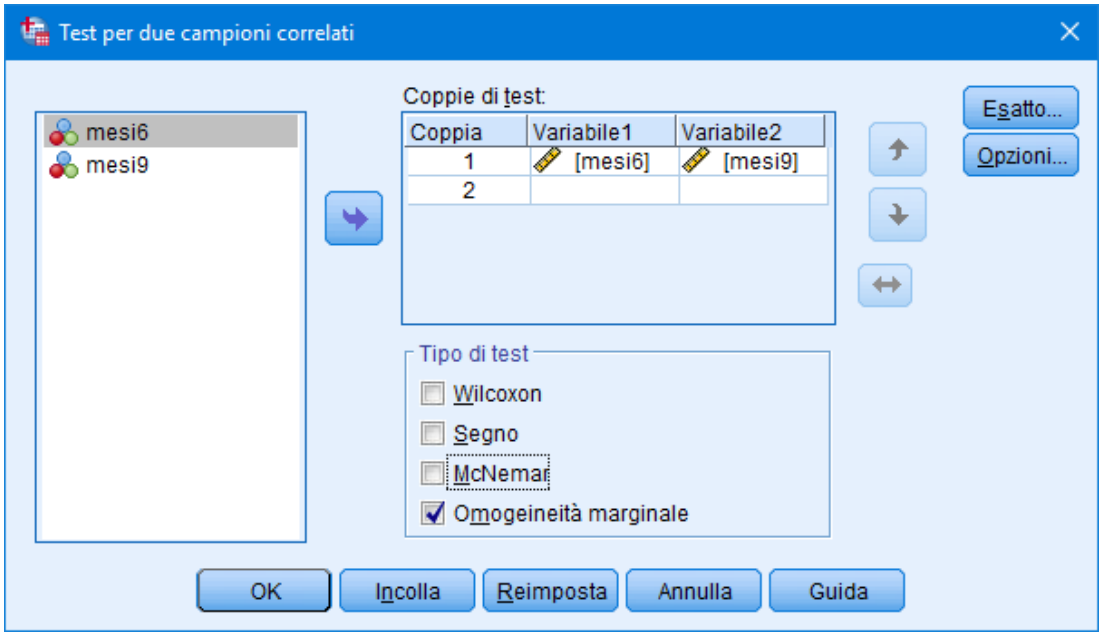
	mesi6x & mesi9x
N	8
Sign. esatta (a due code)	,250[b]

a. Test McNemar

b. Distribuzione binomiale utilizzata.

- Lavora con variabili **dicotomiche** (0, 1)
- Viene creata una tabella 2x2
- la formula è $\chi^2 = \frac{(b-c)^2}{(b+c)}$ dove b e c sono le celle del cambiamento (1-0 e 0-1)
- L'ipotesi H_0 è che le frequenze marginali sono uguali
- Se χ^2 è sig. -> c'è cambiamento

Test dell'Omogeneità marginale



- È il test corrispondente a McNemar per variabili qualitative non dicotomiche

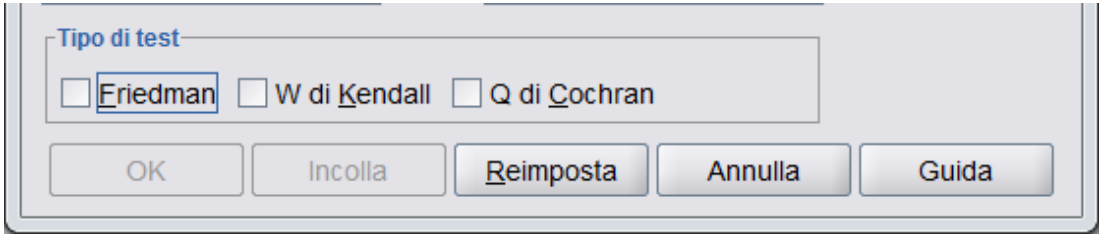
Test di omogeneità marginale

	mesi6 & mesi9
Valori distinti	7
Casi esterni alla diagonale	8
Statistica MH osservata	42,000
Media statistica MH	48,000
Deviazione std. della statistica MH	3,536
Statistica MH std.	-1,697
Sign. asint. (a due code)	,090

k campioni dipendenti/appaiati

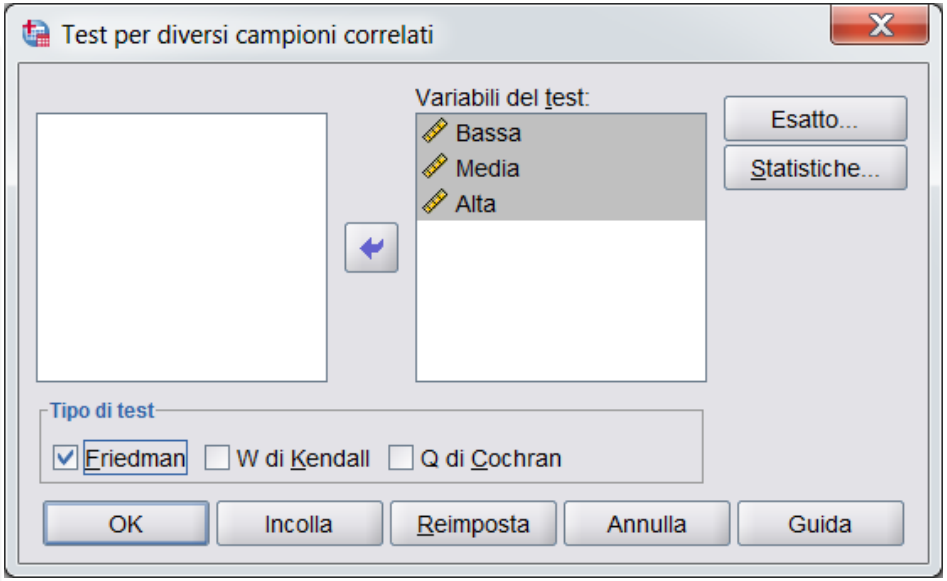
Gruppo	Condizione 1	2	3	4
Caso A	X_{A1}	X_{A2}	X_{A3}	X_{A4}
B	X_{B1}	X_{B2}	X_{A3}	X_{A4}
C	X_{C1}	X_{C2}	X_{A3}	X_{A4}
D	X_{D1}	X_{D2}	X_{A3}	X_{A4}

- Test di Friedman
- W di Kendall
- Q di Cochran



Test di Friedman

	Distrazione		
caso	Bassa	Media	Alta
A	9	6	7
B	15	7	2
C	12	9	5
D	16	8	2
E	22	15	6
F	8	3	4



Classificazioni

	Classificazione della media
Bassa	3,00
Media	1,67
Alta	1,33

Statistiche del test[a]

N	6
Chi-quadrato	9,333
gl	2
Sign. asintotica	,009

a. Test di Friedman

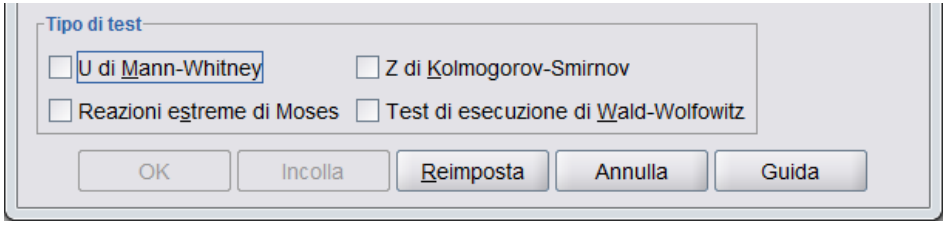
- I dati sono trasformati in ranghi all'interno di ogni variabile
- Vengono confrontate le medie dei ranghi (H_0 sono uguali)
- Il valore calcolato è un chi-quadro con gl=numero variabili -1

2 campioni indipendenti

`Analizza | Test non parametrici | Finestre legacy | 2 campioni indipendenti`

Caso	Gruppo	Var 1
A	1	X_A
B	1	X_B
C	1	X_C
D	1	X_D
E	2	X_E
F	2	X_F
G	2	X_G
H	2	X_H

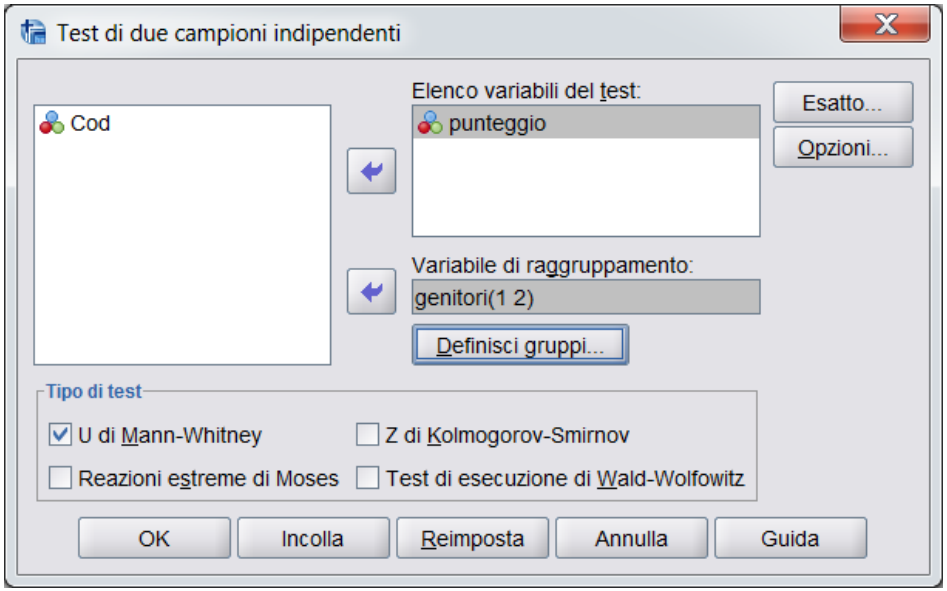
- U di Man-Whitney
- reazioni estreme di Moses
- Z di Kolmogorov-Smirnov
- test delle successioni di Wald-Wolfowitz



U di Man-Whitney

	Punteggi emotivi in base alla famiglia											
2 genitori	12	18	14	10	19	8	15	11	10	13	15	16
1 genitore	6	9	4	13	14	9	8	12	11	9		

- È il corrispettivo del T-test per campioni indipendenti
- Usando i ranghi calcolati sull'intero campione
- E verificando se la media dei ranghi differisce sensibilmente
- Alla differenza della somma si applicano formule diverse per piccoli campioni ($N \leq 20$) o grandi campioni ($N > 20$)
- Per l'interpretazione entrambi gli indici si trasformano in punto z



U di Man-Whitney: SPSS

Classificazioni

	genitori	N	Classificazione della media	Somma delle classificazioni
punteggio	1	10	7,85	78,50
	2	12	14,54	174,50
	Totale	22		

Statistiche del test[a]

	punteggio
U di Mann-Whitney	23,500
W di Wilcoxon	78,500
Z	-2,414
Sign. asintotica (a due code)	,016
Sign. esatta [2*(sig. a una coda)]	,014[b]

a. Variabile di raggruppamento: genitori

b. Non corretto per le correlazioni.

k campioni indipendenti

Caso	Gruppo	Var 1
A	1	X_A
B	1	X_B
C	1	X_C
D	2	X_D
E	2	X_E
F	2	X_F
G	3	X_G
H	3	X_H
I	3	X_I

- Il test H di Kruskal-Wallis
- Jonckheere-Terpstra
- Mediana

Il test H di Kruskal-Wallis

		Punteggi dei partecipanti									
Motivazione	alta	17	14	19	16	18	20	23	21	18	
	media	10	11	8	12	9	11	8	12	9	10
	bassa	3	9	2	5	1	7	6			

- Sostituto dell'Anova a 1 fattore
- permette di confrontare la posizione di 2 o più campioni indipendenti
- i dati sono trasformati in ranghi ignorando i gruppi

Classificazioni

	motivazione	N	Classificazione della media
punteggio	1 Bassa	9	22,00
	2 Media	10	12,20
	3 Alta	7	4,43
	Totale	26	

Statistiche del test[a,b]

	punteggio
Chi-quadrato	21,317
gl	2
Sign. asintotica	,000

a. Test di Kruskal Wallis

b. Variabile di raggruppamento: motivazione