

* Test non parametrici basati sui ranghi

Quando vanno applicati

- I test parametrici (z-test, t-test e analisi di varianza) sono basati sul presupposto che le osservazioni siano tratte da popolazioni di valori normalmente distribuiti, nelle quali le varianze sono approssimativamente uguali.
- Questi metodi sono detti parametrici perché si basano sulla stima dei due parametri della popolazione, la media e la varianza, che definiscono completamente una distribuzione normale.
- Se i dati sperimentali non sono compatibili con queste condizioni preliminari o se addirittura, le osservazioni sono misurate su una scala qualitativa ordinale, i metodi parametrici diventano poco attendibili poiché la media e la varianza non sono sufficienti per una completa descrizione della popolazione.
- In questi casi è possibile utilizzare, invece delle osservazioni, i ranghi, cioè il numero d'ordine delle osservazioni stesse, al fine di calcolare test non parametrici (ovvero test liberi da distribuzione) nella verifica delle ipotesi.

* Test non parametrici basati sui ranghi

Il test per la somma dei ranghi di Mann-Whitney

■ Analogamente al T-test di Student per dati indipendenti, questo test è utilizzato in esperimenti in cui si confrontano due campioni indipendenti, ma la variabile in studio non rispetta le condizioni di applicabilità dei test parametrici.

Illustriamo il procedimento utilizzando un piccolo esperimento effettuato su 7 pazienti con depressione maggiore che mira a valutare l'efficacia di un farmaco antidepressivo. Il livello di gravità è valutato calcolando il punteggio totale della Scala HAM-D dopo un ciclo di un mese in cui a 3 soggetti è stato somministrato un placebo e a 4 il farmaco.

I ranghi sono attribuiti indipendentemente dal gruppo di appartenenza: al punteggio più basso è attribuito rango 1, al più elevato rango 7.

Osservazioni relative ad un esperimento su un farmaco antidepressivo

Placebo (controllo)		Farmaco (trattamento)	
Punteggio Scala HAM-D	Rango	Punteggio Scala HAM-D	Rango
19	3	21	4
24	5	27	7
16	1	18	2
	<hr/>	25	6
	T=9		

* Test non parametrici basati sui ranghi

Il test di Mann-Whitney, considerazioni.

- Se il farmaco avesse diminuito i punteggi, dovremmo aspettarci i ranghi più bassi nel gruppo in trattamento; viceversa se gli avesse aumentati, dovremmo aspettarci i ranghi più bassi nel gruppo di controllo (placebo).
- Come test statistico possiamo utilizzare la somma T dei ranghi del gruppo più piccolo (nell' esempio il gruppo di controllo).
- Nel nostro esempio, il valore $T=9$ è sufficientemente “estremo” da giustificare il rifiuto dell' ipotesi nulla che il farmaco non abbia effetto? Ovvero, prendendo in considerazione la distribuzione di tutti i possibili valori di T , quanto è probabile osservare una somma T altrettanto “estrema” di quella osservata?
- Per stimare questa probabilità potremmo elencare tutti i possibili ranghi attribuibili a ciascuna osservazione calcolando per ogni combinazione la somma dei ranghi del gruppo di controllo.

* Test non parametrici basati sui ranghi

Il test di Mann-Whitney. Possibili attribuzioni di ranghi

Possibili ranghi e somme di ranghi per 3 individui su 7

rango							Somma dei ranghi
1	2	3	4	5	6	7	
x	x	x					6
x	x		x				7
x	x			x			8
x	x				x		9
x	x					x	10
x		x	x				8
x		x		x			9
x		x			x		10
x		x				x	11
x			x	x			10
x			x		x		11
x			x			x	12
x				x	x		12
x				x		x	13
x					x	x	14
	x	x	x				9
	x	x		x			10
	x	x			x		11
	x	x				x	12
	x		x	x			11
	x		x		x		12
	x		x			x	13
	x			x	x		13
	x			x		x	14
	x				x	x	15
		x	x	x			12
		x	x		x		13
		x	x			x	14
		x		x	x		14
		x		x		x	15
		x			x	x	16
			x	x	x		15
			x	x		x	16
				x	x	x	17
				x	x	x	18

Nel nostro esempio, il numero totale di modi diversi con cui i ranghi possono essere attribuiti al gruppo placebo corrisponde alle combinazioni di 7 oggetti a tre a tre, ovvero:

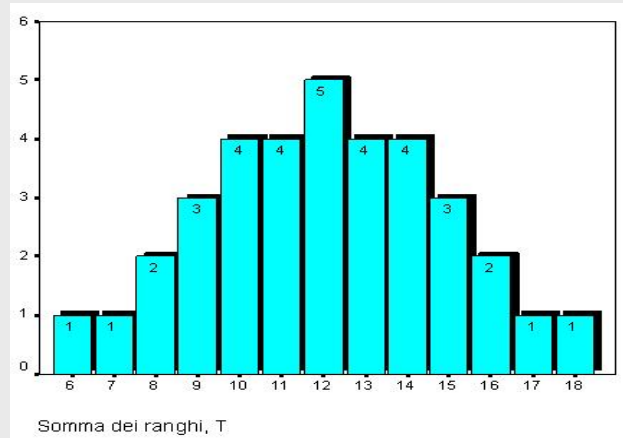
$${}^7C_3 = \frac{7!}{3! \times 4!} = 35$$

La tabella a lato mostra i 35 modi; le crocette identificano i ranghi associabili ai tre soggetti del gruppo placebo.

La colonna sulla destra indica la somma dei ranghi attribuibile al gruppo di controllo per ogni possibile combinazione di ranghi.

* Test non parametrici basati sui ranghi

Il test di Mann-Whitney. Distribuzione di T.



- La figura a lato mostra la distribuzione dei possibili valori di T per ampiezze campionarie pari a 3 e 4.

- L'insieme dei due valori più estremi (T=6 e T=18) costituisce il 5,7% ($p=0,057$) di tutti i possibili modi di combinare i ranghi.

- Possiamo quindi definire questi valori come valori soglia (valori critici) per individuare i valori “estremi” di rifiuto dell'ipotesi nulla di inefficacia del trattamento ($\alpha=0,057$)*.

- Dal momento che nel nostro esempio abbiamo ottenuto T=9, possiamo concludere che il valore non è sufficientemente “estremo” da giustificare il rifiuto dell'ipotesi che il farmaco non abbia effetto sulla depressione.

* A differenza dei test parametrici in cui la distribuzione dei possibili valori della statistica è una distribuzione continua, la distribuzione dei valori T è discreta e quindi non è possibile considerare valori soglia corrispondenti esattamente al 5% dei valori più estremi (livello di significatività $\alpha=0,05$), ma solo quelli che più gli si avvicinano.

* Test non parametrici basati sui ranghi

Il test di Mann-Whitney. Valori critici.

n_1, n_2	Livello di significatività a due code			n_1, n_2	Livello di significatività a due code		
	0.05	0.01	0.001		0.05	0.01	0.001
2, 8	3, 19			4, 9	15, 41	11, 45	
2, 9	3, 21			4, 10	15, 45	12, 48	
2, 10	3, 23			4, 11	16, 48	12, 52	
2, 11	4, 24			4, 12	17, 51	13, 55	
2, 12	4, 26			4, 13	18, 54	14, 58	10, 62
2, 13	4, 28			4, 14	19, 57	14, 62	10, 66
2, 14	4, 30			4, 15	20, 60	15, 65	10, 70
2, 15	4, 32			4, 16	21, 63	15, 69	11, 73
2, 16	4, 34			4, 17	21, 67	16, 72	11, 77
2, 17	5, 35			4, 18	22, 70	16, 76	11, 81
2, 18	5, 37			4, 19	23, 73	17, 79	12, 84
2, 19	5, 39	3, 41		4, 20	24, 76	18, 82	12, 88
2, 20	5, 41	3, 43		4, 21	25, 79	18, 86	12, 92
2, 21	6, 42	3, 45		4, 22	26, 82	19, 89	13, 95
2, 22	6, 44	3, 47		4, 23	27, 85	19, 93	13, 99
2, 23	6, 46	3, 49		4, 24	28, 88	20, 96	13, 103
2, 24	6, 48	3, 51		4, 25	28, 92	20, 100	14, 106
2, 25	6, 50	3, 53					
				5, 5	17, 38	15, 40	
3, 5	6, 21			5, 6	18, 42	16, 44	
3, 6	7, 23			5, 7	20, 45	17, 48	
3, 7	7, 26			5, 8	21, 49	17, 53	
3, 8	8, 28			5, 9	22, 53	18, 57	15, 60
3, 9	8, 31	6, 33		5, 10	23, 57	19, 61	15, 65
3, 10	9, 33	6, 36		5, 11	24, 61	20, 65	16, 69
3, 11	9, 36	6, 39		5, 12	26, 64	21, 69	16, 74
3, 12	10, 38	7, 41		5, 13	27, 68	22, 73	17, 78
3, 13	10, 41	7, 44		5, 14	28, 72	22, 78	17, 83
3, 14	11, 43	7, 47		5, 15	29, 76	23, 82	18, 87
3, 15	11, 46	8, 49		5, 16	31, 79	24, 86	18, 92
3, 16	12, 48	8, 52		5, 17	32, 83	25, 90	19, 96
3, 17	12, 51	8, 55		5, 18	33, 87	26, 94	19, 101
3, 18	13, 53	8, 58		5, 19	34, 91	27, 98	20, 105
3, 19	13, 56	9, 60		5, 20	35, 95	28, 102	20, 110
3, 20	14, 58	9, 63		5, 21	37, 98	29, 106	21, 114
3, 21	14, 61	9, 66	6, 69	5, 22	38, 102	29, 111	21, 119
3, 22	15, 63	10, 68	6, 72	5, 23	39, 106	30, 115	22, 123
3, 23	15, 66	10, 71	6, 75	5, 24	40, 110	31, 119	23, 127
3, 24	16, 68	10, 74	6, 78	5, 25	42, 113	32, 123	23, 132
3, 25	19, 71	11, 76	6, 81				
				6, 6	26, 52	23, 55	
4, 4	10, 26			6, 7	27, 57	24, 60	
4, 5	11, 29			6, 8	29, 61	25, 65	21, 69
4, 6	12, 32	10, 34		6, 9	31, 65	26, 70	22, 74
4, 7	13, 35	10, 38		6, 10	32, 70	27, 75	23, 79
4, 8	14, 38	11, 41		6, 11	34, 74	28, 80	23, 85

Adattata da White [114].

La tabella a fianco mostra i valori critici del test T di Mann-Whitney per la somma dei ranghi.

Essi dipendono dalla numerosità del campione più piccolo (n_1) e dalla numerosità del campione più grande (n_2).

* Test non parametrici basati sui ranghi

Procedimento di calcolo del test di Mann-Whitney.

Ricapitolando, il procedimento per il calcolo del test di Mann-Whitney per la verifica dell'ipotesi che un trattamento non abbia effetto, consiste nei seguenti passi:

- Attribuire un rango a tutte le osservazioni sulla base della loro grandezza, assegnando rango 1 all'osservazione più piccola*.
- Calcolare T, la somma dei ranghi del campione più piccolo**.
- Confrontare il valore ottenuto di T con la distribuzione di tutte le possibili somme dei ranghi, per vedere se il valore ottenuto sia compatibile con l'ipotesi che le differenze fra i due gruppi a confronto siano dovute solo alla casualità della loro composizione.

* A osservazioni con lo stesso valore deve essere assegnato lo stesso rango, pari alla media dei ranghi che sarebbero stati attribuiti alle singole osservazioni se fosse stato possibile distinguerle.

** Se entrambi i campioni hanno la stessa ampiezza, calcoliamo T da uno qualunque di essi.

* Test non parametrici basati sui ranghi

Approssimazione della distribuzione di T a quella di Gauss

- Per esperimenti con campioni più ampi di quelli elencati nella tabella, risulta poco agevole calcolare l'esatta distribuzione di T perché il numero delle possibili combinazioni di ranghi è molto alto.
- Quando il campione più ampio contiene più di 8 elementi, la distribuzione di T è approssimativamente una distribuzione normale con media e deviazione standard rispettivamente:

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \qquad \sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

- Pertanto è possibile trovare il valore standard di T: $Z_T = \frac{T - \mu_T}{\sigma_T}$

e confrontarlo con i valori critici della distribuzione normale, per es. quelli che individuano, tra i valori possibili, una quota pari al 5% dei valori più estremi ($\alpha=0,05$).

* Esercizi 13.6.1*

TEST di MANN-WHITNEY sulla somma dei ranghi: durata di remissione della depressione endogena versus quella neurotica

Durata remissione (giorni)	
Depressione endogena	Depressione neurotica
109	546
214	844
1818	602
140	87
179	794
744	643
105	199
101	91
105	105
1547	479
529	1296
140	279

Ordinamento e ranghi per il totale di 24 pazienti:

Ordine: 87, 91, **101, 105, 105**, 105, **109, 140, 140, 179**, 199, **214**

Rango: 1, 2, **3, 5, 5**, 5, **7, 8.5, 8.5, 10**, 11, **12**

Ordine: 279, 479, **529**, 546, 602, 643, **744**, 794, 844, 1296, **1547, 1818**

Rango: 13, 14, **15**, 16, 17, 18, **19**, 20, 21, 22, **23, 24**

Somma dei ranghi tra i 12 pazienti con *depressione endogena*: **140**

Statisticamente non significativo ($P > 0.05$)

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{12(12 + 12 + 1)}{2} = 150$$

$$\sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{1800} = 42,43$$

$$Z_T = \frac{T - \mu_T}{\sigma_T} = \frac{140 - 150}{42,43} = -0,24$$

I valori critici di Z associati ad $\alpha=0,05$ sono $-1,96$ e $1,96$; pertanto, anche utilizzando l'approssimazione alla distribuzione normale giungiamo alla stessa conclusione.

Nota: i valori critici associati ad $\alpha=0,05$ per $n_1=n_2=12$ (non presenti in tabella) sono 115 e 185

* Esercizi 13.6.2*

In un campione casuale di 15 maschi adulti con livello del colesterolo totale maggiore di 200 si è misurata la percentuale di massa grassa. E così in un secondo campione casuale, da confrontare con il primo, di 10 maschi adulti con livello di colesterolo totale minore di 200. Saggiare l'ipotesi H_0 che la percentuale di massa grassa non è significativamente ($\alpha=0.05$) diversa fra i due gruppi. (nota: utilizzare un test non parametrico)

Colesterolo > 200		Colesterolo < 200	
% massa grassa	Ranghi	% massa grassa	Ranghi
19.5	4	12.2	1
22.2	6	13.1	2
22.5	7	13.6	3
23.0	8	22.1	5
23.5	9	24.8	12
23.6	10	25.4	13.5
24.4	11	25.4	13.5
26.0	15	27.5	20
26.1	16	29.4	24
26.4	17	30.4	25
26.7	18		
27.4	19		
28.5	21		
29.1	22.5		
29.1	22.5		

- La somma dei ranghi nel gruppo più piccolo è 119. I valori estremi per $\alpha=0,05$ e $n_1=10$ e $n_2=15$ sono 94 e 166. Dunque la percentuale di massa grassa nel gruppo con colesterolo <200 non è significativamente diversa da quella del gruppo con colesterolo >200.

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{10(10 + 15 + 1)}{2} = 130$$

$$\sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{325} = 18,03$$

$$Z_T = \frac{T - \mu_T}{\sigma_T} = \frac{119 - 130}{18,03} = -0,61$$

I valori critici di Z associati ad $\alpha=0,05$ sono $-1,96$ e $1,96$; pertanto, anche utilizzando l'approssimazione alla distribuzione normale giungiamo alla stessa conclusione.

* Esercizi 13.6.3 *

rapporti di penetrazione x 10⁻⁶/ min

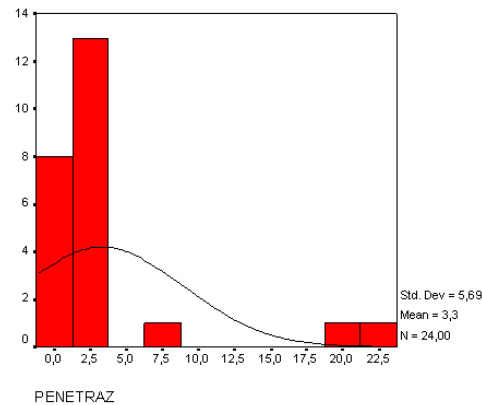
Retina normale	ranghi	Alter. fovea	ranghi
0,5	1	1,2	7
0,7	2,5	1,4	10
0,7	2,5	1,6	13
1,0	4,5	1,7	16
1,0	4,5	1,7	16
1,2	7	1,8	18
1,4	10	2,2	19
1,4	10	2,3	20
1,6	13	2,4	21
1,6	13	6,4	22
1,7	16	19,0	23
1,2	7	23,6	24
Somma ranghi=91		Somma ranghi=209	
Rango medio=7,58		Rango medio=17,42	

Persone affette da certi disturbi della retina manifestano un'eccessiva permeabilità dei vasi sanguigni retinici. Questa permeabilità viene rilevata iniettando in vena del liquido fluorescente e misurandone la quantità che penetra nell'occhio. Gerald Fishman e coll. studiarono un gruppo di soggetti con retine normali e un gruppo con alterazioni evidenti della sola fovea della retina. I rapporti di penetrazione del liquido di contrasto in questi gruppi sono illustrati, assieme alla definizione dei ranghi, nella tabella a lato.

Questi dati confermano l'ipotesi che c'è differenza di permeabilità retinica fra persone con retina normale e persone affette da alterazioni della fovea ?

* Soluzione Esercizio 13.6.3 *

Penetrazione retina



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std.	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
PENETRAZ	24	,50	23,60	3,3042	5,6944	3,106	,472
Valid N (listwise)	24						

$$\mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{12(12 + 12 + 1)}{2} = 150$$

$$\sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{12 \times 12 \times (12 + 12 + 1)}{12}} = \sqrt{300} = 17,32$$

$$Z_T = \frac{T - \mu_T}{\sigma_T} = \frac{91 - 150}{17,32} = -3,41$$

Poiché Z_T è esterno all'intervallo $(-2,58; 2,58)$ possiamo rifiutare l'ipotesi nulla con un rischio di errore $p < 0,01$ e concludiamo che i soggetti con alterazione della fovea hanno un rapporto di penetrazione significativamente maggiore di quello dei soggetti senza alterazioni della retina.

* Test non parametrici basati sui ranghi

Il Test con segno di Wilcoxon

- Analogamente al test t di Student per dati appaiati, il test con segno di Wilcoxon è utilizzato negli esperimenti in cui i dati risultano appaiati (ad es., nel caso di esperimenti che prevedono un unico gruppo di soggetti osservati prima e dopo un certo trattamento) ed in cui non siano rispettate le condizioni di applicabilità dei test parametrici.
- Si calcolano le differenze causate dal trattamento in ciascun soggetto in studio, e si assegna un rango a ciascuna differenza in relazione al valore assoluto; successivamente attribuiamo al rango il segno della differenza e, infine, calcoliamo la somma algebrica dei ranghi in modo da ottenere il test statistico W.
- Confrontiamo poi il valore di W con la distribuzione di tutti i possibili valori di W per campioni di ampiezza uguale a quella del nostro studio, per verificare se le osservazioni sono compatibili con l'ipotesi che il trattamento non abbia avuto effetto.

* Test non parametrici basati sui ranghi

Il Test con segno di Wilcoxon: procedimento di calcolo

Illustriamo il procedimento di calcolo utilizzando un esperimento ipotetico sull'efficacia di un diuretico somministrato a 6 persone. La tabella mostra i risultati di questo esperimento, insieme alla variazione nella diuresi in ciascun soggetto dovuta al trattamento.

Soggetto	Diuresi giornaliera, ml/giorno			Rango delle differenze	Rango con Segno della differenza
	Prima della somministrazione	Dopo la somministrazione	Differenza		
1	1490	1600	110	5	5
2	1300	1850	550	6	6
3	1380	1420	40	2	2
4	1410	1500	90	4	4
5	1350	1400	50	3	3
6	1010	1000	-10	1	-1
					W=19

La diuresi giornaliera è aumentata in 5 soggetti su 6. Questo è sufficiente a giustificare l'asserzione che il farmaco è efficace?

* Test non parametrici basati sui ranghi

Il Test con segno di Wilcoxon: considerazioni

- Se il trattamento non avesse effetto, i ranghi associati a variazioni positive dovrebbero essere pressoché pari ai ranghi associati a variazioni negative e la somma dei ranghi con segno W dovrebbe assumere un valore prossimo a 0.
- Viceversa, quanto più il trattamento modifica i valori della variabile oggetto di studio, tanto più le variazioni tenderanno ad assumere lo stesso segno. In questo caso la somma dei ranghi con segno W tenderà ad assumere un valore positivo o negativo “grande”.

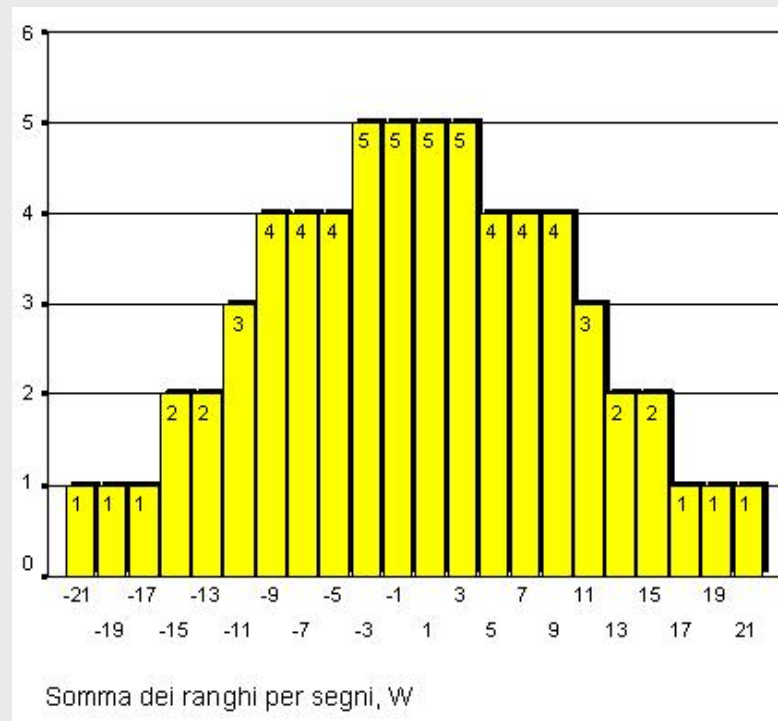
Per delimitare il confine fra “piccolo” e “grande” dobbiamo considerare tutte le possibili combinazioni di ranghi e calcolare per ognuna di esse W , in modo da identificare quali sono i suoi valori più “estremi” di rifiuto dell’ipotesi nulla di inefficacia del trattamento (per un dato il livello di significatività α).

* Test non parametrici basati sui ranghi

Il Test con segno di Wilcoxon: distribuzione di W

La tabella mostra una parte dei possibili modi ($n.tot=2^6=64$) con cui i segni (qui codificati con 0=positivo e con 1=negativo) sono attribuiti ai ranghi associati ai 6 soggetti in studio ed i corrispondenti valori W.

RANGO						Somma W
1	2	3	4	5	6	
0	0	0	0	0	0	21
0	0	0	0	0	1	9
0	0	0	0	1	0	11
0	0	0	0	1	1	-1
0	0	0	1	0	0	13
0	0	0	1	0	1	1
0	0	0	1	1	0	3
0	0	0	1	1	1	-9
0	0	1	0	0	0	15
0	0	1	0	0	1	3
0	0	1	0	1	0	5
0	0	1	0	1	1	-7
0	0	1	1	0	0	7
0	0	1	1	0	1	-5
0	0	1	1	1	0	-3
0	0	1	1	1	1	-15
.
.
1	1	1	1	1	0	-9
1	1	1	1	1	1	-21



Il diagramma a barre mostra la distribuzione completa dei valori di W.

Come si può osservare i valori -19 e 19 delimitano il 6,25% (pari a $4/64$) dei valori più estremi della distribuzione di W.

* Test non parametrici basati sui ranghi

Il Test di Wilcoxon: quando W è sufficientemente estremo da rifiutare H_0

- Va osservato che, come nel test di Mann-Whitney per la somma dei ranghi, la natura discreta della distribuzione dei possibili valori di W implica l'impossibilità di determinare sempre i valori critici associati esattamente al livello di significatività voluto (generalmente $\alpha=0,05$).
- Come si può notare nel nostro esempio, il livello α più prossimo al 5% è 6,25% ($p=4/64$) a cui corrispondono i valori critici -19 e $+19$.
- Pertanto il valore $W=19$ ottenuto ci autorizza a rifiutare l'ipotesi che il trattamento non abbia avuto effetto (cioè che il diuretico non sia efficace). Così facendo ci accogliamo un rischio di errore del 6,25%.

* Test non parametrici basati sui ranghi

Il Test di Wilcoxon: valori critici.

Valori critici del test W di Wilcoxon (test a due code)

n	Valore critico	p	n	Valore critico	p
5	15	0.062	13	65	0.022
6	21	0.032		57	0.048
	19	0.062	14	73	0.020
7	28	0.016		63	0.050
	24	0.046	15	80	0.022
8	32	0.024		70	0.048
	28	0.054	16	88	0.022
9	39	0.020		76	0.050
	33	0.054	17	97	0.020
10	45	0.020		83	0.050
	39	0.048	18	105	0.020
11	52	0.018		91	0.048
	44	0.054	19	114	0.020
12	58	0.020		98	0.050
	50	0.052	20	124	0.020
				106	0.048

Fonte: adattata da F. Mosteller e R. Rourke, *Study, Statistics, Nonparametrics and Order Statistics*, Addison-Wesley, Reading, Mass, 1973, Tab. A-11.

La tabella a fianco mostra i valori di W per esperimenti che coinvolgono fino a 20 unità statistiche.

* Test non parametrici basati sui ranghi

Riepilogo del procedimento di calcolo del test di Wilcoxon.

Ricapitolando, il procedimento per il calcolo del test di Wilcoxon per confrontare gli effetti di un trattamento effettuato su un unico gruppo di soggetti, consiste nel:

- Calcolare per ciascun soggetto la variazione verificatasi nella variabile oggetto di studio.
- Attribuire un rango a tutte le differenze in relazione al loro valore assoluto.
- Assegnare a ciascun rango il segno della differenza corrispondente.
- Sommare i ranghi con segno in modo da ottenere il test statistico W .
- Confrontare il valore di W ottenuto con la distribuzione dei possibili valori di W per decidere se è abbastanza “elevato” da far rifiutare l’ipotesi che il trattamento non ha effetto.

NOTA: ad una eventuale differenza nulla non è attribuito rango e la corrispondente unità statistica non viene presa in considerazione (l’ampiezza campionaria viene diminuita di tante unità quante sono le differenze nulle).

* Test non parametrici basati sui ranghi

Approssimazione della distribuzione di W a quella di Gauss.

- Per esperimenti con campioni più ampi di quelli elencati nella tabella dei valori critici, ci possiamo servire del fatto che la distribuzione di W si adatta, con buona approssimazione, ad una distribuzione normale con media e deviazione standard rispettivamente uguali a:

$$\mu = 0 \qquad \sigma_w = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

dove n rappresenta il numero delle unità statistiche in studio.

- Pertanto è possibile calcolare il valore standard di W : $Z_w = \frac{W - \mu_w}{\sigma_w} = \frac{W}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$

e confrontarlo con i valori critici della distribuzione normale, per es. quelli che individuano, tra i valori possibili, una quota pari al 5% dei valori più estremi ($\alpha=0,05$).

* *Esercizio 13.15* *

I seguenti dati forniscono la perdita di peso dopo un anno di 9 soggetti per cui era stata posta la diagnosi di diabete e che ricevettero Diabinese per il controllo della loro malattia. Usando l'appropriata tecnica non parametrica, si esegua il test di significatività della seguente ipotesi nulla: con il trattamento con Diabinese non vi è alcun cambiamento in peso dopo un anno.

Codice soggetto	Perdita di peso dopo un anno (lb)
1	25
2	13
3	41
4	-11
5	-8
6	4
7	0
8	11
9	-4

* Soluzione Esercizio 13.15*

Codice soggetto	Perdita di peso dopo un anno (lb)	Rango
1	25	7
2	13	6
3	41	8
4	-11	-4.5
5	-8	-3
6	4	1.5
7	0	-
8	11	4.5
9	-4	-1.5

$$W=7+6+8-4.5-3+1.5+4.5-1.5=18$$

Essendo il valore critico associato a $n=8$ pari a 28, W non raggiunge un valore sufficientemente estremo per giustificare il rifiuto dell'ipotesi nulla.

*NOTA: nell'assegnazione dei ranghi una differenza di 0 è ignorata.

Il problema può essere risolto anche sfruttando il fatto che la distribuzione di W si adatta con buona approssimazione alla distribuzione normale con media e dev. standard rispettivamente:

$$\mu_w = 0 \quad \text{e} \quad \sigma_w = \sqrt{\frac{n(n+1)(2n+1)}{6}} \quad \text{ovvero, nel caso specifico:}$$

$$\mu_w = 0 \quad \text{e} \quad \sigma_w = \sqrt{\frac{n(n+1)(2n+1)}{6}} = \sqrt{\frac{8(8+1)(16+1)}{6}} = 14.28$$

$$z_w = \frac{W - \mu_w}{\sigma_w} = \frac{18}{14.28} = 1.26$$

Poiché $1.26 < 1.96$ che è il valore critico per la curva di Gauss associato al livello di confidenza del 95%, non abbiamo ragione di rifiutare l'ipotesi nulla.

* Test non parametrici basati sui ranghi

Il Test di Kruskal-Wallis: procedimento di calcolo

Supponendo per semplicità che ci siano solo tre gruppi di ampiezza campionaria n_1, n_2, n_3 e somme dei ranghi R_1, R_2, R_3 rispettivamente, i ranghi medi nei gruppi saranno allora:

$$\bar{R}_1 = \frac{R_1}{n_1}$$

$$\bar{R}_2 = \frac{R_2}{n_2}$$

$$\bar{R}_3 = \frac{R_3}{n_3}$$

Il rango medio di tutte le $n_1+n_2+n_3=N$ osservazioni è la media dei primi N numeri interi:

$$\bar{R} = \frac{1+2+3+\dots+N}{N} = \frac{N+1}{2}$$

Come indicatore della variabilità tra valori osservati e valori attesi nel caso in cui l'ipotesi che il trattamento non abbia effetto sia corretta, viene utilizzata la somma dei quadrati delle differenze (pesate dalle ampiezze campionarie) tra le medie campionarie e la media generale, ovvero:

$$D = n_1(\bar{R}_1 - \bar{R})^2 + n_2(\bar{R}_2 - \bar{R})^2 + n_3(\bar{R}_3 - \bar{R})^2$$

* Test non parametrici basati sui ranghi

Il Test di Kruskal-Wallis

- Il test di Kruskal-Wallis è utilizzato per analizzare esperimenti nei quali almeno tre gruppi di soggetti sono esposti a differenti trattamenti, ma le osservazioni non rispettano le condizioni di applicabilità dell'analisi di varianza a un criterio di classificazione.
- Esso è un' immediata generalizzazione del test di Mann-Whitney: anche qui i ranghi sono attribuiti a tutte le osservazioni, indipendentemente dal gruppo di appartenenza e viene calcolata la somma dei ranghi per ciascun gruppo.
- Se i trattamenti non hanno effetto, ranghi grandi e piccoli dovrebbero essere equamente distribuiti nei gruppi e il rango medio di ciascun gruppo dovrebbe essere prossimo alla media di tutti i ranghi.
- Tanto maggiore è la differenza esistente fra i ranghi medi osservati in ciascun gruppo e quelli attesi, tanto minore è la probabilità che la tesi di inefficacia dei trattamenti possa essere accolta.

* Test non parametrici basati sui ranghi

Il Test di Kruskal-Wallis: approssimazione alla distribuzione del χ^2

La distribuzione dei possibili valori di D , nell'ipotesi che i trattamenti non abbiano effetto, dipende dalla numerosità del campione.

- E' possibile ottenere un valore del test che sia indipendente dalla numerosità del campione dividendo D per $N(N+1)/12$, ottenendo così la statistica H di Kruskal-Wallis:

$$H = \frac{D}{N(N+1)/12} = \frac{12}{N(N+1)} \sum n_t (\bar{R}_t - \bar{R})^2$$

- Se le dimensioni campionarie non sono troppo ridotte*, la distribuzione χ^2 con $\nu=K-1$ gradi di libertà (dove K è il numero dei gruppi), è una buona approssimazione della distribuzione di H .
- Possiamo quindi verificare l'ipotesi che il trattamento non abbia effetto confrontato il valore di H osservato con i valori critici del χ^2 .

* L'approssimazione è valida in esperimenti con tre gruppi, quando ciascun gruppo contiene almeno 5 elementi e, in esperimenti con 4 gruppi, quando l'intero studio coinvolge più di 10 individui. Per studi con campioni più limitati è necessario utilizzare la distribuzione esatta di H .

* Test non parametrici basati sui ranghi

Il Test di Kruskal-Wallis: riepilogo del procedimento di calcolo

Ricapitolando, il procedimento nell'analisi di esperimenti in cui vengono confrontati più di due gruppi sottoposti a trattamenti diversi, consiste nel:

- Assegnare il rango a ciascuna osservazione indipendentemente dal gruppo di appartenenza.
- Calcolare la statistica H in modo da ottenere un indicatore normalizzato della differenza della media dei ranghi in ciascun gruppo dalla media dei ranghi di tutte le osservazioni.
- Confrontare il valore di H con la distribuzione del χ^2 per $\nu=k-1$ gradi di libertà (dove K rappresenta il numero dei gruppi a confronto) per decidere se è abbastanza “elevato” da far rifiutare l'ipotesi che il trattamento non ha effetto.

NOTA: In analogia con l'analisi di varianza, esistono procedure di confronto multiplo a coppie (test di Student-Newman-Keuls per uguali ampiezze campionarie, test di Dunn per diverse ampiezze campionarie e con un solo gruppo di controllo (test di Dunnett).

* Test di Dunn (confronti multipli)

Nei confronti multipli, quando le ampiezze campionarie sono diverse, si utilizza il test di Dunn*. Per tutti i confronti a due a due si calcola la statistica:

$$Q = \frac{(\bar{R}_A - \bar{R}_B)}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Dove \bar{R}_A e \bar{R}_B sono i ranghi medi dei due gruppi a confronto
 N è la dimensione totale del campione

n_A e n_B sono rispettivamente il numero di osservazioni nei campioni A e B .

I valori critici di Q dipendono dal numero di gruppi trattati K .

* Il test di Dunn può comunque essere usato anche per ampiezze uguali.

* Test di Dunn

Valori critici

Valori critici di Q per confronti multipli con test non parametrici

k	α	
	0,05	0,01
2	1,960	2,576
3	2,394	2,936
4	2,639	3,144
5	2,807	3,291
6	2,936	3,403
7	3,038	3,494
8	3,124	3,570
9	3,197	3,635
10	3,261	3,692
11	3,317	3,743
12	3,368	3,789
13	3,414	3,830
14	3,456	3,868
15	3,494	3,902
16	3,529	3,935
17	3,562	3,965
18	3,593	3,993
19	3,622	4,019
20	3,649	4,044
21	3,675	4,067
22	3,699	4,089
23	3,722	4,110
24	3,744	4,130
25	3,765	4,149

Adattata dalla tabella B.14 di J.H. Zar, *Biostatistical Analysis*, 2^a ed., Prentice-Hall, Englewood Cliffs, N.J., 1984, p. 568.

La tabella a fianco mostra i valori critici di Q per i confronti multipli col test di Dunn in funzione del numero di gruppi a confronto (k) e dei livelli di significatività $\alpha=0,05$ e $\alpha=0,01$.

* *Esercizio 13.20* *

Esempio tratto da "Statistica per discipline biomediche". Stanton A. Glantz. McGraw-Hill

La tabella mostra i risultati di un esperimento mirante a verificare le differenze fra uomini, donne e donne che assumono contraccettivi nella metabolizzazione della caffeina. L'esperimento prevedeva la somministrazione di una compressa contenente 250 mg di caffeina e la rilevazione dell'emivita.

Emivita (ore) della caffeina ematica dopo un'unica dose

Maschi ($n_1=13$)	rango	Femmine (non contraccettivi) ($n_2=9$)	rango	Femmine (con contraccettivi) ($n_3=9$)	rango
2,04	1	5,30	12	10,36	25
5,16	10	7,28	19	13,28	29
6,11	15	8,98	21	11,81	28
5,82	14	6,59	16	4,54	6
5,41	13	4,59	8	11,04	26
3,51	4	5,17	11	10,08	24
3,18	2	7,25	18	14,47	31
4,57	7	3,47	3	9,43	23
4,83	9	7,60	20	13,41	30
11,34	27				
3,79	5				
9,03	22				
7,21	17				
Somma ranghi R_t	146		128		222
Rango medio R_t/n_t	11,23		14,22		24,67

Emivita: tempo necessario perché la concentrazione di caffeina si riduca alla metà del valore iniziale

* Soluzione Esercizi 13.20*

$$\bar{R} = \frac{1+2+3+\dots+N}{N} = \frac{N+1}{2} = \frac{31+1}{2} = 16$$

$$\begin{aligned} D &= n_1(\bar{R}_1 - \bar{R})^2 + n_2(\bar{R}_2 - \bar{R})^2 + n_3(\bar{R}_3 - \bar{R})^2 = \\ &= 13(11,23 - 16)^2 + 9(14,22 - 16)^2 + 9(24,67 - 16)^2 = 1000,82 \end{aligned}$$

$$H = \frac{D}{N(N+1)/12} = \frac{1000,82}{31(31+1)/12} = 12,07$$

$$v = k - 1 = 3 - 1 = 2 \quad \text{gradi di libertà}$$

Per $\alpha=0,01$ abbiamo un valore critico pari a 9,201; pertanto concludiamo che in almeno uno dei tre gruppi, l'eliminazione della caffeina avviene con una velocità significativamente diversa ($p<0,01$). Per vedere in quale/i gruppo/i risiede la differenza è necessario procedere ai confronti multipli post-hoc tramite test di Dunn

* Confronti multipli Esercizi 13.20*

$$Q = \frac{\bar{R}_{contrac} - \bar{R}_{maschi}}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_{contrac}} + \frac{1}{n_{maschi}} \right)}} = \frac{24,67 - 11,23}{\sqrt{\frac{31(31+1)}{12} \left(\frac{1}{9} + \frac{1}{13} \right)}} = 3,409$$

$$Q = \frac{\bar{R}_{contrac} - \bar{R}_{no_contrac}}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_{contrac}} + \frac{1}{n_{no_contrac}} \right)}} = \frac{24,67 - 14,22}{\sqrt{\frac{31(31+1)}{12} \left(\frac{1}{9} + \frac{1}{9} \right)}} = 2,438$$

$$Q = \frac{\bar{R}_{no_contrac} - \bar{R}_{maschi}}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_{no_contrac}} + \frac{1}{n_{maschi}} \right)}} = \frac{14,22 - 11,23}{\sqrt{\frac{31(31+1)}{12} \left(\frac{1}{9} + \frac{1}{13} \right)}} = 0,7583$$

Il valore critico di Q associato a $k=3$ gruppi e $\alpha=0,05$ è **2,394**

Perciò possiamo dire che gli uomini e le donne che non prendono contraccettivi orali eliminano la caffeina in modo simile e che le donne che assumono contraccettivi orali la eliminano in modo significativamente più lento ($\alpha < 0,05$) sia delle donne che non li prendono, sia degli uomini.

* Test non parametrici basati sui ranghi

Il Test di Friedman

- Il test di Friedman è utilizzato per analizzare esperimenti nei quali ciascun soggetto è sottoposto a un certo numero di trattamenti diversi, ma le osservazioni non soddisfano le condizioni di applicabilità dei metodi parametrici.
- Il procedimento attribuisce un rango alle risposte di ciascun soggetto ai trattamenti senza tener conto degli altri soggetti.
- Se l'ipotesi che il trattamento non ha effetto è corretta, allora in ciascun soggetto i ranghi saranno distribuiti in modo casuale e le somme dei ranghi saranno analoghe per ogni gruppo sottoposto ai trattamenti.

Soggetti	Trattamenti			
	1	2	3	4
1	1	2	3	4
2	4	1	2	3
3	3	4	1	2
4	2	3	4	1
5	1	4	3	2
Somma dei ranghi R_t	11	14	13	12

Esempio:

Ranghi dei risultati per un esperimento nel quale 5 soggetti sono sottoposti a 4 trattamenti.

Esempio tratto da "Statistica per discipline biomediche". Stanton A. Glantz. McGraw-Hill

* Test non parametrici basati sui ranghi

Il Test di Friedman: procedimento di calcolo

- Nell' esempio le somme dei ranghi sono tutte molto vicine alla media dei ranghi in ogni soggetto moltiplicata per il numero dei soggetti, ovvero $((1+2+3+4)/4) \times 5 = 12,5$.
- Poiché non è presente una grande variabilità nella somma dei ranghi e nessuna di queste si discosta di molto dal quintuplo del rango medio, la tabella non induce a ritenere che uno dei trattamenti eserciti un effetto sistematico sui soggetti in studio.
- Generalizzando: se ci sono n soggetti e k trattamenti, il rango medio in ogni soggetto è $(k+1)/2$ e, nell' ipotesi di inefficacia dei trattamenti, potremmo attenderci che ciascuna delle somme dei ranghi sia uguale a circa n volte il rango medio per ciascun soggetto, ovvero $n(k+1)/2$.
- Una misura della variabilità della somma dei ranghi può allora essere rappresentata dalla somma dei quadrati delle differenze fra la somme dei ranghi osservate in ciascun gruppo sperimentale e la somma dei ranghi attesa, ovvero:

$$S = \sum [R_t - n(k+1)/2]^2 \quad \text{dove } R_t \text{ indica la somma dei ranghi per il } t\text{-esimo trattamento}$$

* Test non parametrici basati sui ranghi

Il Test di Friedman: approssimazione alla distribuzione del χ^2

- Quanto maggiore è il legame fra i ranghi di ciascun soggetto e i trattamenti, tanto maggiore è il valore della statistica S .
- La distribuzione dei possibili valori di S , nell'ipotesi che i trattamenti non abbiano effetto, dipende dalla numerosità del campione.
- Dividendo S per $nk(k+1)/12$ otteniamo la statistica di Friedman che, in campioni sufficientemente ampi, segue la distribuzione del χ^2 , con $v=k-1$ gradi di libertà, indipendentemente dalla numerosità campionaria:

$$\chi_r^2 = \frac{S}{nk(k+1)/12} = \frac{12 \sum [R_t - n(k+1)/2]^2}{nk(k+1)}$$

- Possiamo quindi verificare l'ipotesi che il trattamento non abbia effetto confrontando il valore osservato della statistica con i valori critici del χ^2 .

NOTA: Quando ci sono 3 trattamenti e non più di 9 soggetti o 4 trattamenti e non più di 4 soggetti ciascuno l'approssimazione è inadeguata ed è pertanto necessario confrontare il valore della statistica con la distribuzione esatta dei valori ottenuta elencando tutte le possibilità.

* Test non parametrici basati sui ranghi

Il Test di Friedman: valori critici.

Valori critici del χ_r^2 di Friedman

k = 3 trattamenti			k = 4 trattamenti		
n	χ_r^2	P	n	χ_r^2	P
3	6,00	0,028	2	6,00	0,042
4	6,50	0,042	3	7,00	0,054
	8,00	0,005		8,20	0,017
5	5,20	0,093	4	7,50	0,054
	6,40	0,039		9,30	0,011
	8,40	0,008	5	7,80	0,049
6	5,33	0,072		9,96	0,009
	6,33	0,052	6	7,60	0,043
	9,00	0,008		10,20	0,010
7	6,00	0,051	7	7,63	0,051
	8,86	0,008		10,37	0,009
8	6,25	0,047	8	7,65	0,049
	9,00	0,010		10,35	0,010
9	6,22	0,048			
	8,67	0,010			
10	6,20	0,046			
	8,60	0,012			
11	6,54	0,043			
	8,91	0,011			
12	6,17	0,050			
	8,67	0,011			
13	6,00	0,050			
	8,67	0,012			
14	6,14	0,049			
	9,00	0,010			
15	6,40	0,047			
	8,93	0,010			

Fonte: adattata da Owen, *Handbook of Statistical Tables*, U.S. Department of Energy, Addison-Wesley, Reading, Mass., 1962. Per gentile concessione.

La tabella a fianco mostra i valori critici di χ^2 di Friedman che, per esperimenti con 3 e 4 trattamenti e diverse ampiezze campionarie

* Test non parametrici basati sui ranghi

Il Test di Friedman: riepilogo del procedimento di calcolo

Ricapitolando, il procedimento nell'analisi di esperimenti in cui gli stessi individui sono sottoposti a più trattamenti, consiste nel:

- Attribuire un rango a ciascuna osservazione in riferimento a ciascun soggetto in studio.
- Calcolare la somma dei ranghi osservati in tutti i soggetti in studio per ciascun trattamento.
- Calcolare il test di Friedman come indicatore della differenza fra le somme dei ranghi osservate e quelle attese nel caso in cui il trattamento non avesse effetto.
- Confrontare il valore ottenuto con i valori critici della distribuzione del χ^2 per decidere se è sufficientemente “elevato” da rifiutare l'ipotesi di inefficacia.

NOTA: In caso di rifiuto di H_0 si può poi procedere con i confronti multipli a due a due o con il gruppo di controllo rispettivamente con adattamenti del test di Student-Newman-Keuls e di Dunnet.

*Test di Student-Newman-Keuls (SNK) (confronti multipli per il Test di Friedman)

Quando il test di Friedman è significativo si può procedere ai confronti multipli a due a due sui trattamenti utilizzando un adattamento del test SNK:

$$q = \frac{R_A - R_B}{\sqrt{\frac{pn(p+1)}{12}}}$$

R_A e R_B sono le somme dei ranghi dei due trattamenti a confronto

n è il numero dei soggetti nell' esperimento

p è il numero dei confronti

Il valore ottenuto di q si confronta col valore critico di q per p confronti (vedi tabella seguente).

Valori critici di q ($\alpha=0,05$)

p	2	3	4	5	6	7	8	9	10
q	2,772	3,314	3,633	3,858	4,030	4,170	4,286	4,387	4,474

* Test di Dunnet

(confronti multipli per il Test di Friedman)

Quando il test di Friedman è significativo si può procedere ai confronti con un solo trattamento di controllo utilizzando un adattamento del test di Dunnet:

$$q' = \frac{R_{con} - R_A}{\sqrt{\frac{pn(p+1)}{6}}}$$

R_{con} e R_B sono le somme dei ranghi del controllo e del trattamento a confronto

n è il numero dei soggetti nell'esperimento

p è il numero dei confronti

Il valore ottenuto di q' si confronta col valore critico di q' per p confronti (vedi tabella seguente).

Valori critici di q' ($\alpha=0,05$)

p	2	3	4	5	6	7	8	9	10
q'	1,96	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69

* *Esercizio 13.21.7* *

Esempio tratto da “Statistica per discipline biomediche”. Stanton A. Glantz. McGraw-Hill

L'elevata pressione nei vasi sanguigni polmonari, definita ipertensione polmonare, è una condizione invalidante e potenzialmente letale. L'elevata pressione in circolo può provocare il passaggio di liquido nei polmoni, inibendo il trasporto gassoso. L'idralazina è un farmaco che provoca rilassamento e dilatazione nei vasi sanguigni, riducendo la pressione.

La tabella riproduce le resistenze¹ polmonari totali che Rubin e Peter avevano utilizzato per verificare se l'idralazina provocasse anche una diminuzione della pressione sanguigna nel circolo polmonare.

Quale fiducia possiamo avere nell'asserire che il farmaco provoca una diminuzione della resistenza?

soggetto	Prima (controllo)		Dopo 48 ore		Dopo 3-6 mesi	
	<i>Unità</i>	<i>Rango</i>	<i>Unità</i>	<i>Rango</i>	<i>Unità</i>	<i>Rango</i>
1	22,2	3	5,4	1	10,6	2
2	17,0	3	6,3	2	6,2	1
3	14,1	3	8,5	1	9,3	2
4	17,0	3	10,7	1	12,3	2
Somma ranghi		12		5		7

Resistenza polmonare totale in 4 soggetti. Fonte: L.J. Rubin e R.H. Peter, “Oral Hydralazine Therapy for Primary Pulmonary Hypertension”. N. Engl. J. Med., 302:69-73, 1980, tabella 2.

¹ La resistenza polmonare totale indica qual'è la pressione necessaria a produrre un dato flusso di sangue attraverso i polmoni. Quanto più l'ipertensione è grave, tanto più è elevata la resistenza polmonare totale. |

* Soluzione Esercizio 13.21.7*

I possibili ranghi sono 1,2,3 ed il rango medio è $(1+2+3)/3=2$.

Poiché ci sono 4 soggetti, se i trattamenti non avessero effetto le somme di questi ranghi dovrebbero essere prossime a $4 \times 2 = 8$. Abbiamo allora:

$$S = \sum [R_t - n(k+1)/2]^2 = (12-8)^2 + (5-8)^2 + (7-8)^2 = 26$$

$$\chi_r^2 = \frac{S}{nk(k+1)/12} = \frac{26}{4 \times 3(3+1)/12} = 6,5$$

L'ampiezza campionaria è troppo esigua per poter utilizzare la distribuzione del χ^2 per $\nu=k-1$ gradi di libertà. E' pertanto necessario confrontare il valore ottenuto con la distribuzione esatta dei valori ottenuta elencando tutte le possibilità per un esperimento con $k=3$ trattamenti ed $n=4$ soggetti.

La probabilità di ottenere un valore maggiore od uguale a 6,5 è, nel caso in cui un trattamento non abbia effetto, pari a 0,042; pertanto possiamo affermare che l'idralazina altera le resistenze polmonari totali ($p=0,042$).

* *Confronti multipli (Dunnet)Esercizio 13.21.7**

Dato il tipo di disegno sperimentale, che prevede misure ripetute in identiche situazioni sperimentali per i 4 soggetti in studio, per i confronti multipli può essere utilizzato il test di Dunnet considerando come controllo la prima misurazione.

$$q_1' = \frac{R_{con} - R_{48ore}}{\sqrt{\frac{pn(p+1)}{6}}} = \frac{12 - 5}{\sqrt{\frac{2 \times 4 \times 3}{6}}} = \frac{7}{\sqrt{4}} = 3,5$$

$$q_2' = \frac{R_{con} - R_{3-6mesi}}{\sqrt{\frac{pn(p+1)}{6}}} = \frac{12 - 7}{\sqrt{\frac{2 \times 4 \times 3}{6}}} = \frac{5}{\sqrt{4}} = 2,4$$

Il valore critico con 2 confronti e un livello di significatività $\alpha=0,05$ è 1,96; pertanto possiamo affermare che abbiamo una riduzione significativa ($p<0,05$) della resistenza polmonare sia dopo 48 ore, che dopo 3-6 mesi.