* 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
	T=9	25	6

* 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
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:

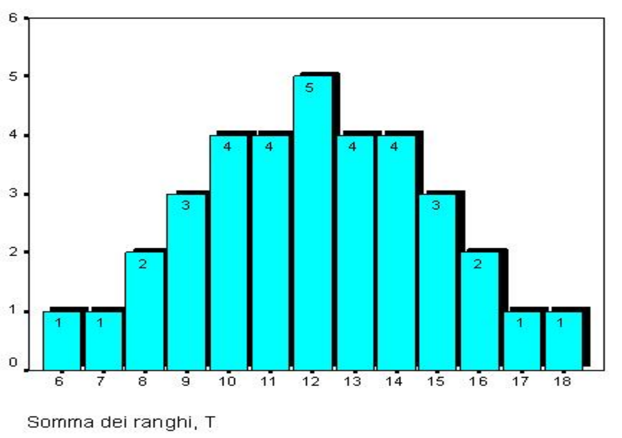
$$_7C_3 = \frac{7!}{3! \times 4!} = 35$$

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 (α=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 α=0,05), ma solo quelli che più gli si avvicinano.

* Test non parametrici basati sui ranghi

Il test di Mann-Whitney. Valori critici.

	Livello di significatività a due code		
n_1, n_2	0.05	0.01	0.001
2, 8	3, 19		
2, 9	3, 21		
2, 10	3, 23		
2, 11	4, 24		
2, 12	4, 26		
2, 13	4, 28		
2, 14	4, 30		
2, 15	4, 32		
2, 16	4, 34		
2, 17	5, 35		
2, 18	5, 37		
2, 19	5, 39	3, 41	
2, 20	5, 41	3, 43	
2, 21	6, 42	3, 45	
2, 22	6, 44	3, 47	
2, 23	6, 46	3, 49	
2, 24	6, 48	3, 51	
2, 25	6, 50	3, 53	
3, 5	6, 21		
3, 6	7, 23		
3, 7	7, 26		
3, 8	8, 28		
3, 9	8, 31	6, 33	
3, 10	9, 33	6, 36	
3, 11	9, 36	6, 39	
3, 12	10, 38	7, 41	
3, 13	10, 41	7, 44	
3, 14	11, 43	7, 47	
3, 15	11, 46	8, 49	
3, 16	12, 48	8, 52	
3, 17	12, 51	8, 55	
3, 18	13, 53	8, 58	
3, 19	13, 56	9, 60	
3, 20	14, 58	9, 63	
3, 21	14, 61	9, 66	6, 69
3, 22	15, 63	10, 68	6, 72
3, 23	15, 66	10, 71	6, 75
3, 24	16, 68	10, 74	6, 78
3, 25	19, 71	11, 76	6, 81
4, 4	10, 26		
4, 5	11, 29		
4, 6	12, 32	10, 34	
4, 7	13, 35	10, 38	
4, 8	14, 38	11, 41	
4, 9	15, 41	11, 45	
4, 10	15, 45	12, 48	
4, 11	16, 48	12, 52	
4, 12	17, 51	13, 55	
4, 13	18, 54	14, 58	10, 62
4, 14	19, 57	14, 62	10, 66
4, 15	20, 60	15, 65	10, 70
4, 16	21, 63	15, 69	11, 73
4, 17	21, 67	16, 72	11, 77
4, 18	22, 70	16, 76	11, 81
4, 19	23, 73	17, 79	12, 84
4, 20	24, 76	18, 82	12, 88
4, 21	25, 79	18, 86	12, 92
4, 22	26, 82	19, 89	13, 95
4, 23	27, 85	19, 93	13, 99
4, 24	28, 88	20, 96	13, 103
4, 25	28, 92	20, 100	14, 106
5, 5	17, 38	15, 40	
5, 6	18, 42	16, 44	
5, 7	20, 45	17, 48	
5, 8	21, 49	17, 53	
5, 9	22, 53	18, 57	15, 60
5, 10	23, 57	19, 61	15, 65
5, 11	24, 61	20, 65	16, 69
5, 12	26, 64	21, 69	16, 74
5, 13	27, 68	22, 73	17, 78
5, 14	28, 72	22, 78	17, 83
5, 15	29, 76	23, 82	18, 87
5, 16	31, 79	24, 86	18, 92
5, 17	32, 83	25, 90	19, 96
5, 18	33, 87	26, 94	19, 101
5, 19	34, 91	27, 98	20, 105
5, 20	35, 95	28, 102	20, 110
5, 21	37, 98	29, 106	21, 114
5, 22	38, 102	29, 111	21, 119
5, 23	39, 106	30, 115	22, 123
5, 24	40, 110	31, 119	23, 127
5, 25	42, 113	32, 123	23, 132
6, 6	26, 52	23, 55	
6, 7	27, 57	24, 60	
6, 8	29, 61	25, 65	21, 69
6, 9	31, 65	26, 70	22, 74
6, 10	32, 70	27, 75	23, 79
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 (α=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 α=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 α=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 (α=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 α=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 α=0,05 sono –1,96 e 1,96; pertanto, anche utilizzando l'approssimazione alla distribuzione normale giungiamo alla stessa conclusione.

* Esercizi 13.6.3*

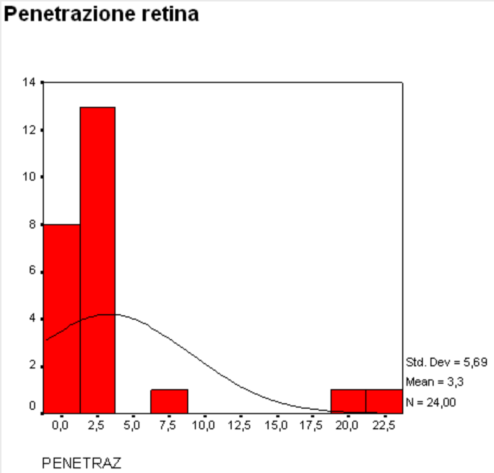
rapporti di penetrazione x 10^{-6}/ min

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*



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.

	Diuresi giornaliera, ml/giorno			Rango delle differenze	Rango con Segno della differenza
Soggetto	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".

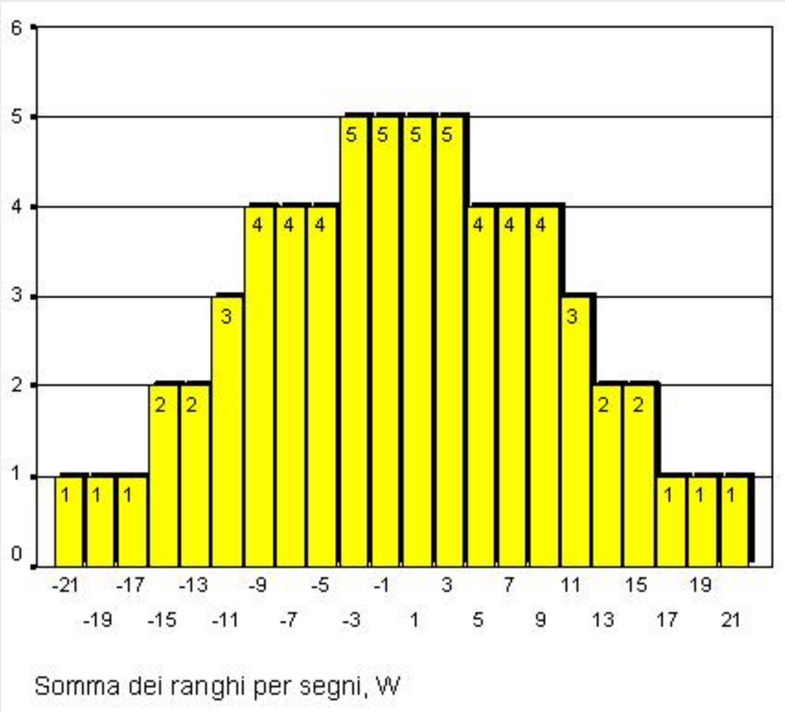
Per delimitare il confine fra "piccolo" e "grande" dobbiamo considerare tutte le possibili combinazioni di ranghi e calcolare per ognuna di esse W, in modo da identificare quali sono i suoi valori più "estremi" di rifiuto dell'ipotesi nulla di inefficacia del trattamento (per un dato il livello di significatività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

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

dove n rappresenta il numero delle unità statistiche in studio.

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

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 (α=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:

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