

CAPITOLO XIX

TEST NON PARAMETRICI PER IL TREND

19.1.	La media mobile e la scelta del test per la tendenza	1
19.2.	Il test di Cox e Stuart (e sue varianti) per il trend nella posizione e nella dispersione	6
19.3.	Test di Jonckheere o Jonckheere-Terpstra per alternative ordinate in k campioni indipendenti	18
19.4.	Test di Mack-Wolfe o ombrella test	33
19.5.	Il test di Page per alternative ordinate in k campioni dipendenti	52

CAPITOLO XIX

TEST NON PARAMETRICI PER IL TREND

19.1. LA MEDIA MOBILE E LA SCELTA DEL TEST PER LA TENDENZA

Anche nella ricerca biologica e ambientale e farmacologica, sono frequenti i casi in cui l'interesse è rivolto all'analisi di serie temporali di dati. In medicina può essere l'esame di un parametro rilevato varie volte nel tempo a un paziente; nella ricerca ambientale, il monitoraggio di un inquinante o di un indice generico di inquinamento; in agraria la quantità di una vitamina presente in un alimento, in funzione del tempo trascorso dalla produzione.

Gli stessi metodi in molti casi possono essere utilizzati anche per analizzare una serie spaziale, territoriale o geografica, cioè di rilevazioni effettuate a distanza crescente da un'origine convenzionale o da un punto qualsiasi. In geologia può essere la presenza di una sostanza in carote di terreno o di roccia, prelevate a profondità differente; nella ricerca ambientale la quantità di polveri in funzione della distanza dal camino che le emette oppure il livello d'inquinamento di un fiume lungo il percorso.

Per l'analisi statistica di una serie temporale o spaziale di dati, sono già stati presentati

- il **test delle successioni**, che valuta la casualità del susseguirsi di eventi binari,
- il test di **Kolmororov-Smirnov** sia per uno che per due campioni, che permette di valutare la differenza per un parametro qualsiasi tra due distribuzioni oppure tra una distribuzione osservata e quella teorica o attesa.

I test presentati in questo capitolo verificano tutti ipotesi sulle variazioni sistematiche della tendenza centrale, misurata di solito con la mediana trattandosi di test non parametrici; ma può essere l'evoluzione della media, della moda o un quantile qualsiasi. Per il test di Cox e Stuart è presentato anche il metodo per testare variazioni sistematiche nella variabilità o dispersione.

Disponendo di una serie di dati rilevati in successione temporale oppure a partire da un'origine spaziale, il primo problema che si pone è l'**individuazione dell'ipotetico andamento "normale"**, cioè delle quantità che dovrebbero aversi in un determinato momento o luogo, in assenza di eventi straordinari o di valori comunque distanti dal comportamento normale.

Una soluzione utilizzata da molti ricercatori è l'uso di una **interpolazione analitica**, consistente spesso in una **retta** oppure in una **curva di secondo** o di **terzo grado**. Ma essa, soprattutto se ricavata con metodi parametrici, richiede condizioni di validità che la distribuzione dei dati non sempre rispetta o che il tipo di scala utilizzato non permette. Inoltre, fondati su tutta la serie di osservazioni, questi metodi tendono ad eliminare in maniera troppo drastica sia i massimi che i minimi della curva reale,

cioè proprio quei valori che caratterizzano una zona o un periodo e che di norma lasciano tracce del loro effetto su un intervallo maggiore.

La soluzione può essere l'uso di **campioni mobili** (*moving sample*). Ad esempio, una sequenza campionaria di 80 dati può essere scomposta in un primo sottocampione, formato dai primi 10 dati. Il secondo sottocampione è sempre formato da 10 dati, ma iniziando dal secondo valore e terminando con l'undicesimo del campione; il terzo, prendendo i dati dal terzo al dodicesimo. Si prosegue fino a quando è possibile avere l'ultimo sottocampione di 10 dati, che ovviamente sarà quello formato dagli ultimi 10 dati del campione totale.

Ogni volta si ottiene un sottocampione diverso, seppure solo in parte minima rispetto a quelli vicini, che si muove lungo la sequenza e ha sempre dimensioni costanti. Di ogni sottocampione si calcola la media aritmetica. La successione delle medie di questo *moving sample* è la *moving average* o **media mobile**.

La **perequazione di una curva** mediante la **media mobile** ha il vantaggio, rispetto all'adattamento di una curva analitica di tipo matematico, di una minore rigidità: in genere, la curva perequata con la media mobile si adatta meglio alle particolarità della curva reale.

La **media mobile**

- **diminuisce** sensibilmente la **variabilità casuale dei singoli dati**
- e quindi **evidenzia la tendenza di fondo**, determinata appunto dalle variazioni sistematiche della quantità media,
- **mantenendo la variabilità delle medie di k dati**, quindi senza eliminarla totalmente.

Inoltre, è semplice **statistica descrittiva**; pertanto non richiede che siano realizzate le condizioni necessarie alla validità di una retta o una curva di regressione, quando calcolata con metodi parametrici.

Il punto fondamentale, per mantenere la variabilità più utile, è la scelta del numero **k** di dati, che formano ogni sottocampione mobile. Poiché il metodo è applicato in particolare quando esistono **fenomeni ciclici**, ma con una **grande variabilità tra osservazioni contigue**, il numero **k** è scelto in **funzione del periodo o ciclo** la cui variabilità si vuole mantenere, eliminando o riducendo molto quella entro il ciclo, spesso determinata dalle singole osservazioni.

A causa del metodo, per un'analisi dettagliata che evidenzi queste regolarità, occorre disporre di una serie abbastanza lunga di dati, comprendente più cicli.

Si supponga di disporre dei consumi giornalieri di acqua per uso domestico, rilevati per 4 settimane, in una località di villeggiatura, al fine di una valutazione corretta dell'incremento avvenuto nel periodo:

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°
I Settimana							II Settimana						
L	M	M	G	V	S	D	L	M	M	G	V	S	D
21	23	22	24	27	35	30	23	22	24	25	28	36	32

15°	16°	17°	18°	19°	20°	21°	22°	23°	24°	25°	26°	27°	28°
III Settimana							IV Settimana						
L	M	M	G	V	S	D	L	M	M	G	V	S	D
24	27	26	28	31	38	40	25	24	25	25	29	43	41

Le semplice lettura dei dati dei 28 gironi, organizzati in settimane, evidenzia massimi ciclici nei giorni di sabato e domenica, con minimi nei primi giorni della settimana.

La procedura di **calcolo della media mobile** è semplice:

- dopo individuato il numero (**k**) di elementi che caratterizzano il ciclo, si prendono i primi **k dati** e si calcola la **media centrata sul dato collocato esattamente a metà del periodo ciclico**; per questo, di regola, il numero **k** di elementi da considerare è dispari;
- la media successiva è stimata spostando di una unità tutto il gruppo di dati, in pratica eliminando il primo valore del sottocampione e aggiungendo il primo a destra: la media è centrata sul dato immediatamente successivo a quello precedente;
- si prosegue fino al valore che sta al centro dell'ultimo ciclo che è possibile stimare.

Ovviamente la serie totale raccolta deve comprendere più di un ciclo.

Ad esempio, i dati riportati nelle due tabelle successive evidenziano una fase settimanale entro la durata quasi mensile della rilevazione. E' quindi conveniente utilizzare una **media mobile a 7 elementi**, detta anche **centrata su 7 elementi**:

- il calcolo può iniziare solo dal 4° giorno:

$$(21 + 23 + 22 + 24 + 27 + 35 + 30) / 7 = 182 / 7 = 26,0$$

- si passa al calcolo della media mobile centrata sul 5° giorno:

$$(23 + 22 + 24 + 27 + 35 + 30 + 23) / 7 = 184 / 7 = 26,3$$

- e si prosegue fino al 25° giorno, l'ultimo possibile,

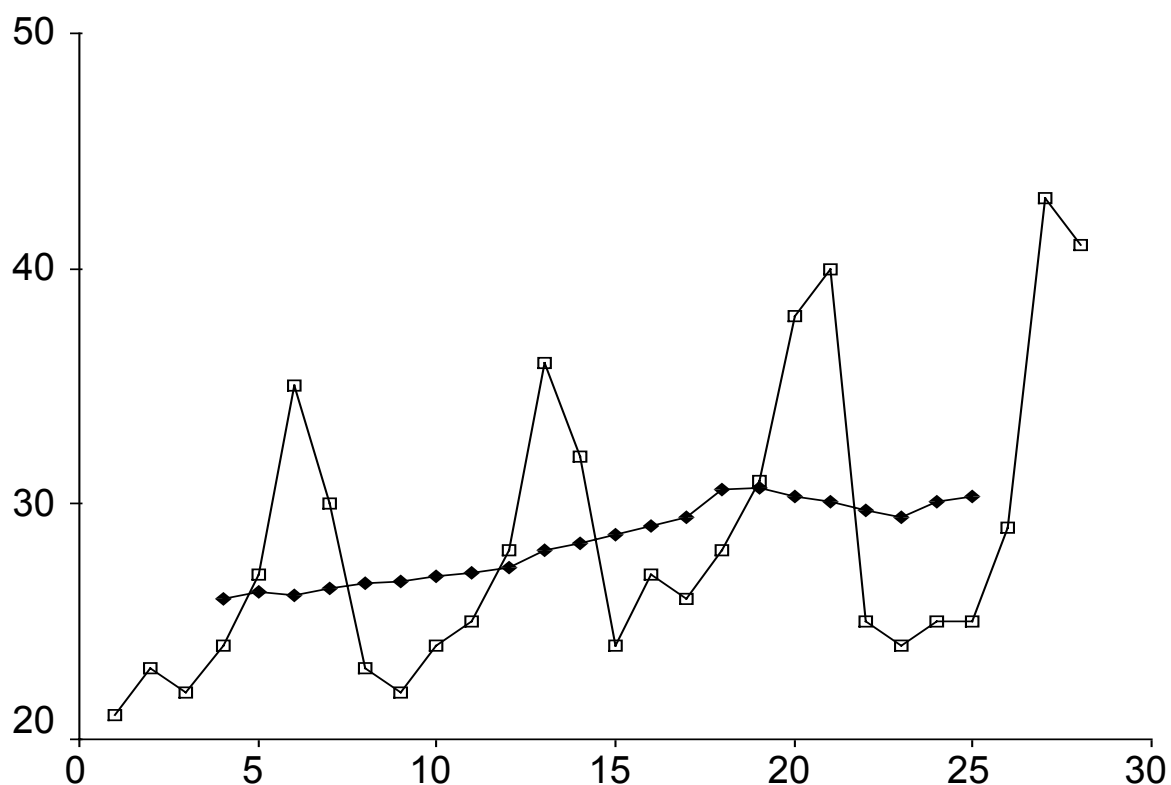
$$(25 + 24 + 25 + 25 + 29 + 43 + 41) = 212 / 7 = 30,3$$

Riportati in tabella, le medie calcolate determinano la serie seguente

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°
---	---	---	26,0	26,3	26,1	26,4	26,6	26,7	26,9	27,1	27,3	28,0	28,3

15°	16°	17°	18°	19°	20°	21°	22°	23°	24°	25°	26°	27°	28°
28,7	29,1	29,4	30,6	30,7	30,3	30,1	29,7	29,4	30,1	30,3	---	---	---

La loro **rappresentazione grafica**, sovrapposta a quella dei dati giornalieri, evidenzia come i picchi siano stati attenuati



Il trend di fondo è reso più evidente.

A differenza di quanto può evidenziare una retta, si nota come i consumi massimi siano avvenuti non alla fine del mese, ma nelle settimane centrate sul 18° (30,6) e sul 19° giorno (30,7).

Ma per rispondere alla domanda se durante il mese esaminato è stata rilevata una **variazione significativa in aumento** oppure **in diminuzione**, si deve ricorrere a test d'inferenza.

Essi possono essere

- **unilaterali**, quando il trend è ipotizzato a priori sulla base della conoscenza teorica del fenomeno o di esperienze precedenti, non sulla base delle variazioni indicate dai dati rilevati,
- **bilaterali**, quando si intende scoprire una non stazionarietà, cioè un aumento oppure una diminuzione, senza ipotesi più specifiche su uno dei due o ritenere un trend più logico o probabile dell'altro.

Tra i test di maggiore utilità sono da ricordare quelli chiamati **test di tendenza o per la significatività del trend**.

La loro applicazione classica è la verifica, ad esempio, dell'ipotesi che

- lungo il corso di un fiume si abbia una **tendenza monotona all'aumento (o alla diminuzione)** dei livelli d'inquinamento;
- in un periodo di cura, i parametri biologici di una o più persone aumentino o diminuiscano;
- la somministrazione di un farmaco o di un tossico determini effetti in crescita o in calo.

La raccolta dei dati può essere fatta, schematicamente, in tre modi diversi, ai quali corrispondono tre tipi di test diversi, per verificare le stesse ipotesi:

- il primo caso è quando sono stati fatte più di una decina di rilevazioni in tempi differenti: si applica il **test di Cox e Stuart**;
- il secondo gruppo di casi è quando sono state individuati stazioni o periodi, in ognuno dei quali sono state fatte alcune rilevazioni, riportate in una tabella come nell'analisi della varianza ad un criterio: si applica il **test di Jonckheere** se il trend può essere schematizzato in una regressione monotonica oppure il **test di Mack-Wolfe** se può essere rappresentato da una parabola;
- il terzo, quando nelle stazioni prefissate le analisi sono state condotte in modo ordinato, tali da poter essere rappresentate in una tabella come per l'analisi della varianza a due criteri di classificazione: si applica il **test di Page**.

Altri test non parametrici che è possibile utilizzare per la verifica del trend, quando i dati sono raccolti singolarmente, sono riportati nel capitolo sulla **correlazione** e sulla **regressione non parametrica**. Quando un parametro è il tempo oppure lo spazio, sono applicabili

- il **test di Daniel**, che ricorre alla correlazione sui ranghi e quindi dimostra una regressione monotonica,
- la **regressione lineare non parametrica**, quando si cerca una variazione lineare,
- la **regressione monotonica**, quando la variazione non è necessariamente lineare o costante.

19.2. IL TEST DI COX E STUART (E SUE VARIANTI) PER IL TREND NELLA POSIZIONE E NELLA DISPERSIONE.

Con un campione di dati, non importa se raccolti con regolarità costante oppure in modo variabile, ma

- in successione temporale,
- oppure disposti lungo un gradiente spaziale,

il test di Cox e Stuart (*Cox-Stuart trend test, the Cox-Stuart sign test for trend*) permette di verificare, per la variabile considerata e con due metodi distinti, se esiste **una tendenza monotona all'aumento oppure alla diminuzione della tendenza centrale e della variabilità**.

In termini più espliciti il test può essere utilizzato per verificare separatamente

- sia ipotesi sulla **tendenza centrale o posizione (*location*)**,
- sia ipotesi sulla **variabilità o dispersione (*dispersion*)**.

Per quanta riguarda **la tendenza centrale**, la metodologia più semplice di Cox e Stuart è del tutto **simile al test dei segni per due campioni dipendenti**, sia nel caso di piccoli che di grandi campioni.

Pure ricorrendo a una trasformazione binaria (che gli autori indicano con 1 e 0, ma che in molti testi si preferisce indicare con + e -), i dati originali dovrebbero essere rilevati con una **scala continua**, per non avere valori identici (o almeno pochi).

Per quanto riguarda la tendenza centrale, il metodo “breve” proposto congiuntamente da D. R. **Cox** dell'Università di Cambridge e Alan **Stuart** della London School of Economics nel 1955 (con l'articolo *Some quick tests for trend in location and dispersion*, pubblicato su **Biometrika**, Vol. 42, pp. 80-95) permette di **valutare** se

- nel **complesso dei dati**,
- tra i **valori iniziali** e quelli **finali**,
- esiste un **incremento** oppure una **diminuzione** significativi,
- pure in presenza di **ampie irregolarità** casuali o cicliche e di un **allontanamento** rilevante **dalla linearità**.

Il test non risulta significativo se i dati hanno una fase d'incremento e una fase altrettanto lunga di decremento o viceversa; è significativo solamente se una delle due è statisticamente prevalente; la verifica verte sulla esistenza di una **regressione monotonica**, cioè un variazione sistematica in aumento oppure in diminuzione che non è necessariamente lineare.

Nei testi internazionali di statistica non parametrica, questo metodo è riportato

- brevemente in quello di P. **Sprent** e N. C. **Smeeton** del 2001 *Applied nonparametric statistical methods* (3rd ed. Chapman & Hall/CRC, XII + 461 p.),
- in modo più esteso in quello di W. J. **Conover** del 1999 *Practical nonparametric statistics* (3rd ed., John Wiley & Sons, New York, VIII + 583 p.).

Spesso è collocato tra i **test per un campione** come sviluppo della **distribuzione binomiale**, sulla quale è fondato. Con tale impostazione, il parametro tempo o spazio non assumono alcuna importanza, purché i singoli dati siano riportati in serie, secondo l'ordine della rilevazione.

In altri manuali, è classificato tra i **test per dati bivariati**, insieme con la regressione lineare o monotonica e della correlazione non parametriche.

Con **N osservazioni indipendenti**

$$X_1, X_2, \dots, X_N$$

in serie ordinata rispetto a un'origine

- l'ipotesi nulla è che non esiste trend

H_0 : la tendenza centrale delle osservazioni da 1 a N è costante

- contro una delle ipotesi alternative,

H'_1 : la serie di osservazioni ha una tendenza centrale non costante (ipotesi bilaterale)

oppure

H''_1 : le osservazioni hanno una tendenza centrale in crescita (ipotesi unilaterale crescente)

oppure

H'''_1 : le osservazioni hanno una tendenza centrale in diminuzione (ipotesi unilaterale decrescente).

Il metodo proposto, estremamente semplice e **fondato sulla distribuzione binomiale**, richiede alcuni passaggi logici, spiegati con facilità attraverso la presentazione di un esempio.

Si supponga di avere rilevato la seguente serie di 12 misure (arrotondate all'unità), raccolte

- in **successione temporale**, come dodici analisi mensili
- o **geografica**, come un indice d'inquinamento lungo un corso d'acqua

17	16	14	17	21	22	21	23	22	18	10	23
----	----	----	----	----	----	----	----	----	----	----	----

per verificare se esiste un incremento significativo nella tendenza centrale dei valori.

In questo caso il **test è unilaterale**, con

H_0 : il valore delle osservazioni da 1 a N è costante o tende a diminuire

H_1 : il valore delle osservazioni è crescente

Per applicare e comprendere il test, è utile seguire i seguenti passaggi:

1 - **Individuare l'osservazione centrale, che corrisponde alla mediana** del tempo o dello spazio della serie di rilevazioni; con $N = 12$ è tra la sesta e la settima misura, quindi tra il valore 22 e il valore 21 del campione.

2 - Se il numero N d'osservazioni è pari (come nell'esempio), separare il campione in due gruppi uguali, ognuno con $N/2$ dati; ad esempio

- il primo gruppo dalla prima misura fino alla sesta comprese
- il secondo gruppo dalla settima misura alla dodicesima

Prima della mediana	17	16	14	17	21	22
Dopo la mediana	21	23	22	18	10	23

accoppiando la serie delle misure **antecedenti la mediana** con quella **ad essa successive**.

3 – In modo più dettagliato, abbinare

- il primo valore dei dati che precedono la mediana (1) al primo valore di quelli successivi alla mediana ($N/2 + 1$), cioè i valori 17 e 21,
- il secondo del primo gruppo (2) con il secondo del secondo gruppo ($N/2 + 2$), cioè i valori 16 e 23,
- fino all'ultimo dato (22 e 23) di ognuno dei due gruppi,
- segnando **il segno** delle $N/2$ differenze, come riportato nella tabella seguente:

Prima della mediana	17	16	14	17	21	22
Dopo la mediana	21	23	22	18	10	23
Segno della differenza	+	+	+	+	-	+

Si ottengono $N/2$ segni, ognuno dei quali può essere **positivo (+) o negativo (-)**, in dipendenza del fatto che tra le coppie di dati a confronto si è realizzato un aumento oppure una diminuzione.

4) Se è vera l'ipotesi nulla H_0 , questa serie di $N/2$ differenze avrà tendenza centrale nulla: in termini più discorsivi, i segni positivi e i segni negativi tenderanno ad avere la stessa frequenza.

Diversamente, avrà

- tendenza centrale positiva (maggioranza significativa di segni positivi), se la serie totale di rilevazioni ha valori crescenti;

- oppure una tendenza centrale negativa (maggioranza significativa di segni negativi), se la serie di rilevazioni ha valori decrescenti.

5) Se il numero **N** di osservazioni non è pari ma dispari, come nella serie successiva di 13 dati,

17	16	14	17	21	22	20	21	23	22	18	10	23
----	----	----	----	----	----	-----------	----	----	----	----	----	----

è necessario eliminare il valore **(N +1)/2**, cioè il settimo (in grassetto), corrispondente alla mediana del tempo o dello spazio.

Successivamente si devono costruire i due gruppi con lo stesso numero di dati e abbinare i valori come nel precedente punto 3; infine si calcolano le **(N-1)/2** differenze:

Prima della mediana	17	16	14	17	21	22
Dopo la mediana	21	23	22	18	10	23
Segno della differenza	+	+	+	+	-	+

6) Alla successione di segni positivi e negativi, si applica il **test dei segni** utilizzando

- nel caso di **campioni piccoli** la **distribuzione binomiale**,

$$P_{(r)} = C_N^r p^r q^{N-r}$$

in cui

- **N** = numero totale di segni
- **r** = numero del segno meno frequente;

Per non effettuare i calcoli e ottenere rapidamente il valore della probabilità, è vantaggioso utilizzare la distribuzione cumulata riportata nella pagina successiva.

PROBABILITA' CUMULATE DELLA DISTRIBUZIONE BINOMIALE

$$C_N^r \cdot p^r \cdot q^{N-r}$$

N = numero di osservazioni

r = numero minore tra segni positivi e negativi

N

r	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0.016	0.008	0.004	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.109	0.062	0.035	0.020	0.011	0.006	0.003	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000
2	0.344	0.227	0.144	0.090	0.055	0.033	0.019	0.011	0.006	0.004	0.002	0.001	0.001	0.000	0.000
3	0.656	0.500	0.363	0.254	0.172	0.113	0.073	0.046	0.029	0.018	0.011	0.006	0.004	0.002	0.001
4	0.891	0.773	0.637	0.500	0.377	0.274	0.194	0.133	0.090	0.059	0.038	0.024	0.015	0.010	0.006
5	0.984	0.938	0.856	0.746	0.623	0.500	0.387	0.291	0.212	0.151	0.105	0.072	0.048	0.032	0.021
6	1.00	0.992	0.965	0.910	0.828	0.726	0.613	0.500	0.395	0.304	0.227	0.166	0.119	0.084	0.058
7		1.00	0.998	0.980	0.945	0.887	0.806	0.710	0.605	0.500	0.402	0.314	0.240	0.180	0.132
8			1.00	0.996	0.989	0.967	0.927	0.867	0.788	0.696	0.598	0.500	0.407	0.324	0.252
9				1.00	0.999	0.994	0.981	0.954	0.910	0.849	0.773	0.686	0.593	0.500	0.412
10					1.00	0.999	0.997	0.989	0.971	0.941	0.895	0.834	0.760	0.676	0.588
11						1.00	1.00	0.998	0.994	0.982	0.962	0.928	0.881	0.820	0.748
12							1.00	1.00	0.999	0.996	0.989	0.976	0.952	0.916	0.868
13								1.00	1.00	1.00	0.998	0.994	0.985	0.968	0.942
14									1.00	1.00	1.00	0.999	0.996	0.990	0.979
15										1.00	1.00	1.00	0.999	0.998	0.994
16											1.00	1.00	1.00	1.00	0.999
17												1.00	1.00	1.00	1.00
18													1.00	1.00	1.00
19														1.00	1.00
20															1.00

Nel caso di **campioni grandi** per la rapidità del calcolo è conveniente utilizzare

- la **distribuzione normale**

$$Z = \frac{|X - N \cdot p| - 0,5}{\sqrt{N \cdot p \cdot q}}$$

con la correzione per la continuità (tra 30 e 100 dati almeno).

Poiché **l'analisi è condotta sulle N/2 differenze**, la serie di rilevazioni del campione deve essere il doppio di quanto necessario alla significatività della distribuzione binomiale, cioè **almeno 6 differenze**; di conseguenza, **serve una serie con almeno 12 osservazioni** per raggiungere la significatività, come è possibile dedurre dalla semplice lettura della tabella della distribuzione binomiale cumulata.

Nel caso in cui una o più **differenze risultino uguali a 0**, si eliminano tali osservazioni non conteggiandole né tra **N** né tra **r**.

E' il motivo per cui (come già evidenziato) si richiede che la scala, con la quale i valori o punteggi sono stimati, sia continua: i valori identici dovrebbero essere nulli o comunque molto rari. Ignorare i confronti con differenza nulla è un metodo semplice e diffuso, applicato in vari test in cui le risposte utilizzate sono di tipo binario. Il metodo solleva obiezioni da parte di alcuni statistici teorici, in quanto anche la risposta nulla contiene una certa quantità d'informazione; ma pure i metodi alternativi hanno aspetti criticabili e inoltre sono metodologicamente più complessi e lunghi, senza apportare miglioramenti effettivi alla potenza del test.

ESEMPIO 1. In un corso d'acqua che attraversa un centro di residenze turistiche, durante due mesi sono stati effettuati 20 prelievi ad intervalli irregolari, per verificare l'ipotesi di un aumento dell'inquinamento d'origine organica dovuto all'incremento delle presenze.

I valori raccolti che possono essere

- sia indicatori complessi o punteggi della presenza di sostanze inquinanti, quindi una scala di rango
- sia misure di concentrazione per litro, quindi una scala di rapporti ma con distribuzione non normale

sono stati disposti in tabella secondo l'**ordine temporale** dei prelievi

Prelievi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Indice	12	13	15	22	16	13	17	17	18	15	14	15	16	13	18	18	18	20	19	19

Con questi dati campionari, si può sostenere che esiste un aumento significativo nella concentrazione delle sostanze inquinanti, durante il periodo dei prelievi?

Risposta. E' un'ipotesi unilaterale crescente, con

H_0 : la posizione delle osservazioni da 1 a N è costante

e

H_1 : la serie dei valori dimostra un aumento della tendenza centrale

La stima della probabilità può essere ottenuta con alcuni passaggi:

1 - Separare la serie complessiva in

- valori prima del tempo mediano e
- valori dopo il tempo mediano,

mantenendo per ogni gruppo lo stesso ordine della serie iniziale

< Mediana del tempo	12	13	15	22	16	13	17	17	18	15
> Mediana del tempo	14	15	16	13	18	18	18	20	19	19
Segno della differenza	+	+	+	-	+	+	+	+	+	+

2 - Benché i valori siano approssimati, in questo esempio è possibile calcolare il segno di tutte le differenze: si ottengono 9+ e 1-

3 - Se fosse vera l'ipotesi nulla H_0 , la probabilità di trovare per caso questa risposta sommata a quella di risposte più estreme è data dalla somma delle relative **distribuzioni binomiali**:

$$P_{(9)} = C_{10}^9 0,5^9 0,5^1$$

e

$$P_{(10)} = C_{10}^0 0,5^{10} 0,5^0$$

Se la somma delle probabilità $P_{(9)} + P_{(10)}$ risulta inferiore a 0.05 si può rifiutare l'ipotesi nulla e concludere che è dimostrata una tendenza significativa all'aumento della variabile X.

Senza perdere tempo nel calcolo delle probabilità, **la tabella dei valori cumulati**

per $N = 10$ e $r = 1$ fornisce una probabilità complessiva $P = 0.011$,
che permette di rifiutare l'ipotesi nulla ed affermare che esiste una tendenza significativa all'aumento dell'inquinamento.

Se il **test** fosse stato **bilaterale**, sarebbe stato necessario moltiplicare questa probabilità per due: si sarebbe ottenuta una probabilità $P = 0.022$ che avrebbe ugualmente permesso di affermare che esiste una **variazione monotona** significativa (senza indicarne la direzione).

Questo **test per la tendenza centrale** ha **alcune varianti**, che possono essere scelte in rapporto alle caratteristiche della serie di osservazioni.

Le due varianti più diffuse sono:

- la suddivisione dei dati non in due ma in tre gruppi, confrontando tra loro solamente i due gruppi estremi mediante la medesima procedura già illustrata,
- la suddivisione dei dati in due gruppi, ma effettuando il confronto a coppie in modo differente: si confronta il primo valore del primo gruppo con l'ultimo del secondo gruppo; il secondo del primo gruppo con il penultimo del secondo gruppo, procedendo quindi con abbinamenti di dati dagli estremi verso l'interno.

Tra questi due metodi, il secondo può esser utile quando le variazioni cicliche sono importanti e non è possibile sovrapporre due cicli completi. Ma quello che ha applicazioni più generali è il primo, già proposto da Cox e Stuart. E' utile in particolare quando i valori sono stati raccolti in un numero dispari di cicli, come nella tabella seguente. In essa è riportato l'esperimento dell'esempio precedente sul livello d'inquinamento in una località di vacanza, ma in un periodo di rilevazioni giornaliere della durata tre settimane consecutive

Settimana	Giorni						
	L	M	M	G	V	S	D
I	21	23	22	24	27	35	30
II	23	22	24	25	28	36	32
III	24	27	26	28	31	38	40

A causa delle variazioni giornaliere, non ha senso dividere il gruppo in due. E' più utile confrontare il lunedì della prima settimana con quello della terza e procedere nello stesso modo per tutti i 7 giorni.

E' importante annotare che questa ciclicità non dovrebbe essere stabilita sulla base del campione di dati raccolti, poiché potrebbe essere casuale e quindi la scelta incidere fortemente sulla probabilità α stimata. L'eliminazione di un gruppo di dati deve avvenire sulla base di conoscenze già acquisite sulla ciclicità del fenomeno.

Confrontando la prima e la terza settimana,

Settimana	Giorni						
	L	M	M	G	V	S	D
I	21	23	22	24	27	35	30
III	24	27	26	28	31	38	40
Differenza	+	+	+	+	+	+	+

risulta che i segni + sono 7 e quelli - sono 0.

Sulla tabella della distribuzione binomiale cumulata con $N = 7$ e $r = 0$, è riportata una probabilità unilaterale $P = 0.008$. Se la domanda era di verificare un aumento del livello d'inquinamento durante tutto il periodo di tre settimane, la conclusione è che esiste un incremento altamente significativo.

Rispetto alla procedura standard di suddivisione in due periodi, questo metodo

- ha sempre lo **svantaggio** di eliminare una parte dei dati e quindi di **perdere informazione e potenza**,
- ma spesso offre il **grande vantaggio**, superiore allo svantaggio, di permettere **confronti omogenei** tra due gruppi di **osservazioni che inoltre distano un tempo maggiore**; quindi di favorire la costruzione di un test più significativo, perché il segno delle differenze tra le coppie di osservazioni risulta molto più stabile, meno influenzato dalle variazioni casuali, con prevalenza numerica più netta di un segno sull'altro.

ESEMPIO 2. (Tratto dall'articolo di Cox e Stuart per il caso di un campione grande). Si assuma di disporre della quantità annuale di pioggia (in pollici) a Oxford dal 1858 al 1952. Si è avuta una variazione significativa nella quantità media di pioggia durante questi 95 anni?

Risposta. Il numero di osservazioni è alto e quindi è vantaggioso suddividerlo in tre parti. Ma poiché 35 non è divisibile per 3, è conveniente

- fare il primo gruppo con le prime 32 osservazioni,
- scartare il gruppo centrale con 31 osservazioni e
- accoppiare le ultime 32 osservazioni con le prime.

Nella pubblicazione sono riportati i risultati dei 32 confronti, dai quali risultano 14 più e 18 meno (in realtà gli autori, pure chiamandolo test dei segni, utilizzano rispettivamente 1 e 0 e trovano una somma uguale a 14).

Il campione è abbastanza grande per utilizzare l'approssimazione alla normale, con la correzione per la continuità

$$Z = \frac{|X - \mu| - 0,5}{\sigma} = \frac{|X - Np| - 0,5}{\sqrt{Npq}}$$

dove

- X è la frequenza minore tra + e - (nell'esempio X = 14)
- Np è la media attesa (nell'esempio N = 32 e p = 0,5)
- Npq è la varianza di una proporzione (nell'esempio 32 x 0,5 x 0,5)

Di conseguenza, si ottiene

$$Z = \frac{|14 - 16| - 0,5}{\sqrt{32 \cdot 0,5 \cdot 0,5}} = \frac{1,5}{2,83} = 0,53$$

un valore $Z = 0,53$.

In una distribuzione bilaterale, poiché in questo caso a priori non è atteso né un aumento né una diminuzione, ma si vuole verificare solo se era intervenuta una variazione significativa, a $Z = 0,53$ corrisponde una probabilità $P = 0,596$. E' molto alta: non solo non permette di rifiutare l'ipotesi nulla, ma spinge a affermare che l'ipotesi nulla è vera. La quantità annuale di pioggia si è mantenuta costante durante il periodo analizzato.

Cox e Stuart nel loro articolo propongono anche un **metodo più complesso**, che utilizza il rango delle differenze, non solo il segno. Ha il vantaggio di utilizzare una quantità maggiore dell'informazione contenuta nelle differenze e quindi di essere più potente; ma per una quantità trascurabile, come stimano i due autori confrontando l'efficienza relativa secondo il metodo di Pitman. Soprattutto ha lo svantaggio di essere più lungo e di poter essere applicato solo con scale continue. Sebbene nell'articolo citato Cox e Stuart lo preferissero, è riportato molto raramente anche nei testi internazionali di statistica non parametrica. Il metodo illustrato, quello più semplice, offre vantaggi pratici rilevanti.

Per l'**analisi della variabilità o dispersione** (*sign tests for trend in dispersion*), il criterio dei confronti a coppie è identico; ma, per ottenere misure di variabilità, si richiede un numero nettamente maggiore di dati, dovendo calcolare la variabilità entro ogni sottoperiodo. E' stato discusso anche da Hans K. Ury nel 1966 con un articolo molto breve, di poche righe, *Large-Sample Sign Tests for*

Trend in Dispersion (pubblicato su **Biometrika**, Vol. 53, pp.289-291). E' una **analisi rapida della dispersione dei dati in una serie storica**, così presentata: *For quickly testing trend in dispersion (in a time series, for example,), Cox e Stuart (1955) have investigated several sign tests applied to the ranges of subset of k observations.*

Disponendo di una **serie larga di osservazioni**, occorre

1 - suddividere l'insieme degli N dati in **tanti sottogruppi con k dati** ognuno;

Cox e Stuart, dopo valutazione dei risultati, suggeriscono anche quale è la **scelta migliore per k**.

Dipende da N, secondo lo schema:

se $N \geq 90$ prendere $k = 5$

se $90 > N \geq 64$ prendere $k = 4$

se $64 > N \geq 48$ prendere $k = 3$

se $48 > N$ prendere $k = 2$

2 – Se **N** non è esattamente divisibile per **k**, **scartare la serie di valori in eccesso** che si trovano **attorno alla mediana** del periodo (o spazio) complessivo considerato.

3 – Per ogni **sottogruppo** di **k** osservazioni **calcolare la devianza**; è possibile utilizzare anche le varianze, ma è più rapido fermarsi al calcolo delle prime; inoltre, trattandosi di sottogruppi con lo stesso numero k di dati e effettuando confronti binari (maggiore o minore), il risultato è identico.

4 – Con il metodo ritenuto classico, dividere le **p** devianze (dove $p = N/k$) in due gruppi per effettuare confronti a coppie; se il numero **p** è dispari, scartare quello centrale.

5 – E' possibile utilizzare il metodo con tre gruppi, scartando quello centrale, non necessariamente delle stesse dimensioni dei due esterni, che invece devono avere lo stesso numero di gruppi sottogruppi.

6 - Contare il numero di segni positivi e negativi dei confronti a coppie, stimando la probabilità con la distribuzione binomiale o normale, in funzione del numero di segni.

ESEMPIO 3. (Tratto dall'articolo di Cox e Stuart). Come nell'esempio precedente, si assuma di disporre della quantità annuale di pioggia a Oxford dal 1858 al 1952. Tuttavia, in questo caso non si vuole verificare la quantità media ma la variabilità annuale. Si è avuta una variazione significativa nella variabilità della quantità di pioggia annuale durante questo periodo?

Risposta. E' un test bilaterale, non essendo espresse teorie sulla variabilità della quantità annuale di piogge. L'ipotesi nulla è che la variabilità sia rimasta costante, mentre l'ipotesi alternativa è che essa sia variata.

1 - Con 95 osservazioni, si possono fare sottogruppi di $k = 5$ dati (come suggerisce la tabella di Cox e Stuart); dalle quantità annuali (misurate in pollici), si ottengono esattamente 19 devianze ($95/5$) di altrettanti quinquenni, di seguito riportati secondo l'ordine temporale delle osservazioni

Quinquenni	1	2	3	4	5	6	7	8	9	10
Devianze	9,64	12,30	12,01	11,45	5,43	13,05	9,86	10,89	6,95	15,05

Quinquenni	11	12	13	14	15	16	17	18	19
Devianze	11,34	6,63	12,19	8,55	4,80	11,00	7,76	7,03	10,98

2 – Il numero 19 non è divisibile per 2; si potrebbe eliminare solo il valore centrale e fare 9 accoppiamenti.

Cox e Stuart nel loro articolo preferiscono eliminare i 5 centrali ottenendo i seguenti 7 confronti a coppie

Prime 7	9,64	12,30	12,01	11,45	5,43	13,05	9,86
Ultime 7	12,19	8,55	4,80	11,00	7,76	7,03	10,98
Differenze	+	-	-	-	+	-	+

3 – Con 3 + e 4 –, nella tabella della probabilità binomiale cumulata si cerca il valore di probabilità P per $N = 7$ e $r = 3$; risulta molto alto; inoltre, per un test bilaterale, deve essere raddoppiato.

4 – Cox e Stuart concludono affermando che benché vi sia una piccola indicazione (4– e 3+) che la variabilità tende a diminuire nel tempo, il test suggerisce che facilmente si tratta di una variazione casuale. Poiché anche con l'altro metodo giungono alle stesse conclusioni, ovviamente con probabilità non identiche ma molto simili, essi scrivono: *Thus although there is a slight indication that the dispersion decreases with time, both tests suggest that this could easily be a sampling fluctuation.*

Benché il metodo di Cox e Stuart sia apparentemente diverso e molto più breve, i concetti sono simili al test di correlazione non parametrica τ di Kendall. In questo ultimo, come è illustrato nel capitolo sulla correlazione non parametrica, per calcolare le precedenze i confronti sono tra ogni valore e tutti quelli successivi. Quindi questo test è meno potente e gli stessi autori stimano una potenza pari a circa

0,78. Ma è applicabile in una varietà di situazioni sperimentali molto più grande, permette di utilizzare informazioni più generali e è molto più robusto.

Questi concetti su una potenza minore ma una robustezza maggiore e la possibilità di applicazioni a situazioni reali più frequenti sono applicabili anche ai test successivi.

19.3. TEST DI JONCKHEERE O JONCKHEERE-TERPSTRA PER ALTERNATIVE ORDINATE IN K CAMPIONI INDIPENDENTI

Nel caso di **k campioni indipendenti**, con dati disposti come nell'analisi della varianza ad un criterio di classificazione, quando si suppone che i gruppi siano ordinati secondo il valore delle loro mediane, con il **test di Jonckheere** è possibile **verificare l'ipotesi se i vari campioni o gruppi abbiano tendenze centrali in accordo con la sequenza fissata a priori**.

E' chiamato anche test di **Jonckheere-Terpstra o delle alternative ordinate (*ordered alternatives*)**, in quanto proposto quasi contemporaneamente ed in modo indipendente da T. J. **Terpstra** nel 1952 (nell'articolo *The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking* pubblicato su **Indagationes Mathematicae** Vol. 14, pp. 327-333) e da A. R. **Jonckheere** nel 1954 (con un articolo intitolato *A distribution-free k-sample test against ordered alternatives* pubblicato su **Biometrika** vol. 41, pp. 133-145).

Del Department of Psychology, University College, London, A. R. Jonckheere introduce l'uso del suo test con questa presentazione. Un ricercatore, sulla base della sua esperienza o di una teoria, spesso è nella condizione di attendersi una serie ordinata (*rank order*) negli effetti di differenti trattamenti sperimentali. Si supponga, per esempio, di determinare gli effetti di differenti livelli di stress su test di abilità manuale, come il numero di errori commessi nelle quattro situazioni classificate di stress alto, medio, basso e minimo. Lo sperimentatore spera di dimostrare risultati significativamente differenti e che le persone con livello di stress più alto commettano effettivamente un numero di errori maggiore. Per una presentazione più analitica e per osservare l'uso di un linguaggio tecnicamente corretto nella introduzione di una pubblicazione scientifica che utilizzi il test, è utile leggere le parole dell'autore: *The experimenter hopes to refute the hypothesis that the four samples of data could be considered as randomly drawn from the same population. However, experience or theory lead him to expect that the number of errors made increases with the degree of stress (his alternative hypothesis). Hence he desires a test of significance especially sensitive to those differences which, while tending to reject the null hypothesis that the samples are all drawn from the same population, at the same time lends support to the specified alternative in question.*

L'**analisi della varianza** (ANOVA) e il **test non parametrico corrispondente** (il test di Kruskal-Wallis) **non permettono di rispondere a questa domanda**: il loro risultato è indipendente dall'ordine con il quale sono disposte le medie dei trattamenti. In questo caso, invece, serve una combinazione tra

- i concetti dell'analisi della varianza, con la quale condivide l'ipotesi nulla,
- e i concetti della correlazione non parametrica tra ordine osservato e ordine atteso dei vari gruppi a confronto, con la quale condivide l'ipotesi alternativa.

Ma per il tipo di scala utilizzato, quale l'intensità dello stress, non è possibile utilizzare la forma classica di analisi della regressione.

Quando si dispone solamente di due gruppi è possibile effettuare test bilaterali e test unilaterali. Con più gruppi, l'analisi della varianza permette solo ipotesi alternative bilaterali; **questo metodo offre l'opportunità di eseguire test con ipotesi alternativa unilaterale** anche nel caso di più gruppi.

Ha ovvie applicazioni nello studio di **serie temporali o spaziali**: *Suppose, for instance, that on each of n successive occasions, any one of k events may occur. Then we can test the hypothesis that the k events occur randomly in the series of n occasions, against the alternative that they tend to occur in a particular ordered time sequence.*

In modo più formale, l'**ipotesi nulla** H_0 è che le **mediane (me)** dei **k** gruppi a confronto siano **uguali**,

$$H_0: me_1 = me_2 = me_3 = \dots = me_k$$

mentre l'**ipotesi alternativa** H_1 è che esse siano **allineate in base al valore**, con un **ordine fissato a priori**,

- che può essere il seguente

$$H_1: me_1 \leq me_2 \leq me_3 \leq \dots \leq me_k$$

- oppure quello in direzione opposta.

Se si rifiuta l'ipotesi nulla, secondo vari testi recenti tra cui quello di W. J. **Conover** del 1999 (*Practical nonparametric statistics*, 3rd ed. John Wiley & Sons, New York, VIII + 584 p.) e quello di P. **Sprent** e N. C. **Smeeton** del 2001 (*Applied nonparametric statistical methods*, 3rd ed. Chapman & Hall/CRC, London, XII + 461 p.) si accetta che tra i valori centrali dei gruppi a confronto **esista almeno una ineguaglianza** in senso stretto (*where at least inequality is strict*).

Jonckheere in realtà scrive

$$H_1: me_1 < me_2 < me_3 < \dots < me_k$$

e afferma che se si rifiuta l'ipotesi nulla l'ordine è in accordo con quanto ipotizzato nella H_1 .

Il test richiede che i campioni siano misurati con una **scala ordinale o più precisa**, come quelle di intervallo oppure di rapporti, in modo da **non avere valori con lo stesso rango (ties)**. In realtà è accettato che ne sia **presente un numero limitato**.

Per una presentazione semplice dei dati e per facilitare l'interpretazione dei risultati, di norma i valori vengono disposti in una tabella identica a quella dell'**analisi della varianza ad un solo criterio di classificazione**:

GRUPPI						
1	2	3	...	j	...	k
$X_{1,1}$	$X_{1,2}$	$X_{1,3}$...	$X_{1,j}$...	$X_{1,k}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$...	$X_{2,j}$...	$X_{2,k}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$X_{i,1}$	$X_{i,2}$	$X_{i,3}$...	$X_{i,j}$...	$X_{i,k}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$X_{n1,1}$	$X_{n2,2}$	$X_{n3,3}$...	$X_{ni,j}$...	$X_{nk,k}$

Ma i **k** campioni a confronto sono **disposti in ordine progressivo**, in accordo con l'ipotesi alternativa formulata a priori sui valori delle mediane rispettive. E' errato osservare nei dati raccolti una certa successione di medie e quindi collocare i gruppi in quell'ordine, utilizzando il test per valutarne la significatività. Aumenta molto la probabilità α di commettere un errore di Tipo I, cioè di trovare una differenza che in realtà non esiste.

Il test **utilizza il metodo delle precedenze**; è un **procedimento lungo**, quando si confrontano più gruppi, poiché analizza tutte le loro combinazioni semplici 2 a 2; con k gruppi sono $k(k-1)/2$.

Per ogni coppia di gruppi, si deve contare il numero di precedenze, ossia quante volte ogni osservazione dell'i-esimo gruppo precede le osservazioni del gruppo j-esimo, secondo il procedimento già proposto nel 1947 da Mann e Whitney con il test U., con il quale è concettualmente collegato, anche nella presentazione di Jonckheere.

Nel caso di **due valori uguali (ties)**, il punteggio delle precedenze deve essere **aumentato di 0,5**.

L'applicazione del metodo a un esempio permette di fornire una spiegazione semplice e chiara.

Si supponga di dover analizzare 3 gruppi (A, B e C), composti rispettivamente 3, 4 e 3 osservazioni, come nella tabella sottostante,

1 - **in cui**, come primo passo, **i valori devono essere ordinati in modo crescente entro ogni gruppo**:

Gruppo A	Gruppo B	Gruppo C
7,5	9,1	10,1
8,9	9,9	15,4
12,3	14,3	16,2
---	18,2	---

Con 3 gruppi (A, B, e C) i confronti possibili sono 3 (A-B, A-C, B-C).

2 – Successivamente occorre **riportare il numero di precedenze relativo ad ogni dato**

	A-B	A-C	B-C
	4	3	3
	4	3	3
	2	2	2
	---	---	0
Somme	10	8	8

di un gruppo rispetto a tutti i dati del gruppo a confronto;

nella tabella precedente,

- il primo 4 del confronto A-B indica che il valore 7,5 del gruppo A precede nell'ordine naturale i 4 valori di B;
- il valore 2 (terza posizione) del confronto A-B indica che il valore 12,3 del gruppo A precede nell'ordine naturale 2 valori di B (14,3 e 18,2);
- mentre nel confronto B-C il quarto numero è 0, perché il valore 18,2 del gruppo B non precede nessun valore del gruppo C (essendo maggiore).

3 - Il terzo passo è **la somma di tutte le precedenze**, che fornisce la stima dell'**indice J**.

Con i dati dell'esempio, **J** risulta uguale a **26** ($10 + 8 + 8$).

4 - **Quando è vera l'ipotesi nulla, il valore di J tende a essere uguale a un valore medio μ_J** , determinato

- dal numero totale **N** e
- dal numero **n_i** di dati

- in ognuno dei **k** gruppi,
secondo la relazione

$$\mu_J = (N^2 - \sum_{i=1}^k n_i^2)/4$$

Per **piccoli campioni**, la distribuzione campionaria di **J** è fornita da tabelle, riportate nelle 3 pagine seguenti.

Nella prima colonna delle tabelle è riportato il numero di dati che formano i **k** campioni a confronto. L'ordine delle loro dimensioni è ininfluente.

Per esempio, con 3 gruppi in cui il primo ha 6 osservazioni, il secondo 8 ed il terzo 6, si usa la terza riga della terza tabella, che riporta i valori di **J** per **k** uguali a 6, 6, 8. (cioè, rispettivamente per la varie probabilità, i seguenti valori critici: 85, 90, 99, 103).

Si rifiuta l'ipotesi nulla alla probabilità prefissata, quando il valore calcolato è uguale o superiore (\geq) a quello critico riportato nelle tabelle.

Ritornando ai dati dell'esempio con 3, 3 e 4 dati, il valore critico

- alla probabilità $\alpha = 0.05$ è **26**
- alla probabilità $\alpha = 0.01$ è **29**.

Il valore calcolato, uguale a 26, permette di rifiutare l'ipotesi nulla alla probabilità $P \leq 0.05$.

In realtà la probabilità **P** è spesso inferiore a quanto indicato, poiché con pochi dati la distribuzione delle probabilità non varia in modo continuo: essa è calcolata mediante il calcolo combinatorio e il valore critico prescelto non è mai superiore al valore α prefissato.

Per **campioni grandi**, la distribuzione asintotica di **J** è approssimativamente normale, con la **media** μ_J già presentata

$$\mu_J = (N^2 - \sum_{i=1}^k n_i^2)/4$$

e **varianza** σ_J^2 uguale a

$$\sigma_J^2 = \frac{N^2 \cdot (2N + 3) - \sum_{i=1}^k (n_i^2 \cdot (2n_i + 3))}{72}$$

Valori critici di J nel test di Jonckheere

per 3 gruppi con osservazioni da 3 a 8

(PRIMA PARTE)

Dimensioni del campione	α			
	0.10	0.05	0.01	0.005
3 3 3	20	22	24	25
3 3 4	24	26	29	30
3 3 5	28	30	33	35
3 3 6	32	34	38	40
3 3 7	36	38	42	44
3 3 8	40	42	47	49
3 4 4	29	31	34	36
3 4 5	33	35	39	41
3 4 6	38	40	44	46
3 4 7	42	45	49	52
3 4 8	47	50	55	57
3 5 5	38	41	45	47
3 5 6	43	46	51	53
3 5 7	48	51	57	59
3 5 8	53	57	63	65
3 6 6	49	52	57	60
3 6 7	54	58	64	67
3 6 8	60	64	70	73
3 7 7	61	64	71	74
3 7 8	67	71	78	81
3 8 8	74	78	86	89

Valori critici di J nel test di Jonckheere
per 3 gruppi con osservazioni da 4 a 8 e da 5 a 8
(SECONDA PARTE)

Dimensioni del campione	α			
	0.10	0.05	0.01	0.005
4 4 4	34	36	40	42
4 4 5	39	41	45	48
4 4 6	44	47	51	54
4 4 7	49	52	57	60
4 4 8	54	57	63	66
4 5 5	44	47	52	55
4 5 6	50	53	58	61
4 5 7	56	59	65	68
4 5 8	61	65	71	75
4 6 6	56	60	66	69
4 6 7	62	66	73	76
4 6 8	68	73	80	83
4 7 7	69	73	81	84
4 7 8	76	80	88	92
4 8 8	83	88	97	100
5 5 5	50	54	59	62
5 5 6	57	60	66	69
5 5 7	63	67	73	76
5 5 8	69	73	80	84
5 6 6	63	67	74	77
5 6 7	70	74	82	85
5 6 8	77	81	89	93
5 7 7	77	82	90	94
5 7 8	85	89	98	102
5 8 8	92	98	107	111

Valori critici di J nel test di Jonckheere

da 3 a 6 gruppi con osservazioni da 2 a 5

(TERZA PARTE)

Dimensioni del campione	α			
	0.10	0.05	0.01	0.005
6 6 6	71	75	82	86
6 6 7	78	82	91	94
6 6 8	85	90	99	103
6 7 7	86	91	100	103
6 7 8	94	99	109	113
6 8 8	102	108	118	122
7 7 7	94	99	109	113
7 7 8	102	108	119	123
7 8 8	111	117	129	133
8 8 8	121	127	139	144
2 2 2 2	18	19	21	22
2 2 2 2 2	28	30	33	34
2 2 2 2 2 2	40	43	46	49
3 3 3 3	37	39	43	45
3 3 3 3 3	58	62	68	70
3 3 3 3 3 3	85	89	97	101
4 4 4 4	63	66	72	76
4 4 4 4 4	100	105	115	119
4 4 4 4 4 4	146	153	166	171
5 5 5 5	95	100	109	113
5 5 5 5 5	152	159	173	178
5 5 5 5 5 5	223	233	251	258

dove

- N = numero complessivo di dati raccolti
- n_j = numero di dati del j-esimo gruppo.

Poiché le alternative sono ordinate, come a lungo argomentato nella introduzione **il test** è sempre a **una coda**.

ESEMPIO 1 (PER GRANDI CAMPIONI)

Si intende verificare se le aree più vicine ad un forno inceneritore hanno una ricaduta maggiore di polveri e di sostanze inquinanti, con un gradiente non necessariamente lineare ma monotonicamente crescente, nonostante la variabilità nella direzione e nell'intensità dei venti. A questo scopo, è stata misurata la quantità di sostanze inquinanti ricadute in alcune aree e la loro distanza dalla sorgente di emissione.

L'area è stata suddivisa in 4 zone (A, B, C, D) concentriche, di raggio progressivamente maggiore passando dal gruppo A al gruppo D.

Per abbreviare e semplificare i calcoli,

- **i gruppi sono già stati ordinati** secondo l'ordine crescente atteso per le quantità mediane, espresso nell'ipotesi H_1

$$H_1 : me_D \leq me_C \leq me_B \leq me_A$$

- e i **valori** d'inquinamento raccolti in ogni zona sono **già ordinati in ordine crescente** entro ogni gruppo, come nella tabella seguente

ZONE				
	D	C	B	A
	12	28	31	35
	15	30	36	40
	18	38	39	52
	20	48	44	67
	38	60	54	78
	47	66	57	83
	48	70	63	88
	51	71	77	101
	90	--	87	119
	108	--	123	--
	--	--	124	--
n_j	10	8	11	9
	N = 38			

Per i calcoli successivi, servono

- i 4 valori di n_j (10, 8, 11, 9), corrispondenti al numero di osservazioni entro ogni gruppo e
- il loro numero totale $N = 38$

Risposta. Per verificare se la quantità di polveri ricadute è significativamente maggiore nelle aree presso la fonte d'emissione,

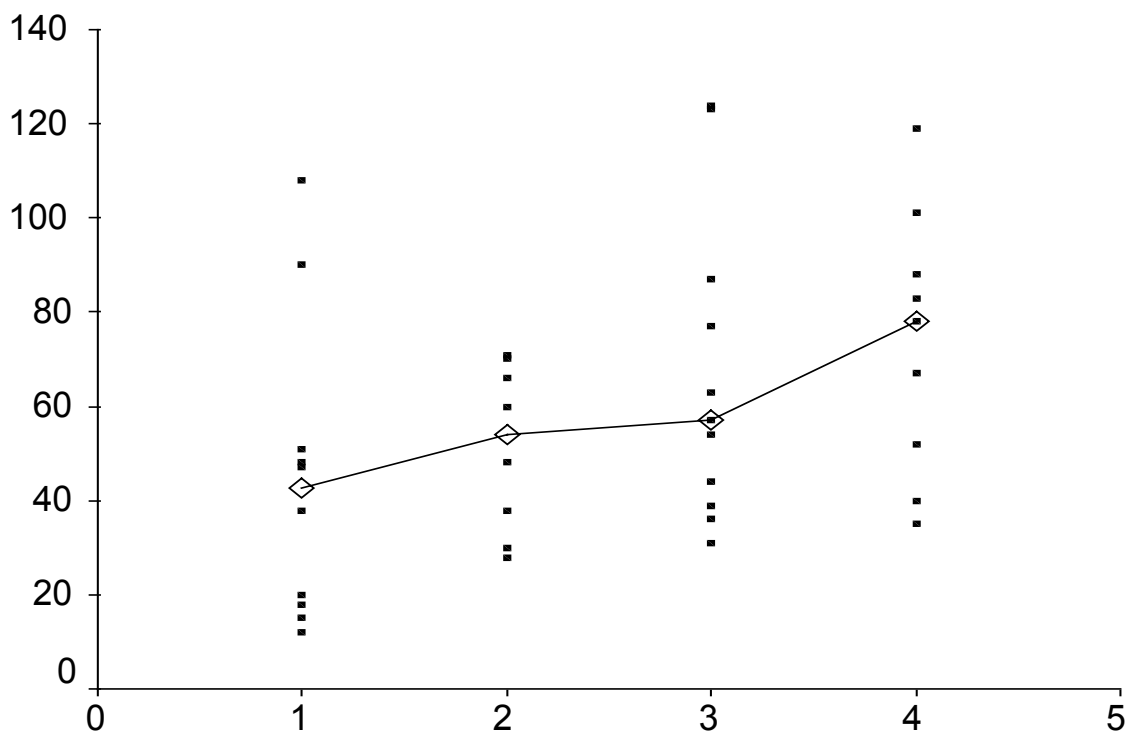
- l'**ipotesi nulla** H_0 è che all'aumento della distanza si abbia una quantità (θ) mediana costante

$$H_0 : \theta_A = \theta_B = \theta_C = \theta_D$$

- mentre l'**ipotesi alternativa** H_1 è che campioni raccolti in aree più distanti dall'inceneritore abbiano **quantità (θ) progressivamente minori** di polveri, anche se non è quantificabile il tipo di relazione (lineare, quadratica, ecc. ...)

$$H_1 : \theta_D \leq \theta_C \leq \theta_B \leq \theta_A$$

Per verificare le caratteristiche della distribuzione dei dati e per la successiva interpretazione dei risultati, è sempre utile costruire il grafico con le singole osservazioni di ogni gruppo e la loro mediana.



In questo caso, passando dalla zona 1 alla 4 (cioè da D a A) si evidenzia una tendenza alla crescita, se l'attenzione viene posta solo sulle mediane. Ma se si prende in considerazione tutto il campo di variazione l'impressione è meno netta. Il test permette una conclusione oggettiva, analizzando l'informazione contenuta nell'insieme dei dati.

Si osserva inoltre una forte variabilità tra i gruppi, minima nel secondo, che non permetterebbe l'uso di tecniche parametriche.

La metodologia di Jonckheere richiede il calcolo delle precedenze, confrontando tutte le possibili coppie di zone campionate; con 4 gruppi sono 6

D>C D>B D>A C>B C>A B>A.

come riportate nella tabella sottostante

CALCOLO DELLE PRECEDENZE						
	D>C	D>B	D>A	C>B	C>A	B>A
	8	11	9	11	9	9
	8	11	9	11	9	8
	8	11	9	9	8	8
	8	11	9	7	7	7
	5,5	9	8	5	6	6
	5	7	7	4	6	6
	4,5	7	7	4	5	6
	4	7	7	4	5	5
	0	2	2	--	--	3
	0	2	1	--	--	0
	--	--	--	--	--	0
U_{ij}	51	78	68	55	55	58

Nella prima colonna, il confronto **D>C** riporta 4 volte il punteggio 8: infatti i primi 4 valori del gruppo **D** sono minori degli 8 valori del gruppo **C**. Il punteggio 5,5 deriva dal fatto che il valore 38 del gruppo **D** precede 5 valori del gruppo **C** (48, 60, 66, 70,71) ed è appaiato dal valore 38 del gruppo **C**.

Anche da questo approccio,

- è evidente che i valori rilevati devono essere misurati in una scala continua per rendere minima la frequenza di dati appaiati;
- **nel caso di valori uguali, il punteggio aumenta di 0,5** dopo aver contato quanti dati sono superiori a quello di confronto.

Il valore di **J** è dato dalla somma di tutte le precedenze

$$J = 51 + 78 + 68 + 55 + 55 + 58 = 365$$

e risulta uguale a 365.

Se l'ipotesi nulla fosse vera, questo valore dovrebbe tendere alla media μ_j

$$\mu_j = \frac{N^2 - \sum n_j^2}{4} = \frac{38^2 - (10^2 + 8^2 + 11^2 + 9^2)}{4} = \frac{1444 - (100 + 64 + 121 + 81)}{4} = \frac{1078}{4} = 269,5$$

che, con i dati del campione, risulta uguale a 269,5

mentre la sua varianza σ_j^2

$$\sigma_j^2 = \frac{N^2 \cdot (2N + 3) - \sum n_j^2 \cdot (2n_j + 3)}{72} =$$

$$= \frac{38^2 \cdot (76 + 3) - (10^2 \cdot (20 + 3) + 8^2 \cdot (16 + 3) + 11^2 \cdot (22 + 3) + 9^2 \cdot (18 + 3))}{72} =$$

$$= \frac{114076 - (2300 + 1216 + 3025 + 1701)}{72} = \frac{105834}{72} = \mathbf{1469,9}$$

risulta uguale a 1469,9

e quindi la deviazione standard σ_j

$$\sigma_j = \sqrt{1469,9} = \mathbf{38,34}$$

è uguale a 38,34.

Da questi dati deriva

$$Z = \frac{J - \mu_j}{\sigma_j} = \frac{365 - 269,5}{38,34} = \frac{95,5}{38,34} = 2,49$$

un valore di $Z = 2,49$.

Ricordando che **nella distribuzione normale per un test unilaterale al valore di Z uguale a 2,49 corrisponde una probabilità P = 0.006**, si rifiuta l'ipotesi nulla.

Di conseguenza, si accetta l'ipotesi alternativa: la tendenza centrale dei 4 gruppi è significativamente maggiore per le aree più vicine alla fonte d'inquinamento.

Il test di Jonckheere è utile soprattutto nel caso del **confronto tra rapporti o percentuali**, quando sono calcolati su dimensioni di **campioni di dimensioni sensibilmente differenti**. Con tali dati infatti non è possibile applicare i test parametrici, poiché i valori a confronto non hanno tutti la stessa attendibilità. La trasformazione delle percentuali mediante arcoseno (vedi paragrafo sulle trasformazioni) serve solamente per rendere omogenee le varianze di gruppi che hanno medie differenti; ma per test parametrici le percentuali a confronto debbono essere calcolate su totali simili, come già ripetutamente ricordato in vari esempi.

ESEMPIO 2 (CAMPIONE PICCOLO).

Nei test di tossicità cronica sono usate varie specie acquatiche, in particolare Ciprinidi, Salmoidi, Cladoceri e Ditteri. Gli animali vengono esposti a sostanze tossiche per tempi generalmente lunghi, al fine di consentire che siano sottoposti all'azione del principio attivo non solo gli adulti, ma anche gli stadi larvali e giovanili.

Tra le stime di tossicità, è frequente l'impiego

- del rapporto tra il numero di individui sopravvissuti sul totale degli immessi,
- la percentuale di femmine che arrivano alla riproduzione,
- la quota di uova che si schiudono.

Con una serie di esperimenti, è stato calcolato il numero di uova rimaste sterili su quelle deposte, per stimare l'effetto di una sostanza ritenuta tossica sulla loro schiusa.

DOSE		
1	2	3
9/84	100/108	8/52
4/35	12/91	7/38
3/22	9/57	8/41
5/76	9/125	3/41
5/29	12/64	15/91
2/15	---	11/86
11/87	---	---

Si intende verificare se all'aumento della dose (1, 2, 3) si ha una quota crescente di uova sterili (misurate con x/y , dove x è il numero di uova sterili e y è il numero totale di uova deposte).

Risposta. Per verificare l'ipotesi nulla

$$H_0 : \theta_1 = \theta_2 = \theta_3$$

mentre l'**ipotesi alternativa** H_1 è

$$H_1 : \theta_1 \leq \theta_2 \leq \theta_3$$

1- la prima operazione consiste nel calcolare i rapporti e riportare i loro valori

DOSE		
1	2	3
0,107	0,093	0,153
0,114	0,132	0,184
0,136	0,158	0,195
0,066	0,072	0,073
0,172	0,187	0,165
0,133	---	0,128
0,126	---	---

2 - Successivamente si devono ordinare i valori entro ogni gruppo in modo crescente, per facilitare le operazioni di confronto tra i gruppi

DOSE		
1	2	3
0,066	0,072	0,073
0,107	0,093	0,128
0,114	0,132	0,153
0,126	0,158	0,165
0,133	0,187	0,184
0,136	---	0,195
0,172	---	---

3 - Con 3 gruppi, si hanno 3 combinazioni di confronti a coppie:

1 contro ($>$) 2,

1 contro ($>$) 3,

2 contro ($>$) 3.

Per ognuna di esse si devono contare le precedenze, che sono riportate, insieme con la somma per colonna e il totale, nella tabella sottostante

	1>2	1>3	2>3
	5	6	6
	3	5	5
	3	5	4
	3	5	3
	2	4	1
	2	4	--
	1	2	--
R _j	19	31	19

4 - Il valore di **J**, dato dalla somma di tutte le precedenze,

$$\mathbf{J} = 19 + 31 + 19 = 69$$

risulta uguale a 69.

Il valore di **J** tabulato per 3 gruppi con 5, 6 e 7 osservazioni alla probabilità $\alpha = 0.05$ è 74.

Il valore di J calcolato (69) è inferiore a quello tabulato; di conseguenza, non si può rifiutare l'ipotesi nulla. All'aumentare della dose di sostanza inquinante, con questi dati non è dimostrato che si abbia un aumento significativo della proporzione di uova sterili.

In letteratura si trovano **applicazioni del test di Jonckheere-Terpstra**

- **a campioni estremamente piccoli,**
- **con gruppi in cui compare addirittura una sola osservazione.**

Soprattutto l'ultima è una condizione che **vieta l'uso del test parametrico** F di Fisher.

ESEMPIO 3. (UN GRUPPO CON UN SOLO DATO) Una dimostrazione dell'uso del test con un solo dato in un gruppo è l'articolo pubblicato nel 1989 su **Applied Statistics** (vol. 38, pp. 495-506) di D. V. **Hinkley** (*Modified profile likelihood in trasformed linear models*) e riportato anche dal testo di P. **Sprenst** del 1993 *Applied Nonparametric Statistical Methods* (2nd ed., Chapman & Hall, London) e nella edizione successiva scritta con N. C. **Smeeton** nel 2001.

Per valutare l'aumento delle rotture meccaniche alla crescita della velocità (20, 25, 30 35 miglia orarie), è presentata l'analisi dei seguenti dati, con solamente 10 osservazioni suddivise in 4 gruppi e dei quali uno con una sola osservazione.

Velocità (mph)			
20	25	30	35
48	33 59 48 56	60 101 67	85 107

L'applicazione del test di Jonckheere dimostra che con questi dati è possibile rifiutare l'ipotesi nulla al livello di significatività $P = 0,0011$. La probabilità è stata ricavata con una distribuzione esatta di permutazione dei ranghi, che è proposta nei pacchetti statistici di molti programmi informatici.

19.4. TEST DI MACK-WOLFE O UMBRELLA TEST.

Uguale al test di **Kruskal-Wallis** e al test di **Jonckheere-Terpstra**

- sia per il **tipo di dati**, riportati in tabelle come nell'ANOVA a un criterio,
- sia per l'**ipotesi nulla**

$$H_0: me_1 = me_2 = \dots = me_p = \dots = me_K$$

il test proposto da Gregory. A. **Mack** e Douglas. A. **Wolfe** nel 1981 (con l'articolo *K-sample Rank Tests for Umbrella Alternatives*, pubblicato su **Journal of the American Statistical Association** Vol. 76, pp. 175-181), si distingue per l'ipotesi alternativa

$$H_1: me_1 \leq me_2 \leq \dots \leq me_{p-1} \leq me_p \geq me_{p+1} \geq \dots \geq me_{k-1} \geq me_K$$

con **almeno una ineguaglianza prima e dopo il valore massimo** (*where at least inequality is strict*). In termini discorsivi, l'ipotesi alternativa afferma che **esiste un punto di picco entro le classi estreme**.

I due autori lo chiamano *umbrella alternatives test*, un nome fantasioso (tradotto in italiano con **test dell'ombrello**). Essi motivano tale scelta con la figura che assume la rappresentazione grafica della successione delle mediane (*because of the configuration of the corresponding population medians*).

L'ipotesi alternativa **H₁** può essere anche nella **direzione opposta**, per verificare se la serie delle mediane ha una **forma concava**, cioè un **valore significativamente minore nelle classi centrali**

$$H_1: me_1 \geq me_2 \geq \dots \geq me_{p-1} \geq me_p \leq me_{p+1} \leq \dots \leq me_{k-1} \leq me_K$$

con **almeno una ineguaglianza prima e dopo il valore minimo**.

E' utile ricordare che

- nel test di **Kruskal-Wallis** l'ipotesi alternativa è bilaterale

$$H_1: \text{non tutte le me sono uguali}$$

- mentre nel test di **Jonckheere-Terpstra** è unilaterale (le mediane tendono a crescere),

$$H_1: me_1 \leq me_2 \leq \dots \leq me_{p-1} \leq me_p \leq me_{p+1} \dots \leq me_{k-1} \leq me_K$$

con almeno una ineguaglianza tra valore iniziale e valore finale.

Il test di **Mack-Wolfe** può essere visto come una generalizzazione del test di **Jonckheere-Terpstra**, nel quale la diseguaglianza è solamente tra valore iniziale e finale. Inoltre, mentre il test di Jonckheere-Terpstra può essere interpretato come analogo ad una regressione lineare o a una curva di secondo grado con una crescita e una successiva tendenza asintotica, il test di **Mack-Wolfe** è analogo a una **regressione quadratica**, con forma a parabola esprimibile dalla funzione di secondo grado

$$y_i = a + bx_i + cx_i^2 + e_i$$

Questo metodo di inferenza statistica è utile, ad esempio, per valutare

- se la dose crescente di un farmaco determina effetti prima positivi e poi negativi quando si supera il valore di soglia;
- se la quantità di un fertilizzante prima aumenta la produzione agricola e oltre certe dosi la riduce;
- se la presenza di ozono nell'aria o di un inquinante nei corpi idrici è massima nei mesi estivi;
- quale è l'effetto dell'età nella risposta a certi stimoli biologici;
- l'evoluzione degli effetti di un farmaco nelle ore successive alla sua somministrazione.

Mack e Wolfe nell'articolo di presentazione del test insistono su un concetto, che ritengono molto importante, per gli effetti non quantificabili che ha sulle probabilità α stimate, cioè sulla probabilità di rifiutare l'ipotesi nulla quando in realtà essa è vera: l'ipotesi alternativa **non può essere costruita o espressa sulla base dell'osservazione dei dati.**

L'ipotesi alternativa, cioè la scelta tra

- **test di Kruskal-Wallis**
- **test di Jonckheere-Terpstra**
- **test di Mack-Wolfe**

deve essere **formulata solamente a priori**, sulla base di una conoscenza teorica e quindi in anticipo rispetto alla raccolta dei dati: "... *using any a priori information we might have about the alternative (to H_0) that is of primary importance*".

Per l'uso che abitualmente viene fatto dei test statistici in molti settori della gestione e interpretazione scientifica dei dati, cioè per la loro applicazione a dati già raccolti eventualmente da altri, è necessario insistere su questo concetto. Se un ricercatore scrive che **i dati raccolti hanno evidenziato** un minimo (o massimo) collocato tra i gruppi estremi e che il successivo test è risultato significativo, evidenzia un approccio errato. Egli deve:

- esprimere l'ipotesi per cui è ragionevole (in campo, biologico, ambientale, medico, ecc. ...) attendersi che tra le mediane esista una regressione di tipo quadratico oppure lineare o infine una semplice differenza tra valori non ordinati;
- descrivere come, per verificare la sua ipotesi, ha raccolto i dati;
- riportare il risultato del test a supporto della sua conoscenza del problema e come dimostrazione empirica che è stato trovato l'effetto atteso.

In quasi tutti i programmi informatici è riportato il test di **Kruskal-Wallis**; in molti anche quello di **Jonckheere-Terpstra**; questo il test di **Mack-Wolfe** è spesso ignorato, nonostante la frequenza dei casi che ne richiedono l'uso nella ricerca applicata.

Tra i test internazionali di statistica, il test è riportato nel volume di Myles **Hollander** e Douglas A. **Wolfe** (coautore dell'articolo citato e professore nel Department of Statistics della Ohio State University) del 1999 **Nonparametric Statistical Methods** (2nd ed. John Wiley & Sons, New York, 787 p). La metodologia descritta in questo paragrafo ne segue le indicazioni (per ulteriori approfondimenti si rimanda al testo citato).

L'applicazione di questo metodo richiede una distinzione preliminare, dalla quale dipendono due procedure differenti: se **il punto di picco** (massimo o minimo)

- è noto
- è ignoto.

CON PICCO NOTO

La metodologia è spiegata in modo semplice, illustrando un esempio in tutti i suoi passaggi logici.

1 - Si supponga di voler verificare se la quantità di ozono raggiunge il massimo durante i mesi più caldi, individuati nel periodo luglio – agosto. A questo scopo, le quantità medie rilevate in un anno (poche osservazioni per motivi pratici di presentazione del metodo) sono state raggruppate in periodi di due mesi, ottenendo la seguente distribuzione

Indici di inquinamento idrico durante un anno					
(1) Gen. – Feb.	(2) Mar. – Apr.	(3) Mag. – Giu.	(4) Lug. – Ago.	(5) Set. - Ott.	(6) Nov. – Dic.
16,1	19,8	24,5	33,9	24,4	11,7
13,5	7,3	35,4	34,1	14,2	3,6
6,7	24,1	28,3	46,7	15,6	10,8
15,8		35,9	26,7	15,4	
10,2		29,9			
11,3					
15,4					
$n_1 = 7$	$n_2 = 3$	$n_3 = 5$	$n_4 = 4$	$n_5 = 4$	$n_6 = 3$

con il numero (n_i) di osservazioni per periodo riportate nell'ultima riga.

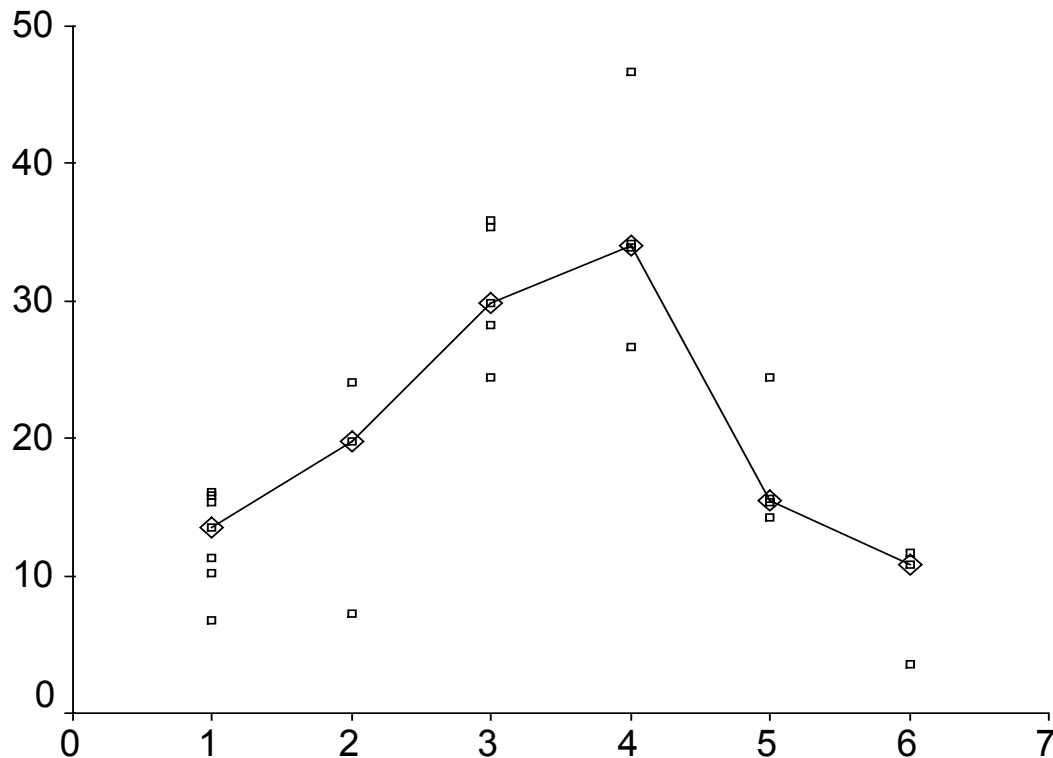
2 - Il test serve per verificare l'ipotesi nulla

$$H_0: me_1 = me_2 = me_3 = me_4 = me_5 = me_6$$

contro l'ipotesi alternativa

$$H_1: me_1 \leq me_2 \leq me_3 \leq me_4 \geq me_5 \geq me_6$$

Il periodo di massimo è stato **preventivamente individuato** nel gruppo 4, corrispondente ai mesi di luglio – agosto.



Come già suggerito in più occasioni, la rappresentazione grafica dei dati è sempre di estrema utilità per comprendere

- la forma di distribuzione dei dati,
- la successione delle mediane e quindi la variazione della tendenza centrale.

Il grafico mostra che l'ipotesi appare verosimile.

Il test serve appunto per verificare se tale impressione visiva è supportata da una interpretazione scientifica dei dati, cioè dalla probabilità di avere trovato per caso la distribuzione espressa nell'ipotesi H_1 , qualora H_0 fosse vera

3 - Come già illustrato nel test di Jonckheere, questo metodo utilizza il calcolo delle precedenze stimandone il numero tra coppie di gruppi.

Di conseguenza, per i conteggi successivi è utile, seppure non obbligatorio, ordinare i dati per rango entro ogni gruppo:

Indici di inquinamento idrico durante un anno (ordinati entro gruppo)					
(1) Gen. – Feb.	(2) Mar. – Apr.	(3) Mag. – Giu.	(4) Lug. – Ago.	(5) Set. - Ott.	(6) Nov. – Dic.
6,7	7,3	24,5	26,7	14,2	3,6
10,2	19,8	28,3	33,9	15,4	10,8
11,3	24,1	29,9	34,1	15,6	11,7
13,5		35,4	46,7	24,4	
15,4		35,9			
15,8					
16,1					

4 - Definiti

- il numero di gruppi, in questo caso $k = 6$,
- la posizione che individua il massimo atteso, in questo caso $p = 4$

il metodo proposto sostanzialmente applica due volte il test di Jonckheere:

- dal gruppo **1** al gruppo **p** (nell'esempio dal gruppo 1 al 4)
- dal gruppo **p** al gruppo **k** (nell'esempio dal gruppo 4 al 6)
- ricordando che **p** è preso in considerazione due volte e che **nella seconda parte la stima delle precedenze deve essere nell'ordine opposto** (per mantenerla sempre crescente).

Si ottengono le seguente 9 stime di precedenze, per confronti tra coppie di gruppi:

- le prime 6, che sono prodotte dalle combinazioni da 1 a 4
- le altre 3, prodotte dalle combinazioni da 6 a 4

Valori di U per confronti tra coppie di gruppi e calcolo di A									
1	2	3	4	5	6	7	8	8	Tot
U_{12}	U_{13}	U_{14}	U_{23}	U_{24}	U_{34}	U_{54}	U_{64}	U_{65}	A_4
1	7	7	3	3	1	4	3	3	
7	7	7	3	3	3	4	3	3	
7	7	7	3	3	3	4	3	3	
	7	7	3	3	5	4	3	3	
	7		3						
15	35	28	15	12	12	16	12	12	157

Come già con la metodologia di **Jonckheere**,

- nella prima colonna (U_{12}), il valore 1 è determinato dall'osservazione che, ordinando per ranghi insieme il gruppo 1 e il gruppo 2, il dato 7,3 del gruppo 2 è preceduto solamente dal primo valore (6,7) del gruppo 1; gli altri due 7 sono determinati dall'osservazione che i restanti due dati del gruppo 2 sono preceduti entrambi dai 7 valori del gruppo 1: il totale della colonna è $U_{12} = 7$;
- nell'ultima colonna (U_{65}), il valore 3 ripetuto quattro volte è determinato dall'osservazione che, ordinando per ranghi insieme il gruppo 6 e il gruppo 5, i quattro dati del gruppo 5 (cioè 14,2 15,4 15,6 24,4) sono tutti preceduti in una distribuzione congiunta, ordinata in modo crescente, dai tre dati del gruppo 6 (cioè 3,6 10,8 11,7); il loro totale è $U_{65} = 7$;

5 - Sommando i risultati delle 9 colonne si ottiene

$$A_p = \sum U_{ij}$$

$$A_4 = 15 + 35 + 28 + 15 + 12 + 12 + 16 + 12 + 12 = 157$$

Il numero di precedenze, centrate sulla colonna 4, è $A_4 = 157$.

Se per la serie ordinata delle mediane fosse stato **ipotizzato non un andamento convesso ma concavo**, la stima dei valori di U avrebbe dovuta essere condotta in modo opposto, calcolando

- non

U_{12}	U_{13}	U_{14}	U_{23}	U_{24}	U_{34}	U_{54}	U_{64}	U_{65}
----------	----------	----------	----------	----------	----------	----------	----------	----------

- ma

U_{21}	U_{31}	U_{41}	U_{32}	U_{42}	U_{43}	U_{45}	U_{46}	U_{56}
----------	----------	----------	----------	----------	----------	----------	----------	----------

In modo più esplicito

- l'indice **12** indica quante volte ogni valore del gruppo **2** è **preceduto** da valori del gruppo **1**
- l'indice **21** indica quante volte ogni valore del gruppo **1** è **preceduto** da valori del gruppo **2**

6 - Se è vera l'ipotesi nulla,

il valore di A_p tende al valore medio (μ_{Ap})

$$\mu_{Ap} = \frac{N_1^2 + N_2^2 - \sum_{i=1}^k n_i^2 - n_p^2}{4}$$

Con

- N_1 = numero totale di osservazioni dal gruppo **1** al gruppo **p**; nell'esempio: $7 + 3 + 5 + 4 = 19$
- N_2 = numero totale di osservazioni dal gruppo **p** al gruppo **k**; nell'esempio: $4 + 4 + 3 = 11$
- n_i = numero di osservazioni in ognuno dei **k** gruppi; nell'esempio: 7, 3, 5, 4, 4, 3
- n_p = numero di osservazioni nel gruppo **p**; nell'esempio: 4

si ottiene

$$\mu_{A4} = \frac{19^2 + 11^2 - (7^2 + 3^2 + 5^2 + 4^2 + 4^2 + 3^2) - 4^2}{4} = \frac{361 + 121 - 124 - 16}{4} = \frac{342}{4} = 85,5$$

il valore medio $\mu_{A4} = 85,5$.

7 - Se invece è vera l'ipotesi alternativa **H₁** espressa a priori, il valore di **A_p** tende a crescere e a divenire significativamente maggiore di quello medio stimato. Infatti il numero di precedenze, calcolato come richiesto dal metodo, è massimo se i valori maggiori sono posti esattamente nell'ordine ipotizzato.

In questo calcolo delle precedenze, è importante osservare che **se il picco indicato a priori non corrisponde realmente al gruppo che contiene le osservazioni di rango maggiore**, il numero di precedenze totale (**A_p**) si riduce notevolmente e **il test difficilmente risulta significativo**.

8 – Nel caso di **piccoli campioni**, (k che varia da 4 a 6 con n che varia da 2 a 5) il valore **A_p** calcolato deve essere confrontato con quelli critici riportati nella tabella.

Nella sua lettura è importante ricordare che la stima dei valori critici non è influenzata dal fatto che il gruppo di picco sia simmetrico rispetto a quello riportato.

Se il numero di gruppi è $k = 4$, il gruppo di picco **p** può essere indifferentemente **3 o 2**;

se il numero di gruppi è $k = 5$, il gruppo di picco **p** può essere indifferentemente **4 o 2**;

se il numero di gruppi è $k = 6$, il gruppo di picco **p** può essere indifferentemente **4 o 3** oppure **5 o 2**.

Ovviamente il valore di picco non può essere in uno dei due gruppi estremi, altrimenti il test diventerebbe quello di **Jonckheere-Terpstra**.

Il **caso di 3 gruppi** ($k = 3$) è particolare; pertanto è stato trattato a parte (nelle pagine successive).

VALORI CRITICI (A_p) PER IL TEST DI MACK-WOLFE
CON PICCO NOTO

- k = Numero di gruppi
- p = Rango del picco
- n = Numero di dati per gruppo; sono considerati solamente campioni bilanciati

$k = 4$			$k = 5$			$k = 6$		
n	$p = 3$		n	$p = 3$		n	$p = 4$	
2	$\alpha=.10$	13	2	$\alpha=.10$	19	2	$\alpha=.10$	26
	$\alpha=.05$	14		$\alpha=.05$	20		$\alpha=.05$	28
	$\alpha=.01$	16		$\alpha=.01$	22		$\alpha=.01$	31
3	$\alpha=.10$	27	3	$\alpha=.10$	38	3	$\alpha=.10$	54
	$\alpha=.05$	29		$\alpha=.05$	41		$\alpha=.05$	58
	$\alpha=.01$	32		$\alpha=.01$	45		$\alpha=.01$	64
4	$\alpha=.10$	45	4	$\alpha=.10$	64	4	$\alpha=.10$	92
	$\alpha=.05$	48		$\alpha=.05$	68		$\alpha=.05$	98
	$\alpha=.01$	53		$\alpha=.01$	75		$\alpha=.01$	107
5	$\alpha=.10$	67	5	$\alpha=.10$	97	5	$\alpha=.10$	140
	$\alpha=.05$	71		$\alpha=.05$	103		$\alpha=.05$	148
	$\alpha=.01$	79		$\alpha=.01$	113		$\alpha=.01$	161
			n	$p = 4$		n	$p = 5$	
			2	$\alpha=.10$	21	2	$\alpha=.10$	31
				$\alpha=.05$	23		$\alpha=.05$	33
				$\alpha=.01$	25		$\alpha=.01$	37
			3	$\alpha=.10$	43	3	$\alpha=.10$	65
				$\alpha=.05$	46		$\alpha=.05$	69
				$\alpha=.01$	51		$\alpha=.01$	76
			4	$\alpha=.10$	73	4	$\alpha=.10$	111
				$\alpha=.05$	78		$\alpha=.05$	117
				$\alpha=.01$	86		$\alpha=.01$	127
			5	$\alpha=.10$	111	5	$\alpha=.10$	169
				$\alpha=.05$	117		$\alpha=.05$	177
				$\alpha=.01$	129		$\alpha=.01$	192

L'esempio utilizzato non è formato da campioni bilanciati, per i quali sono stati calcolati i valori critici. Di conseguenza, per il calcolo della probabilità è necessario utilizzare la distribuzione normale, come per i grandi campioni.

Per la stima delle probabilità, nel testo indicato è stato proposto anche un metodo fondato sul calcolo combinatorio. Esso offre il vantaggio di dare stime esatte, ma presenta il grande svantaggio di essere più complesso e di richiedere calcoli molto più lunghi. Anche il problema dei ties qui è stato trascurato, poiché il test richiede scale continue e la procedura diventa più complessa. Per un approfondimento su questi temi si rimanda al testo citato.

9 – Nel caso di **grandi campioni**, la probabilità (P) è stimata in modo approssimato con la distribuzione normale standardizzata

$$Z = \frac{A_p - \mu_{Ap}}{\sqrt{\sigma_{Ap}^2}}$$

dove

$$\mu_{Ap} = \frac{N_1^2 + N_2^2 - \sum_{i=1}^k n_i^2 - n_p^2}{4}$$

$$\sigma_{Ap}^2 = \frac{2(N_1^3 + N_2^3) + 3(N_1^2 + N_2^2) - \left[\sum_{i=1}^k n_i^2(2n_i + 3) \right] - n_p^2(2n_p + 3) + 12n_p N_1 N_2 - 12n_p^2 N}{72}$$

e

- $N_1 = \sum_{i=1}^p n_i$ (nell'esempio: $7 + 3 + 5 + 4 = 19$)
- $N_2 = \sum_{i=p}^k n_i$ (nell'esempio: $4 + 4 + 3 = 11$)
- $N = \sum_{i=1}^k n_i$ (**differente da $N_1 + N_2$** : nell'esempio: $7 + 3 + 5 + 4 + 4 + 3 = 26$)

Con i dati raccolti e trascurando la media attesa già stimata,

- $2 \cdot (N_1^3 + N_2^3) = 2 \cdot (19^3 + 11^3) = 2 \cdot (6859 + 1331) = 2 \cdot 8190 = 16380$
- $3 \cdot (N_1^2 + N_2^2) = 3 \cdot (19^2 + 11^2) = 3 \cdot (361 + 121) = 3 \cdot 482 = 1446$

$$\sum_{i=1}^k n_i^2(2n_i + 3) = 7^2(2 \cdot 7 + 3) + 3^2(2 \cdot 3 + 3) + 5^2(2 \cdot 5 + 3) + 4^2(2 \cdot 4 + 3) + 4^2(2 \cdot 4 + 3) + 3^2(2 \cdot 3 + 3)$$

$$- \sum_{i=1}^k n_i^2(2n_i + 3) = (49 \cdot 17) + (9 \cdot 9) + (25 \cdot 13) + (16 \cdot 11) + (16 \cdot 11) + (9 \cdot 9)$$

$$- \sum_{i=1}^k n_i^2(2n_i + 3) = 833 + 81 + 325 + 176 + 176 + 81 = 1672$$

$$- n_p^2(2n_p + 3) = 4^2 \cdot (2 \cdot 4 + 3) = 16 \cdot 11 = 176$$

$$- 12n_p N_1 N_2 = 12 \cdot 4 \cdot 19 \cdot 11 = 10032$$

$$- 12n_p^2 N = 12 \cdot 4^2 \cdot 26 = 12 \cdot 16 \cdot 26 = 4992$$

$$\sigma_{Ap}^2 = \frac{16380 + 1446 - 1672 - 176 + 10032 - 4992}{72} = \frac{21018}{72} = 291,917$$

si calcola

$$Z = \frac{157 - 85,5}{\sqrt{291,917}} = \frac{71,5}{17,086} = 4,18$$

un valore $Z = 4,18$.

In una coda della distribuzione normale standardizzata ad esso corrisponde una probabilità molto piccola, approssimativamente $P = 0.0002$. Si rifiuta l'ipotesi nulla: si conferma che la distribuzione ha un massimo nei mesi di luglio-agosto, come ipotizzato a priori.

Quando $k = 3$, il test è del tutto uguale al test di Jonckheere-Terpstra, se $p = 3$ oppure se $p = 1$.

Ovviamente nel primo caso la disposizione delle mediane è ordinata in aumento, mentre nel secondo è in diminuzione

Se $p = 2$ (sia come valore **massimo** che come valore **minimo**) è vantaggioso ricorrere al test **U di Mann-Whitney**: concettualmente è uguale, ma sotto l'aspetto pratico è più rapido. A questo scopo, si deve

- aggregare in un gruppo unico i dati del gruppo 1 e del gruppo 3, per confrontarlo con il gruppo 2,
- per verificare l'ipotesi unilaterale, scelta sulla base della direzione attesa.

Ad esempio, per verificare se effettivamente **la classe centrale (B) è caratterizzata da un minimo**, disponendo di tre gruppi di dati

A	B	C
2,8	2,4	2,9
2,6	2,7	2,8
2,7	2,3	3,2
3,1	2,4	
	2,5	

è sufficiente formare un gruppo unico delle due classi estreme (A + C) e verificare l'ipotesi

$$H_0: me_{(A+C)} \leq me_{(B)} \quad \text{contro } H_1: me_{(A+C)} > me_{(B)}$$

ESEMPIO (CAMPIONE PICCOLO E BILANCIATO)

Si supponga di voler verificare la teoria che, nell'arco della vita di un individuo, un determinato parametro (l'intelligenza, la forza, la quantità di una proteina nel sangue, ...) abbia un generico andamento a parabola, con un massimo raggiunto tra i quarantenni (nota che la **classe d'età è specificata**):

CLASSI D'ETA'				
20-29 (1)	30-39 (2)	40-49 (3)	50-59 (4)	60-74 (5)
19,73	20,96	21,09	20,25	13,99
21,05	21,54	21,81	21,03	20,32
21,17	22,42	22,53	21,61	20,33

I dati sono in accordo con questa teoria?

Risposta. Definiti

- il numero di gruppi: in questo caso $k = 5$,
 - la posizione che individua il massimo atteso: in questo caso $p = 3$
- si applica due volte il test di Jonckheere:
- dal gruppo **1** al gruppo **3**,
 - dal gruppo **3** al gruppo **5**.

Si ottengono le seguenti 6 stime di precedenza, per confronti tra coppie di gruppi:

- le prime 3, che sono prodotte dalle combinazioni dal gruppo 1 al 3
- le altre 3, prodotte dalle combinazioni dal gruppo 5 al 3

U_{12}	U_{13}	U_{23}	U_{54}	U_{53}	U_{43}	A_3
1	2	1	1	3	2	
3	3	2	3	3	3	
3	3	3	3	3	3	
7	8	6	7	9	8	45

Nella stima delle precedenze ricordare che, ad esempio,

- U_{23} è quante volte ogni dato del gruppo **3** è preceduto dai dati del gruppo **2**,
- U_{54} è quante volte ogni dato del gruppo **4** è preceduto dai dati del gruppo **5**.

Sommando i risultati delle 6 colonne, si ricava

$$A_p = \sum U_{ij}$$

$$A_3 = 7 + 8 + 6 + 7 + 9 + 8 = 45$$

che il numero di precedenze, centrate sulla colonna 3, è $A_3 = 45$.

Nella tabella dei **valori critici** di A_p con **picco noto**

per

$$k = 5, \quad p = 3, \quad n = 3$$

si trova

- con $\alpha = 0.10$ il valore $A = 38$
- con $\alpha = 0.05$ il valore $A = 41$
- con $\alpha = 0.01$ il valore $A = 45$

Poiché il valore calcolato (45) è uguale a quello riportato nella tabella per la probabilità $\alpha = 0.01$, si rifiuta l'ipotesi nulla. E' dimostrata l'esistenza di questo picco nella classe d'età dei quarantenni.

CON PICCO IGNOTO

Quando, raccolte più serie di misure campionarie, si vuole verificare se (H_1) entro i gruppi estremi esiste un picco **senza sapere a priori in quale esso sia collocato** (quindi si rifiuta l'ipotesi nulla H_0 che le loro mediane sono uguali), si ricorre al metodo proposto di **Mack-Wolfe per un picco ignoto**.

Il metodo resta identico, se il picco **ipotizzato in H_1** è un massimo oppure un minimo. Qui è illustrato quello per verificare l'esistenza di un valore massimo significativo; se l'interesse fosse rivolto a un minimo, è sufficiente modificare simmetricamente la procedura nella formazione dei gruppi a confronto per calcolare le precedenze, come descritto per il caso di un picco noto.

Il metodo è un po' più complesso del precedente, in particolare quando i campioni non sono bilanciati; di conseguenza, l'illustrazione è limitata al caso più semplice, in cui tutti i gruppi hanno lo stesso numero di osservazioni.

Poiché i calcoli sono lunghi, è stato scelto un campione con un numero limitato di casi (3) e un numero ridotto di gruppi (5).

Ancora per semplificare la procedura, si richiede una **scala continua**, in modo che la distribuzione dei dati **non contenga ties**. Inoltre i campioni sono bilanciati. Anche la presenza di campioni con un numero differente di osservazioni, in particolare se abbinata a ties, rende il metodo sensibilmente più complesso.

La metodologia è presentata con un caso tratto dalla letteratura, in tutti i suoi passaggi logici elementari. Inoltre, per un confronto semplice tra le due impostazioni logiche e le due differenti metodologie, è ripreso l'ultimo esempio sviluppato per un picco noto.

1 - Si supponga di voler verificare la teoria che nell'arco della vita di un individuo, un determinato parametro (l'intelligenza, la forza, la quantità di una proteina nel sangue, ...) abbia un generico andamento a parabola, senza **che la teoria si spinga a identificare anche la classe d'età**.

A questo scopo, con i dati raccolti sono stati formati 5 gruppi

- ordinati per classi d'età crescente (i livelli devono sempre essere ordinabili per rango);
- entro ogni gruppo, anche i dati sono stati ordinati in modo crescente, come nella tabella:

CLASSI D'ETA'				
1	2	3	4	5
20-29	30-39	40-49	50-59	60-74
19,73	20,96	21,09	20,25	13,99
21,05	21,54	21,81	21,03	20,32
21,17	22,42	22,53	21,61	20,33

Questa disposizione dei dati facilita

- sia la identificazione del rango occupato dal gruppo in cui esiste la mediana maggiore,
- sia i successivi calcoli delle precedenze tra coppie di gruppi, in accordo con il metodo U di Mann-Whitney.

2 - Il passo successivo è la individuazione (a posteriori) tra i **k** campioni del gruppo (**p**) con il picco.

Il metodo è ovvio: **si calcolano le mediane dei campioni** e si confrontano. Con i dati dell'esempio, il gruppo che presenta la **mediana campionaria** maggiore è il 3, in cui il valore centrale è 21,81 (**p = 3**). La serie di operazioni successive serve per valutare se effettivamente, mediante il calcolo di tutte le precedenze, i dati dei gruppi confermano tale indicazione approssimata, fondata sulla sola mediana. La parte conclusiva del test serve per decidere se tale picco è reale, cioè significativo alla probabilità prefissata. Pertanto, si verifica l'ipotesi nulla

$$H_0: me_1 = me_2 = me_3 = me_4 = me_5$$

contro l'ipotesi alternativa

$$H_1: me_1 \leq me_2 \leq me_3 \geq me_4 \geq me_5$$

3 – Si stimano tutti i possibili valori **U di Mann-Whitney**, tra coppie di gruppi.

Sono le combinazioni 2 a 2 di $k = 5$, cioè 10:

$$C_5^2 = \frac{5!}{(5-2)!2!} = 10$$

Si ottengono i seguenti U_{iq} (numero di osservazioni del campione **i** che precedono le osservazioni del campione **q**, come riportati nella colonna 3):

(1)	(2)	(3)	(4)
GRUPPI	PRECEDENZE	U	U'
1-2	1 + 3 + 3	$U_{12} = 7$	$U_{21} = 2$
1-3	2 + 3 + 3	$U_{13} = 8$	$U_{31} = 1$
1-4	1 + 1 + 3	$U_{14} = 5$	$U_{41} = 4$
1-5	0 + 1 + 1	$U_{15} = 2$	$U_{51} = 7$
2-3	1 + 2 + 3	$U_{23} = 6$	$U_{32} = 3$
2-4	0 + 1 + 2	$U_{24} = 3$	$U_{42} = 6$
2-5	0 + 0 + 0	$U_{25} = 0$	$U_{52} = 9$
3-4	0 + 0 + 1	$U_{34} = 1$	$U_{43} = 8$
3-5	0 + 0 + 0	$U_{35} = 0$	$U_{53} = 9$
4-5	0 + 1 + 1	$U_{45} = 2$	$U_{54} = 7$

4- Si devono stimare anche le **precedenze tra coppie di gruppi in ordine opposto rispetto a quello appena effettuato**; nella simbologia utilizzata nel test di Mann-Whitney è il valore di **U'**; esso può essere ricavato rapidamente, ricordando la relazione

$$U' = n_i \cdot n_q - U$$

Nei gruppi bilanciati dell'esempio, con $n_i = n_q = 3$

si ottiene

$$U' = (3 \cdot 3) - U$$

la cui serie completa è riportata nella colonna 4.

Come in precedenza,

- l'indice **12** vuole dire quante volte ogni valore del gruppo **2** è **preceduto** da valori del gruppo **1**
- l'indice **21** vuole dire quante volte ogni valore del gruppo **1** è **preceduto** da valori del gruppo **2**

5 – Da questa duplice serie di **U** e **U'** si ricavano i valori $U_{.q}$ (indicati nella colonna 1 della tabella successiva) di ogni gruppo,

con

$$U_{.q} = \sum_{i \neq q} U_{iq} \quad \text{per } q = 1, \dots, k$$

Utilizzando gli addendi riportati nella colonna 2 (attenzione agli indici per distinguere **U** da **U'**),

(1)	(2)	(3)	(4)
$U_{.1}$	$U_{21} + U_{31} + U_{41} + U_{51}$	$2 + 1 + 4 + 7$	$= 14$
$U_{.2}$	$U_{12} + U_{32} + U_{42} + U_{52}$	$7 + 3 + 6 + 9$	$= 25$
$U_{.3}$	$U_{13} + U_{23} + U_{43} + U_{53}$	$8 + 6 + 8 + 9$	$= 31$
$U_{.4}$	$U_{14} + U_{24} + U_{34} + U_{54}$	$5 + 3 + 1 + 7$	$= 16$
$U_{.5}$	$U_{15} + U_{25} + U_{35} + U_{45}$	$2 + 0 + 0 + 2$	$= 4$

i cui valori campionari corrispondono a quelli riportati nella colonna 3, si ottengono i 5 totali $U_{.q}$ che sono riportati nella colonna 4.

6 – Tra essi, il valore massimo è quello della riga 3: $U_{.3} = 31$ che quindi contiene il picco, a conferma di quanto era stato possibile stimare con il semplice confronto tra le mediane campionarie.

Con campioni bilanciati, come nel caso in esame, questa individuazione è sicura.

Con campioni non bilanciati, si può ottenere che numeri differenti di confronti conducano a risultati differenti di $U_{.q}$. Di conseguenza, si rendono necessari alcuni tentativi (indicati con **r**) per trovare la soluzione corretta.

Il valore di $U_{.q}$ ha una distribuzione di frequenza con i seguenti parametri:

- il **valore medio atteso** (μ_{Uq}) è

$$\mu_{Uq} = \frac{n_q \cdot (N - n_q)}{2}$$

- la **sua varianza** (σ_{Uq}^2) è

$$\sigma_{Uq}^2 = \frac{n_q \cdot (N - n_q) \cdot (N + 1)}{12}$$

Con i dati del campione $n_q = 3$ e $N = 15$

si stimano

$$\mu_{Uq} = \frac{3 \cdot (15 - 3)}{2} = \frac{36}{2} = 18$$

$$\sigma_{Uq}^2 = \frac{3 \cdot (15 - 3) \cdot (15 + 1)}{12} = \frac{3 \cdot 12 \cdot 16}{12} = 48$$

Senza approfondire ulteriormente questi aspetti, appunto perché è stato considerato solamente il caso con campioni bilanciati, si passa alla **stima di A_3^* con picco ignoto**.

7 – Con $\mathbf{r} = \mathbf{1}$, si ricorre alla stessa metodologia utilizzata per il picco noto. Da

U_{12}	U_{13}	U_{23}	U_{54}	U_{53}	U_{43}	\mathbf{A}_3
1	2	1	1	3	2	
3	3	2	3	3	3	
3	3	3	3	3	3	
7	8	6	7	9	8	45

si ottiene

$$\mathbf{Ap} = \sum U_{ij}$$

$$A_3 = 7 + 8 + 6 + 7 + 9 + 8 = 45$$

il risultato $\mathbf{A}_3 = \mathbf{45}$.

8 – Dalle formule generali

$$\mu_{Ap} = \frac{N_1^2 + N_2^2 - \sum_{i=1}^k n_i^2 - n_p^2}{4}$$

VALORI CRITICI DI A_p^* PER IL TEST DI MACK-WOLFE

IPOTESI NULLA CON PICCO IGNOTO

- k = numero di gruppi (da 3 a 6)
- n = numero di osservazioni per gruppo (da 2 a 10, eccetto per $k = 3$)

n	$k = 3$		$k = 4$		$k = 5$		$k = 6$	
2	---	---	$\alpha=.10$	1,915	$\alpha=.10$	1,894	$\alpha=.10$	1,964
	---	---	$\alpha=.05$	2,195	$\alpha=.05$	2,191	$\alpha=.05$	2,226
	---	---	$\alpha=.01$	2,554	$\alpha=.01$	2,619	$\alpha=.01$	2,643
3	$\alpha=.10$	1,889	$\alpha=.10$	1,903	$\alpha=.10$	1,969	$\alpha=.10$	2,040
	$\alpha=.05$	2,324	$\alpha=.05$	2,213	$\alpha=.05$	2,239	$\alpha=.05$	2,242
	$\alpha=.01$	2,556	$\alpha=.01$	2,700	$\alpha=.01$	2,725	$\alpha=.01$	2,733
4	$\alpha=.10$	1,850	$\alpha=.10$	1,912	$\alpha=.10$	1,963	$\alpha=.10$	1,939
	$\alpha=.05$	2,196	$\alpha=.05$	2,180	$\alpha=.05$	2,195	$\alpha=.05$	2,265
	$\alpha=.01$	2,635	$\alpha=.01$	2,708	$\alpha=.01$	2,744	$\alpha=.01$	2,862
5	$\alpha=.10$	1,849	$\alpha=.10$	1,951	$\alpha=.10$	1,960	$\alpha=.10$	1,989
	$\alpha=.05$	2,166	$\alpha=.05$	2,221	$\alpha=.05$	2,222	$\alpha=.05$	2,251
	$\alpha=.01$	2,694	$\alpha=.01$	2,738	$\alpha=.01$	2,716	$\alpha=.01$	2,851
6	$\alpha=.10$	1,787	$\alpha=.10$	1,903	$\alpha=.10$	1,972	$\alpha=.10$	1,964
	$\alpha=.05$	2,101	$\alpha=.05$	2,160	$\alpha=.05$	2,227	$\alpha=.05$	2,242
	$\alpha=.01$	2,668	$\alpha=.01$	2,646	$\alpha=.01$	2,749	$\alpha=.01$	2,817
7	$\alpha=.10$	1,836	$\alpha=.10$	1,898	$\alpha=.10$	1,951	$\alpha=.10$	1,950
	$\alpha=.05$	2,158	$\alpha=.05$	2,205	$\alpha=.05$	2,240	$\alpha=.05$	2,257
	$\alpha=.01$	2,674	$\alpha=.01$	2,744	$\alpha=.01$	2,761	$\alpha=.01$	2,808
8	$\alpha=.10$	1,837	$\alpha=.10$	1,890	$\alpha=.10$	1,937	$\alpha=.10$	1,981
	$\alpha=.05$	2,082	$\alpha=.05$	2,184	$\alpha=.05$	2,216	$\alpha=.05$	2,256
	$\alpha=.01$	2,633	$\alpha=.01$	2,756	$\alpha=.01$	2,765	$\alpha=.01$	2,819
9	$\alpha=.10$	1,800	$\alpha=.10$	1,891	$\alpha=.10$	1,925	$\alpha=.10$	1,978
	$\alpha=.05$	2,111	$\alpha=.05$	2,201	$\alpha=.05$	2,236	$\alpha=.05$	2,266
	$\alpha=.01$	2,623	$\alpha=.01$	2,794	$\alpha=.01$	2,786	$\alpha=.01$	2,770
10	$\alpha=.10$	1,825	$\alpha=.10$	1,876	$\alpha=.10$	1,943	$\alpha=.10$	1,982
	$\alpha=.05$	2,112	$\alpha=.05$	2,172	$\alpha=.05$	2,249	$\alpha=.05$	2,278
	$\alpha=.01$	2,662	$\alpha=.01$	2,771	$\alpha=.01$	2,772	$\alpha=.01$	2,863

VALORI CRITICI DI A_p^* PER IL TEST DI MACK-WOLFE

IPOTESI NULLA CON PICCO IGNOTO

- k = numero di gruppi (da 7 a 10)
- n = numero di osservazioni per gruppo (da 2 a 10)

n	$k = 7$		$k = 8$		$k = 9$		$k = 10$	
2	$\alpha=.10$	1,992	$\alpha=.10$	2,016	$\alpha=.10$	1,999	$\alpha=.10$	2,021
	$\alpha=.05$	2,233	$\alpha=.05$	2,292	$\alpha=.05$	2,287	$\alpha=.05$	2,315
	$\alpha=.01$	2,756	$\alpha=.01$	2,723	$\alpha=.01$	2,789	$\alpha=.01$	2,818
3	$\alpha=.10$	1,982	$\alpha=.10$	2,021	$\alpha=.10$	2,027	$\alpha=.10$	2,026
	$\alpha=.05$	2,286	$\alpha=.05$	2,297	$\alpha=.05$	2,283	$\alpha=.05$	2,315
	$\alpha=.01$	2,782	$\alpha=.01$	2,821	$\alpha=.01$	2,815	$\alpha=.01$	2,802
4	$\alpha=.10$	1,999	$\alpha=.10$	2,039	$\alpha=.10$	2,031	$\alpha=.10$	2,031
	$\alpha=.05$	2,279	$\alpha=.05$	2,310	$\alpha=.05$	2,305	$\alpha=.05$	2,331
	$\alpha=.01$	2,802	$\alpha=.01$	2,866	$\alpha=.01$	2,864	$\alpha=.01$	2,910
5	$\alpha=.10$	2,017	$\alpha=.10$	2,022	$\alpha=.10$	2,035	$\alpha=.10$	2,025
	$\alpha=.05$	2,312	$\alpha=.05$	2,289	$\alpha=.05$	2,310	$\alpha=.05$	2,319
	$\alpha=.01$	2,831	$\alpha=.01$	2,798	$\alpha=.01$	2,917	$\alpha=.01$	2,874
6	$\alpha=.10$	1,974	$\alpha=.10$	2,027	$\alpha=.10$	2,041	$\alpha=.10$	2,027
	$\alpha=.05$	2,280	$\alpha=.05$	2,339	$\alpha=.05$	2,325	$\alpha=.05$	2,297
	$\alpha=.01$	2,785	$\alpha=.01$	2,885	$\alpha=.01$	2,887	$\alpha=.01$	2,912
7	$\alpha=.10$	1,997	$\alpha=.10$	2,034	$\alpha=.10$	2,030	$\alpha=.10$	2,046
	$\alpha=.05$	2,294	$\alpha=.05$	2,321	$\alpha=.05$	2,341	$\alpha=.05$	2,347
	$\alpha=.01$	2,823	$\alpha=.01$	2,928	$\alpha=.01$	2,925	$\alpha=.01$	2,922
8	$\alpha=.10$	1,988	$\alpha=.10$	2,034	$\alpha=.10$	2,037	$\alpha=.10$	2,031
	$\alpha=.05$	2,282	$\alpha=.05$	2,315	$\alpha=.05$	2,325	$\alpha=.05$	2,343
	$\alpha=.01$	2,889	$\alpha=.01$	2,875	$\alpha=.01$	2,879	$\alpha=.01$	2,895
9	$\alpha=.10$	1,986	$\alpha=.10$	2,021	$\alpha=.10$	2,027	$\alpha=.10$	2,050
	$\alpha=.05$	2,276	$\alpha=.05$	2,305	$\alpha=.05$	2,293	$\alpha=.05$	2,380
	$\alpha=.01$	2,826	$\alpha=.01$	2,874	$\alpha=.01$	2,883	$\alpha=.01$	2,948
10	$\alpha=.10$	2,008	$\alpha=.10$	2,028	$\alpha=.10$	2,059	$\alpha=.10$	2,032
	$\alpha=.05$	2,338	$\alpha=.05$	2,333	$\alpha=.05$	2,340	$\alpha=.05$	2,351
	$\alpha=.01$	2,919	$\alpha=.01$	2,893	$\alpha=.01$	2,888	$\alpha=.01$	2,905

$$\sigma_{Ap}^2 = \frac{2(N_1^3 + N_2^3) + 3(N_1^2 + N_2^2) - \left[\sum_{i=1}^k n_i^2 (2n_i + 3) \right] - n_p^2 (2n_p + 3) + 12n_p N_1 N_2 - 12n_p^2 N}{72}$$

con

- $N_1 = \sum_{i=1}^p n_i$ (nell'esempio: $3 + 3 + 3 = 9$)
- $N_2 = \sum_{i=p}^k n_i$ (nell'esempio: $3 + 3 + 3 = 9$)
- $N = \sum_{i=1}^k n_i$ (**diversa da $N_1 + N_2$** : nell'esempio: $3 + 3 + 3 + 3 + 3 = 15$)
- $n_i = n_p = 3$ e $k = 5$

si ricavano

$$\mu_{A3} = \frac{9^2 + 9^2 - (3^2 + 3^2 + 3^2 + 3^2 + 3^2) - 3^2}{4} = \frac{81 + 81 - 45 - 9}{4} = \frac{108}{4} = 27$$

$$\sigma_{A3}^2 = \frac{2 \cdot (9^3 + 9^3) + 3 \cdot (9^2 + 9^2) - 5 \cdot [3^2 \cdot (2 \cdot 3 + 3)] - 3^2 \cdot (2 \cdot 3 + 3) + 12 \cdot 3 \cdot 9 \cdot 9 - 12 \cdot 3^2 \cdot 15}{72}$$

$$\sigma_{A3}^2 = \frac{(2 \cdot 1458) + (3 \cdot 162) - (5 \cdot 9 \cdot 9) - (9 \cdot 9) + 2916 - 1620}{72} = \frac{4212}{72} = 58,5$$

- la media attesa (nella condizione che H_0 sia vera) $\mu_{A3} = 27$
- e la sua varianza $\sigma_{A3}^2 = 58,5$

9 – Infine , mediante la formula generale

$$A_p^* = \frac{A_3 - \mu_{A3}}{\sqrt{\sigma_{A3}^2}}$$

e i risultati già ricavati

$$A_3^* = \frac{45 - 27}{\sqrt{58,5}} = \frac{18}{7,649} = 2,353$$

si ottiene $A_3^* = 2,353$.

Questo risultato deve essere confrontato con i **valori critici riportati nella tabella per picco ignoto**.

Con $k = 5$ e $n = 3$ alla probabilità $\alpha = 0.05$ è 2,239 e alla probabilità $\alpha = 0.01$ è 2,725.

Poiché il valore calcolato (2,353) è maggiore di quello critico alla probabilità 0.05, si rifiuta l'ipotesi nulla. L'analisi dimostra che esiste una evidenza sufficiente per affermare che questa caratteristica aumenta con l'età verso le classi centrali, per poi diminuire con l'invecchiamento.

Un altro aspetto importante da rilevare è che **con picco noto la probabilità ($P < 0.01$) era risultata inferiore a quella ricavata ora con picco ignoto ($P < 0.05$)**. E' un concetto ampiamente discusso nei confronti multipli e che evidenzia i vantaggi che si ottengono nell'impostare l'analisi statistica sulla verifica di ipotesi formulate a priori, prima della raccolta dei dati, che dovrebbero sempre essere campionati con tale finalità specifica.

19.5. IL TEST DI PAGE O DELLE ALTERNATIVE ORDINATE IN K CAMPIONI DIPENDENTI

Il **test di Page** è simile al test di Jonckheere, ma serve quando le rilevazioni dei dati sono condotte in modo programmato, tale che i **k** gruppi a confronto abbiano osservazioni raccolte nelle stesse condizioni. Nell'analisi della varianza con la quale condivide l'ipotesi nulla

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

è il caso di **k campioni dipendenti** o a **blocchi randomizzati**

I dati devono essere impostati in una tabella a doppia entrata:

		T R A T T A M E N T I						
		1	2	3	...	j	...	k
B L O C C H I	1	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$...	$X_{1,j}$...	$X_{1,k}$
	2	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$...	$X_{2,j}$...	$X_{2,k}$
	3	$X_{3,1}$	$X_{3,2}$	$X_{3,3}$...	$X_{3,j}$...	$X_{3,k}$
	---	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	I	$X_{i,1}$	$X_{i,2}$	$X_{i,3}$...	$X_{i,j}$...	$X_{i,k}$
	---	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
	N	$X_{N,1}$	$X_{N,2}$	$X_{N,3}$...	$X_{N,j}$...	$X_{N,k}$

Questo test, derivato da quello di Friedman già noto come *distribution-free test for ordered alternatives based on Friedman rank sums*, è stato proposto da Ellis Batten **Page** dell'Università del Connecticut nel 1963 con l'articolo *Ordered hypotheses for multiple treatments: a significance test for linear ranks* (pubblicato sulla rivista **Journal of the American Statistical Association** Vol. 58, pp. 216-230).

Il test di **Page**, come il test di **Friedman**, serve per verificare l'ipotesi nulla

$$H_0: me_1 = me_2 = me_3 = \dots = me_K$$

ma contro l'**ipotesi alternativa unilaterale**

$$H_1: me_1 \leq me_2 \leq me_3 \leq \dots \leq me_K$$

o con un ordine delle mediane opposto.

Nel riassunto del suo articolo, Page scrive: *A ranking statistic L is presented as test of a monotonic relationship among the treatment groups in the two-way analysis of variance.*

Il test è **solo unilaterale** e quindi occorre definire a priori l'ordine o rango atteso dei valori delle mediane.

E' quindi **più potente del test di Friedman**.

Accettare l'ipotesi alternativa H_1 significa che **esiste almeno una disuguaglianza (<) valida in senso stretto**. La **procedura è fondata sulla somma dei ranghi di ogni gruppo**. La scala utilizzata per determinare i valori raccolti con il campione deve essere di **tipo almeno ordinale**.

Per proporre il metodo, evidenziando affinità e differenze con il test di Friedman, Page presenta la seguente tabella di un esperimento fittizio

	K Trattamenti				
N	A	B	C	D	
Righe	1	2	3	4	Y
I	2	1	3	4	
II	1	3	4	2	
III	1	3	2	4	
IV	1	4	2	3	
V	3	1	2	4	
VI	1	2	4	3	
$\sum^N X_{ij} =$	9	14	17	20	= 60
$Y_j \cdot \sum^N X_{ij} =$	9	28	51	80	= 168 = L

(In realtà Page indica con **n** il numero di trattamenti e con **m** il numero di righe; ma per utilizzare la stessa simbologia usata nei test precedenti, anche qui è usato **k** per il numero di trattamenti e **N** per quello delle righe)

Il **test di significatività** utilizza il χ^2 . Tale verifica è stata introdotta nel 1937 da **Milton Friedman** con l'articolo *The use of ranks to avoid the assumption of normality implicit in the analysis of variance* (pubblicato su **Journal of the American Statistical Association**, Vol. 32 pp. 675-701) e sviluppata nel 1940 con l'articolo *A comparison of alternative tests of significance for the test of m rankings* (pubblicato su **Annals of Mathematical Statistics**, Vol. 11, pp. 86-92). Successivamente è stata esaminata da altri, tra i quali M. G. **Kendall** e B. B. **Smith** nel 1939 con l'articolo *The problem of m rankings* (pubblicato su **Annals of Mathematical Statistics**, Vol. 10, pp. 275-287).

La formula generale

$$\chi^2 = \sum \frac{(O_{ss.} - Att.)^2}{Att.}$$

applicata ai ranghi diventa

$$\chi_r^2 = \frac{6 \sum (O_{ss.} - Att.)^2}{\sum O_{ss.}}$$

dove i gdl sono $N - 1$ e il valore atteso è

$$Att. = \frac{\sum O_{ss.}}{k}$$

Con i dati dell'esempio riportato da Page
dove

$$Att. = \frac{60}{4} = 15$$

$$\chi_r^2 = \frac{6[(9-15)^2 + (14-15)^2 + (17-15)^2 + (20-15)^2]}{60} = 6,60$$

si ottiene $\chi^2 = 6,60$ con 3 gdl.

Il test **non risulta significativo**, poiché il valore critico alla probabilità $\alpha = 0.05$ è $\chi_3^2 = 7,82$.

In questo caso, il test di Friedman non permette di rifiutare l'ipotesi nulla.

Ellis Batten **Page** propone la statistica

$$L = \sum \left(Y_j \cdot \sum X_{ij} \right)$$

dove Y_j è il rango della colonna.

Con i dati dell'esempio

$$L = (1 \cdot 9) + (2 \cdot 14) + (3 \cdot 17) + (4 \cdot 20) = 168$$

si ottiene $L = 168$

I valori critici sono forniti direttamente dalla Page nell'articolo citato. Essi sono riportati nelle tabelle seguenti, con variazioni che riguardano solo la impostazione grafica.

Nel caso di **piccoli campioni**, la significatività di L è fornita solo dalle tabelle.

L 'entrata è determinata

- dal **numero di colonne o gruppi (k)** e
- dal **numero di dati (N) in ogni colonna**.

Con i dati dell'esempio,

in cui $k = 4$ e $N = 6$,

la tabella dei valori critici alla probabilità

- $\alpha = 0.05$ riporta un valore uguale a **163**
- $\alpha = 0.01$ riporta un valore uguale a **167**
- $\alpha = 0.001$ riporta un valore uguale a **172**

Avendo calcolato $L = 168$ nel caso dell'esempio precedente è possibile rifiutare l'ipotesi nulla con probabilità $P < 0.01$. E' da evidenziare che con il test di Friedman non si rifiutava l'ipotesi nulla nemmeno alla probabilità $\alpha = 0.05$.

Nel caso di **grandi campioni**, la statistica L si distribuisce in modo approssimativamente normale

$$Z = \frac{L - \mu_L}{\sigma_L}$$

dove, in tabelle a doppia entrata di

- k trattamenti o colonne ,
- N righe o osservazioni per colonna,

la media attesa μ_L è

$$\mu_L = \frac{N \cdot k \cdot (k + 1)^2}{4}$$

e σ_L è

$$\sigma_L = \sqrt{\frac{N \cdot (k^3 - k)^2}{144(k - 1)}}$$

Valori critici di L proposti da Page

(Sono significativi i valori uguali o maggiori di quelli riportati alle probabilità prefissate)

I valori per **campioni piccoli**, riportati entro caselle sono basati su distribuzioni esatte.

I valori per **campioni grandi**, riportati senza casella, sono basati su approssimazioni alla normale

N	K = 3			K = 4			K = 5		
	α			α			α		
	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
2	28	---	---	58	60	---	103	106	109
3	41	42	---	84	87	89	150	155	160
4	54	55	56	111	114	117	197	204	210
5	66	68	70	137	141	145	244	251	259
6	79	81	83	163	167	172	291	299	307
7	91	93	96	189	193	198	338	346	355
8	104	106	109	214	220	225	384	393	403
9	116	119	121	240	246	252	431	441	451
10	128	131	134	266	272	278	477	487	499
11	141	144	147	292	298	305	523	534	546
12	153	156	160	317	324	331	570	581	593
13	165	169	172	343	350	358	615	628	642
14	178	181	185	368	376	384	661	674	689
15	190	194	197	394	402	410	707	721	736
16	202	206	210	420	427	436	754	767	783
17	215	218	223	445	453	463	800	814	830
18	227	231	235	471	479	489	846	860	876
19	239	243	248	496	505	515	891	906	923
20	251	256	260	522	531	541	937	953	970

(continua)

Valori critici di L proposti da Page

(Sono significativi i valori uguali o maggiori di quelli riportati alle probabilità prefissate)

I valori per **campioni piccoli**, riportati entro caselle sono basati su distribuzioni esatte.

I valori per **campioni grandi**, riportati senza casella, sono basati su approssimazioni alla normale

N	K = 6			K = 7			K = 8		
	α			α			α		
	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
2	166	173	178	252	261	269	362	376	388
3	244	252	260	370	382	394	532	549	567
4	321	331	341	487	501	516	701	722	743
5	397	409	420	603	620	637	869	893	917
6	474	486	499	719	737	757	1037	1063	1090
7	550	563	577	835	855	876	1204	1232	1262
8	625	640	655	950	972	994	1371	1401	1433
9	701	717	733	1065	1088	1113	1537	1569	1603
10	777	793	811	1180	1205	1230	1703	1736	1773
11	852	869	888	1295	1321	1348	1868	1905	1943
12	928	946	965	1410	1437	1465	2035	2072	2112
13	1003	1022	1044	1525	1553	1585	2201	2240	2285
14	1078	1098	1121	1639	1668	1702	2367	2407	2453
15	1153	1174	1197	1754	1784	1818	2532	2574	2622
16	1228	1249	1274	1868	1899	1935	2697	2740	2790
17	1303	1325	1350	1982	2014	2051	2862	2907	2958
18	1378	1401	1427	2097	2130	2167	3028	3073	3126
19	1453	1476	1503	2217	2245	2283	3193	3240	3294
20	1528	1552	1579	2325	2360	2399	3358	3406	3461

(continua)

Valori critici di L proposti da Page

(Sono significativi i valori uguali o maggiori di quelli riportati alle probabilità prefissate)

I valori per **campioni piccoli**, riportati entro caselle sono basati su distribuzioni esatte.
I valori per **campioni grandi**, riportati senza casella, sono basati su approssimazioni alla normale

N	K = 9			K = 10		
	α			α		
	0.05	0.01	0.001	0.05	0.01	0.001
2	500	520	544	670	696	726
3	736	761	790	987	1019	1056
4	971	999	1032	1301	1339	1382
5	1204	1236	1273	1614	1656	1704
6	1436	1472	1512	1927	1972	2025
7	1668	1706	1750	2238	2288	2344
8	1900	1940	1987	2549	2602	2662
9	2131	2174	2223	2859	2915	2980
10	2361	2407	2459	3169	3228	3296
11	2592	2639	2694	3478	3541	3612
12	2822	2872	2929	3788	3852	3927
13	3052	3104	3163	4097	4164	4241
14	3281	3335	3397	4405	4475	4556
15	3511	3567	3631	4714	4786	4869
16	3741	3798	3864	5022	5097	5183
17	3970	4029	4098	5330	5407	5496
18	4199	4260	4330	5638	5717	5808
19	4428	4491	4563	5946	6027	6121
20	4657	4722	4796	6253	6337	6433

Sostituendo in

$$Z = \frac{L - \mu_L}{\sigma_L}$$

e semplificando,

si ottiene la formula abbreviata

$$Z_L = \frac{12 \cdot L - 3 \cdot N \cdot k \cdot (k+1)^2}{k \cdot (k^2 - 1)} \cdot \sqrt{\frac{k-1}{N}}$$

che permette un calcolo più rapido e semplice del valore di Z.

Poiché **il test di Page** è a **una coda**, i valori critici fondamentali sono:

- **Z = 1,645** alla probabilità **$\alpha = 0.05$**
- **Z = 2,328** alla probabilità **$\alpha = 0.01$**
- **Z = 3,09** alla probabilità **$\alpha = 0.001$**

Nelle precedenti **tabelle dei valori critici**, entro le caselle sono riportati i valori stimati con il metodo esatto, valido per campioni piccoli. Nella zone senza le caselle, sono riportati i valori ricavati con la distribuzione normale, valida per grandi campioni. Pubblicati dalla Page fino a campioni di dimensioni $k = 10$ e $N = 50$, questi valori permettono di **verificare la significatività di L** senza l'uso della normale. Il ricorso alla distribuzione Z unilaterale ha lo svantaggio di richiedere calcoli molto più lunghi, per passare da L a Z, ma il vantaggio di determinare probabilità P più precise.

Per spiegare la metodologia in modo ancora più semplice e illustrare altri concetti importanti nell'applicazione del test, è utile sviluppare un esempio in tutti i suoi passaggi logici e metodologici.

ESEMPIO 1. Si supponga di voler **sperimentare gli effetti di una sostanza tossica sulla crescita di una specie** vegetale. A questo scopo, sono stati collocati i semi

- in 5 colture con concentrazione crescente (36, 54, 72, 108, 144) del principio attivo e
- il test è stato ripetuto da 3 ricercatori (A, B, C).

Nella tabella sottostante sono riportati i valori medi delle 4 repliche fatte da ogni ricercatore per ogni concentrazione (l'uso di valori medi vieta l'uso della statistica parametrica):

	CONCENTRAZIONE (g/l)				
Ricercatore	36	54	72	108	144
A	7,63	8,14	7,76	7,17	7,46
B	8,00	8,15	7,73	7,57	7,68
C	7,93	7,87	7,74	7,80	7,21

E' possibile affermare che l'aumento della concentrazione della sostanza tossica (da 36 a 144) inibisce la crescita della specie?

Risposta. La metodologia inferenziale del test di Page richiede alcuni passaggi.

1 – Dapprima si deve formulare l'ipotesi nulla

$$H_0: me_{36} = me_{54} = me_{72} = me_{108} = me_{144}$$

secondo la quale le mediane della crescita alle 5 diverse concentrazioni sono uguali
e l'ipotesi alternativa unilaterale

$$H_1: me_{144} \leq me_{108} \leq me_{72} \leq me_{54} \leq me_{36}$$

secondo la quale la crescita della specie aumenta con il diminuire della concentrazione.

2 – L'ipotesi alternativa definisce l'ordine con il quale devono essere riportati i dati nella tabella.

Con i dati dell'esempio, è necessario invertire l'ordine delle colonne (i trattamenti a confronto), che pertanto diventa

	CONCENTRAZIONE (g/l)				
Ricercatore	144	108	72	54	36
A	7,46	7,17	7,76	8,14	7,63
B	7,68	7,57	7,73	8,15	8,00
C	7,21	7,80	7,74	7,87	7,93

3 – Come nel test di Friedman, i valori devono essere **trasformati nel loro rango, entro la stessa riga**. Si assegna quindi il rango minore (1) al punteggio più basso e si aumenta di 1 per ogni punteggio successivo della stessa riga fino a **k**, uguale al numero di trattamenti o gruppi a confronto. Nel caso di due o più valori uguali entro la stessa riga, secondo la procedura abituale, il punteggio viene calcolato sulla media dei loro ranghi.

Con i dati dell'esempio, la trasformazione dei valori in ranghi delle tre prime righe e la loro somma, riportata nella quarta riga, diventa

	CONCENTRAZIONE (g/l)				
Ricercatore	144	108	72	54	36
A	2	1	4	5	3
B	2	1	3	5	4
C	1	3	2	4	5
R_j	5	5	9	14	12

4 - Il test utilizza i totali dei ranghi R_j , effettuati per colonna.

Se è vera l'ipotesi nulla H_0 , i punteggi saranno distribuiti a caso entro la stessa riga e pertanto le somme dei ranghi (R_j) per colonna tenderanno ad essere uguali. Questi valori dipenderanno dal numero N di dati presenti in ogni colonna.

5 - Nel test di Page, dalla somma degli R_j si ricava **L**, definita come la somma dei ranghi di ogni colonna, moltiplicata per la posizione attribuita alla colonna, come nella formula seguente

$$L = \sum j \cdot R_j = 1 \cdot R_1 + 2 \cdot R_2 + 3 \cdot R_3 + \dots + k \cdot R_k$$

Con i dati dell'esempio

$$L = 1 \times 5 + 2 \times 5 + 3 \times 9 + 4 \times 14 + 5 \times 12 = 5 + 10 + 27 + 56 + 60 = 158$$

L risulta uguale a 158.

6 - Nel caso di **piccoli campioni**, la significatività di L è fornita dalle tabelle.

L'entrata è determinata

- dal **numero di colonne o gruppi (k)** e
- dal **numero di dati (N) in ogni colonna**.

Con i dati dell'esempio,

in cui $k = 5$ e $N = 3$,

la tabella dei valori critici alla probabilità

- $\alpha = 0.05$ riporta un valore uguale a **150**
- $\alpha = 0.01$ riporta un valore uguale a **155**
- $\alpha = 0.001$ riporta un valore uguale a **160**

Poiché il valore calcolato è $L = 158$, **si rifiuta l'ipotesi nulla con probabilità $P < 0.01$** .

E' importante osservare come l'affermazione riguardi il trend generale, cioè la diminuzione dei valori medi di crescita nel loro complesso, non tutti i risultati dei trattamenti; infatti la crescita media è stata maggiore con dose 54 rispetto a dose 36.

TIES. I dati dovrebbero essere misurati su una scala continua, per cui **non dovrebbero esistere valori con lo stesso rango**. Page non ha proposto correzioni per i ties, affermando che **la scala deve essere continua**. Tuttavia un numero molto limitato di valori sono accettabili, poiché non alterano sensibilmente il risultato e forniscono una risposta più cautelativa. In altri termini l'errore che si commette con i ties rende il test meno significativo

Nel **caso di 2 soli gruppi ($k = 2$)**, l'alternativa al test di Page è il test **U di Mann-Whitney per 2 campioni dipendenti, nell'ipotesi unidirezionale** (Infatti verifica la direzione della tendenza centrale con due gruppi).

Nelle situazioni in cui i trattamenti hanno un ordine naturale, il test di Page è preferibile a quello di Friedman, che risponde solo alla domanda se le mediane dei vari trattamenti sono differenti.

Una volta che sia stata rifiutata l'ipotesi nulla, per ottenere una informazione più completa sulla serie delle mediane, è possibile **verificare tra quali esista una differenza significativa, ricorrendo ai confronti multipli basati sulla somma dei ranghi di Friedman**.

ESEMPIO 2 (PER GRANDI CAMPIONI). Lungo il corso d'acqua che attraversa una città, sono state collocate 6 stazioni (A, B, C, D, E, F) di rilevazione dell'inquinamento. In ognuna delle 6 stazioni, per 15 giorni è stata fatta una misura del carico inquinante.

I valori campionati, classificati per stazione e per giorno di rilevazione, sono riportati nella tabella a due entrate:

GIORNI	STAZIONI					
	A	B	C	D	E	F
1	20	18	24	22	29	38
2	32	37	34	31	39	38
3	18	23	19	25	23	26
4	9	7	14	11	12	11
5	29	37	32	59	40	45
6	38	25	27	47	45	45
7	8	15	7	12	15	13
8	18	13	22	26	23	22
9	32	36	37	35	48	40
10	23	25	26	25	32	56
11	6	8	12	9	10	10
12	24	18	20	27	25	27
13	13	18	14	14	19	26
14	18	26	19	19	29	32
15	14	12	25	56	54	75

Si intende verificare se, con l'accumulo degli scarichi urbani, lungo il percorso (dalla stazione A alla F) si ha un aumento significativo del carico inquinante.

Risposta.

1 - Dapprima si devono definire l'ipotesi nulla

$$H_0: me_A = me_B = me_C = me_D = me_E = me_F$$

e l'ipotesi alternativa

$$H_1: me_A \leq me_B \leq me_C \leq me_D \leq me_E \leq me_F$$

intendendo con essa che tra valore iniziale e valore finale c'è almeno una discontinuità significativa.

2 - Successivamente, poiché i trattamenti sono già nell'ordine naturale, si devono trasformare i valori in ranghi, considerando in modo indipendente ogni riga, come nella tabella sottostante:

GIORNI	STAZIONI					
	A	B	C	D	E	F
1	2	1	4	3	5	6
2	2	4	3	1	6	5
3	1	3,5	2	5	3,5	6
4	2	1	6	3,5	5	3,5
5	1	3	2	6	4	5
6	3	1	2	6	4,5	4,5
7	2	5,5	1	3	5,5	4
8	2	1	3,5	6	5	3,5
9	1	3	4	2	6	5
10	1	2,5	4	2,5	5	6
11	1	2	6	3	4,5	4,5
12	3	1	2	5,5	4	5,5
13	1	4	2,5	2,5	5	6
14	1	4	2,5	2,5	5	6
15	2	1	3	5	4	6
R _j	25,0	37,5	47,5	56,5	71,5	82,5

3 - Successivamente, si calcolano le somme dei ranghi per colonna (**R_j**), come nell'ultima riga della tabella precedente.

4 - Dalla somma dei ranghi per colonna (**R_j**) si ottiene il valore di **L**

$$L = (1 \cdot 25,0) + (2 \cdot 37,5) + (3 \cdot 47,5) + (4 \cdot 56,5) + (5 \cdot 71,5) + (6 \cdot 82,5) = \\ = 25 + 75 + 142,5 + 226 + 357,5 + 495 = \mathbf{1321}$$

che risulta uguale a 1321.

5 - Per valutare la sua significatività, si stima il valore di Z_L

$$Z_L = \frac{(12 \cdot 1321) - (3 \cdot 15 \cdot 6 \cdot 7^2)}{6 \cdot (6^2 - 1)} \cdot \sqrt{\frac{5}{15}} = \frac{15852 - 13230}{210} \cdot \sqrt{0,333} = 12,48 \cdot 0,577 = 7,20$$

che risulta uguale a 7,20.

Il valore di Z_L ottenuto è particolarmente elevato, superiore a quelli riportati nella tabella della distribuzione normale unilaterale: corrisponde ad una probabilità P di circa 1 su un milione.

Pertanto, si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa: dalla stazione A alla stazione F si ha un aumento altamente significativo del carico inquinante, anche se non è possibile dire nulla sulla curva di crescita.

Nella tabella dei valori critici, con $k = 6$ e $N = 15$ alla probabilità $\alpha = 0.001$ è riportato $L = 1197$. Permette di rifiutare l'ipotesi nulla con probabilità $P < 0.001$ quindi con una stima molto meno precisa. Infine l'eventuale ricorso ai confronti multipli con il test di Friedman permette di verificare tra quali stazioni esiste una differenza significativa.

ESEMPIO 3 (PER PICCOLI CAMPIONI). In un centro urbano sono stati attivati 5 punti di osservazione (A, B, C, D, E) per rilevare l'inquinamento dovuto al traffico e per ogni zona sono state effettuate 4 osservazioni, a distanza di 4 ore ognuna, dalle ore 6 alle ore 18.

Per esporre alla cittadinanza la situazione dell'inquinamento atmosferico in modo chiaro e sintetico, l'ufficio comunale ha fornito le misure su una scala ordinale, con giudizi della qualità dell'aria riportati in modo simbolico,

-- pessima
 - insufficiente
 = sufficiente
 + buona
 ++ ottima

come nella tabella sottostante

STAZIONE	ORE			
	6	10	14	18
A	-	--	+	=
B	--	=	-	=
C	-	=	+	+
D	=	+	=	++
E	-	+	=	+

Si vuole verificare se, nel corso di una giornata con blocco del traffico, dal mattino alla sera vi è stata una diminuzione nei tassi d'inquinamento.

Risposta. I dati devono essere

- trasformati in ranghi entro la stessa riga e
- sommati per colonna (R_j),

STAZIONE	ORE			
	6	10	14	18
A	2	1	4	3
B	1	3,5	2	3,5
C	1	2	3,5	3,5
D	1,5	3	1,5	4
E	1	3,5	3	3,5
R_i	6,5	13,0	14,0	17,5

al fine di calcolare il valore di **L**

$$L = 1 \times 6,5 + 2 \times 13 + 3 \times 14 + 4 \times 17,5 = 144,5$$

che risulta uguale a 144,5.

Per **k = 4** e **N = 5**,

la tabella sinottica dei **valori critici della L di Page** riporta

- alla probabilità $\alpha = 0.05$ il valore 137
- alla probabilità $\alpha = 0.01$ il valore 141
- alla probabilità $\alpha = 0.001$ il valore 145.

Il valore di **L** calcolato (144,5) è compreso tra quello relativo alla probabilità $\alpha = 0.01$ e quello alla probabilità $\alpha = 0.001$.

Di conseguenza, si rifiuta l'ipotesi nulla con probabilità $P < 0.01$ di commettere un errore di I tipo.

Si accetta l'ipotesi alternativa: nel complesso delle stazioni esiste un miglioramento delle condizioni dell'aria al trascorrere delle ore (dalle 6 alle 18), durante la giornata di blocco del traffico.

ESEMPIO 4 (CONFRONTI TRA L E Z). E' ovvio che, quando il campione non è troppo piccolo, in tutti i test non parametrici esista corrispondenza tra le probabilità indicate dai valori critici riportate nelle tabelle sinottiche e quelle stimate con la distribuzione normale.

E' una relazione che spesso sfugge allo studente e al ricercatore, che non abbiano acquisito sufficiente familiarità con la statistica. A questo scopo, è didatticamente utile l'esercizio di valutare le probabilità corrispondenti ai valori critici.

Per il test di Page dalla tabella dei valori critici ricaviamo che

per **k** uguale a 10 ed **N** uguale a 12,

- alla probabilità $\alpha = 0.05$ il valore critico è $L = 3788$,
- alla probabilità $\alpha = 0.01$ è $L = 3852$,

- alla probabilità $\alpha = 0.001$ è $L = 3927$.

Poiché (con k uguale a 10 ed N uguale a 12) il campione è di dimensioni relativamente grandi, le probabilità calcolate con la distribuzione normale, ovviamente solo in una coda come richiede il test, dovrebbero essere molto simili.

Infatti la formula generale

$$Z_L = \frac{12 \cdot L - 3 \cdot N \cdot k \cdot (k+1)^2}{k \cdot (k^2 - 1)} \cdot \sqrt{\frac{k-1}{N}}$$

per $N = 12$ e $k = 10$ diventa

$$Z = \frac{12 \cdot L - 3 \cdot 12 \cdot 10 \cdot (10+1)^2}{10 \cdot (10^2 - 1)} \cdot \sqrt{\frac{10-1}{12}}$$

nella quale l'unico valore da inserire è L , fornito dalle tabelle.

Per la probabilità $\alpha = 0.05$ il valore critico di L è 3788;

con esso, il risultato di Z diviene

$$Z = \frac{(12 \cdot 3788) - (360 \cdot 11^2)}{10 \cdot 99} \cdot \sqrt{\frac{9}{12}} = \frac{45456 - 43560}{990} \cdot \sqrt{0,75} = \frac{1896}{990} \cdot 0,866 = 1,66$$

uguale a 1,66.

Nella tabella dei valori critici in una coda della distribuzione normale, a $Z = 1,66$ corrisponde la probabilità $\alpha = 0.0485$. E' un valore molto vicino a 0.05 indicato per la significatività di L .

Per la probabilità $\alpha = 0.01$ a L si sostituisce il valore critico di 3852;

il risultato di Z diviene

$$Z = \frac{(12 \cdot 3852) - (360 \cdot 11^2)}{10 \cdot 99} \cdot \sqrt{\frac{9}{12}} = \frac{46224 - 43560}{990} \cdot \sqrt{0,75} = \frac{2664}{990} \cdot 0,866 = 2,33$$

uguale a 2,33 al quale, in una coda della distribuzione, corrisponde la probabilità $\alpha = 0.009$.

Per la probabilità $\alpha = 0.001$ a L si sostituisce il valore critico di 3927;

il risultato di Z diviene

$$Z = \frac{(12 \cdot 3927) - (360 \cdot 11^2)}{10 \cdot 99} \cdot \sqrt{\frac{9}{12}} = \frac{47124 - 43560}{990} \cdot \sqrt{0,75} = \frac{3564}{990} \cdot 0,866 = 3,11$$

uguale a 3,11 al quale corrisponde la probabilità $\alpha = 0.00099$.

Con la distribuzione normale, i valori critici hanno fornito probabilità molto vicine, quasi coincidenti, con quelle riportate nella tabella.

E' una dimostrazione empirica del fatto che, in assenza dei valori critici, **il ricorso alla distribuzione normale fornisce risposte attendibili**, se i campioni non sono eccessivamente piccoli.

Per campioni di dimensioni maggiori, la coincidenza è perfetta, a meno delle approssimazioni introdotte nei calcoli. Infatti la Page stima i valori critici di L fino a $N = 50$ e $k = 10$ utilizzando i valori critici di Z. Ad esempio per $\alpha = 0.01$ suggerisce la misura approssimata $Z = 2,33$.

Per valutare la significatività di L in grandi campioni, Page suggerisce

$$\chi_L^2 = \frac{[12L - 3Nk(k+1)^2]^2}{Nk^2(k^2 - 1)(k+1)}$$

che è distribuito approssimativamente come un chi-quadrato con 1 grado di libertà.

Ma il chi-quadrato serve per test bilaterali, mentre il test di Page è unilaterale: di conseguenza la probabilità stimata con il chi-quadrato deve essere dimezzata.

*χ_L^2 is distributed approximately as chi-square with 1 degree of freedom. The ordinary use of χ^2 tables would be equivalent to a two-sided test of agreement. If a one-sided test is desired, **as will almost always be the case**, the probability discovered in the chi-square table should be **halved**.*