ISTISAN 1982/27

Guida alla elaborazione statistica dei dati
Correlazione e Regressione

F. Chiarotti e F. Taggi

Laboratorio di Epidemiologia e Biostatistica

Roma, novembre 1982

INDICE

PRESENTAZIONE

Questo rapporto costituisce una sezione di una "Guida alla Elaborazione Statistica dei dati", in corso di realizzazione presso il reparto di Metodologie e Modelli Biostatistici del Laboratorio di Epidemiologia e Biostatistica.

La Guida in oggetto dovrebbe essere, una volta completata, uno strumento di lavoro per ricercatori ed operatori sanitari e, più in generale, per tutti coloro che hanno a che fare con dei dati e vogliono da questi trarre indicazioni generali e/o specifiche sui fenomeni in studio, usando procedure di Elaborazione Elettronica dei Dati (EDP).

La struttura della Guida risente, oltre che del carattere non definito dell'utente, anche delle particolari macchine utilizzate nella soluzione dei problemi stessi. Per una serie di ragioni, derivanti da decisioni prese negli scorsi anni e confortate da successive esperienze, tale scelta è caduta su piccoli computer grafici (4051 Tektronix), programmabili in linguaggio BASIC. La caratteristica essenziale del linguaggio BASIC (la sua grande capacità di "conversare" con l'utilizzatore) si è dimostrata di grande utilità, in particolare nel permettere ad utenti non specificatamente interessati agli aspetti tecnici dell'EDP un accesso quasi immediato ad una consistente parte di programmi di elaborazione disponibili.

Dato il carattere elementare dei problemi affrontati nella Guida, tali macchine BASIC sono apparse i migliori strumenti utilizzabili nel risolvere i problemi trattati. La possibilità di avere, oltre all'elaborazione numerica, anche rappresentazioni grafiche delle elaborazioni eseguite accresce, peraltro, il valore di tali strumenti di lavoro, in particolare azionando un meccanismo di autocritica nell'utente stesso: nulla è più immediato di un grafico, specie quando si debba valutare quanto le significatività statistiche eventualmente trovate siano anche significative a livello pratico.

La Guida è stata, nel progetto, suddivisa in 7 sezioni, ciascuna indipendente dalle altre:

1 - Statistica Descrittiva e Programmi di interesse generale;
2 - Analisi di 1 campione;
3 - Analisi di 2 campioni dipendenti;
4 - Analisi di 2 campioni indipendenti;
5 - Analisi di k campioni dipendenti;
6 - Analisi di k campioni indipendenti;
7 - Correlazione e regressione.

Nella sezione esaminata in questo rapporto sono specificamente trattati alcuni fra i più noti e rilevanti tests non-parametrici e parametrici per l'analisi della correlazione fra k serie di osservazioni (con $k \geq 2$).

I tests non-parametrici costituiscono una valida alternativa nei confronti dei più conosciuti tests parametrici, specialmente nel caso di piccoli campioni, per i quali non sia possibile ipotizzare la normalità delle variabili oggetto di esame; inoltre, essi sono applicabili anche quando le variabili siano misurate secondo una scala nominale od ordinale.

Per ogni statistica presentata vengono esaminati brevemente gli aspetti teorici e più diffusamente le modalità d'uso del programma; queste vengono inoltre concretamente illustrate mediante una applicazione pratica su dati talora simulati (ad es.: il programma per la trasformazione in ranghi) ma più frequentemente ricavati da studi scientifici già pubblicati, dei quali viene indicato di volta in volta il riferimento bibliografico.

Chiunque fosse interessato alla lista dei programmi in questione, può farne richiesta al nostro reparto. Suggerimenti e commenti sui programmi realizzati saranno i benvenuti.

Roma, 5 novembre 1982

Flavia CHIAROTTI
Franco TAGGI

Reparto di Metodologie
e Modelli Biostatistici

ISTITUTO SUPERIORE DI SANITA'

LABORATORIO DI EPIDEMIOLOGIA E BIOSTATISTICA

REPARTO DI BIOSTATISTICA

GUIDA ALLA ELABORAZIONE STATISTICA

Nastro G.E.S. 7

1 - Trasformazione in ranghi
2 - Coefficiente di contingenza
3 - Coefficiente di correlazione per ranghi di Spearman: Rs
4 - Coefficiente di correlazione per ranghi di Kendall: Tau
5 - Coefficiente di correlazione parziale per ranghi di Kendall: Txy.z
6 - Coefficiente di concordanza di Kendall: W
7 - Test sui tassi in presenza di un trend di Armitage
8 - Test sul coefficiente di correlazione r di Pearson; H0: R=0
9 - Test sul coefficiente di correlazione r di Pearson; H0: R=R0
10 - Test sui coefficienti di correlazione r di Pearson; H0: R1=R2
11 - Analisi della regressione
12 - Test sui coefficienti di regressione; H0: B1=B2
13 - Analisi della regressione: adattamento programma Operators Manual (X vs Y DATA PLOT)

Inserire il numero del programma desiderato : *

1 - TRASFORMAZIONE IN RANGHI

ASPETTI TEORICI

La trasformazione in ranghi è operazione utile ai fini del calcolo della maggior parte delle statistiche distribution-free (o non-parametriche).

Dato un vettore $x = (x_1, x_2, \ldots, x_n)$, il vettore dei ranghi $r = (r_1, r_2, \ldots, r_n)$ si ottiene ordinando gli elementi del vettore x in senso non decrescente $(x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)})$ ed attribuendo a ciascun elemento il numero naturale i corrispondente al posto da esso occupato nella scala ordinata.

Nel caso di valori ex-aequo, il rango si ottiene come media aritmetica dei numeri naturali corrispondenti ai posti occupati dai dati ex-aequo nella scala ordinata.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione del programma; a richiesta dell'utente viene stampata una spiegazione più dettagliata.

INPUT DATI

L'immissione può avvenire da tastiera o da nastro, a scelta dell'utente.

Le informazioni richieste sono:

1 - numero dei dati a cui il programma deve essere applicato;

2 - opzione nastro: il numero del file in cui i dati sono registrati (RICORDA: occorre inserire il nastro in cui i dati sono registrati);

opzione tastiera: i singoli dati, che vanno inseriti separatamente, premendo dopo ognuno il tasto RETURN.

Vi è la possibilità di sostituire i dati, qualora vi sia stato errore nell'immissione; in questo caso occorre indicare il numero del dato da sostituire, ed inserire poi il dato corretto.

Vi è la possibilità di registrare i dati, inserendo il nastro per la registrazione ed indicando il numero del file in cui si vuole che i dati vengano conservati.

OUTPUT

Vengono stampati i quattro vettori:

1 - dati originari;

2 - ranghi associati ai dati originari;

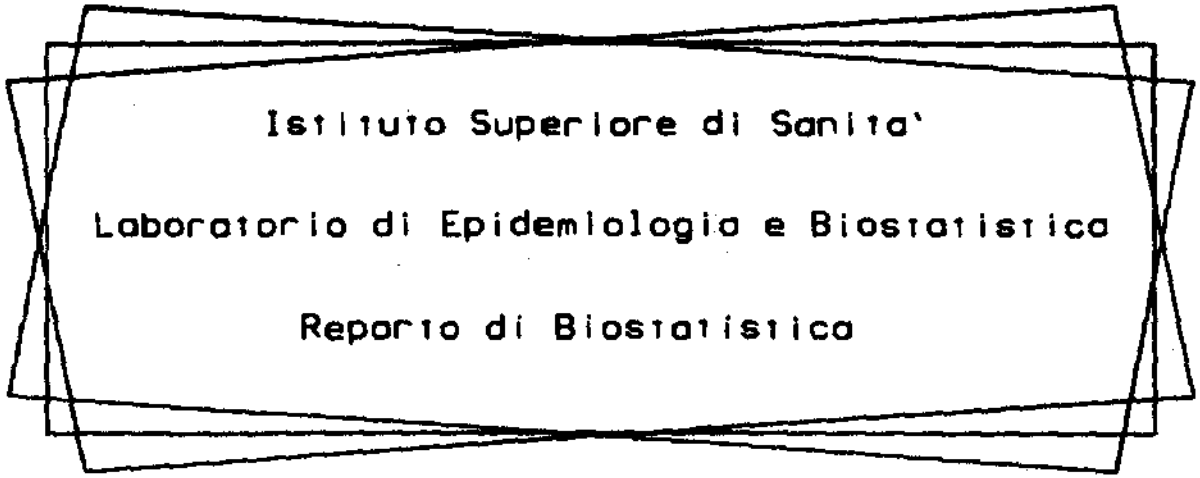
3 - dati ordinati;

4 - ranghi associati ai dati ordinati.

APPLICAZIONE

Illustriamo ora con due esempi l'utilizzazione del programma; i dati a cui esso viene applicato sono riportati nella tabella seguente.

Elem. n.	1° esempio (senza valori ex-aequo)	2° esempio (con valori ex-aequo)
1	7	14
2	41	3
3	3	27
4	2	38
5	88	15
6	15	21
7	24	27
8	9	40
9	62	15
10	30	27



Guida alla Elaborazione Statistica
Nostro G.E.S. 7

*** Programma per l'ordinamento e la trasformazione in ranghi ***

I valori numerici inseriti vengono ordinati dal minore al maggiore e vengono poi trasformati in ranghi. Al termine si avranno 4 vettori :
1) Dati originari ; 2) Ranghi associati ai dati originari ;
3) Dati ordinati ; 4) Ranghi associati ai dati ordinati .

Vuoi saperne di piu' ? s

****

Spiegazione Test

****

Lo trasformazione in ranghi dei dati ordinati consiste nell'attribuzione a ciascuno di essi del numero naturale corrispondente al posto da esso occupato nella scala ordinata.

I valori ex-aequo hanno lo stesso rango, che si ottiene come media aritmetica dei numeri naturali corrispondenti ai posti che i dati ex-aequo occupano nella scala ordinata.

Per continuare e cambiare pagina premere il tasto RETURN

TRASFORMAZIONE IN RANGHI (G.E.S. 7)

Numero dati 10

Immissione dati. Tastiera (1), Nastro (2) 1

DATO no. 1 ---> 7
DATO no. 2 ---> 41
DATO no. 3 ---> 3
DATO no. 4 ---> 2
DATO no. 5 ---> 88
DATO no. 6 ---> 15
DATO no. 7 ---> 24
DATO no. 8 ---> 9
DATO no. 9 ---> 62
DATO no. 10 ---> 30

Volete sostituire dei dati ? (si,no) n

Volete registrare i dati ? (si,no) n

****

DATI ORIGINARI	RANGHI	DATI ORDINATI	RANGHI
7	3	2	1
41	8	3	2
3	2	7	3
2	1	9	4
88	10	15	5
15	5	24	6
24	6	30	7
9	4	41	8
62	9	62	9
30	7	88	10

```
TRASFORMAZIONE IN RANGHI     (C.E.S. 7)

Numero dati    10

Immissione dati. Tastiera (1),    Nastro (2)    1

DATO no.    1    --->     14
DATO no.    2    --->     3
DATO no.    3    --->     27
DATO no.    4    --->     38
DATO no.    5    --->     15
DATO no.    6    --->     21
DATO no.    7    --->     27
DATO no.    8    --->     40
DATO no.    9    --->     15
DATO no.    10    --->     27

Volete sostituire dei dati ?    (si,no)    n

Volete registrare i dati ?    (si,no)    n

**********************************************************************

DATI ORIGINARI     RANGHI              DATI ORDINATI      RANGHI

 14                 2                   3                  1
 3                  1                   14                 2
 27                 7                   15                 3.5
 38                 9                   15                 3.5
 15                 3.5                 21                 5
 21                 5                   27                 7
 27                 7                   27                 7
 40                 10                  27                 7
 15                 3.5                 38                 9
 27                 7                   40                 10
```

2 - COEFFICIENTE DI CONTINGENZA

ASPETTI TEORICI

Il coefficiente di contingenza può essere utilizzato per valutare il grado di associazione tra due serie di attributi.

Esso utilizza le informazioni contenute nelle frequenze associate agli attributi considerati, e viene calcolato da una tabella di contingenza, applicando la formula:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

dove

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - A_{ij})^2}{A_{ij}}$$

r = numero delle righe della tabella di contingenza

c = numero delle colonne della tabella di contingenza

O_{ij} = frequenza osservata nella casella (i, j)

A_{ij} = frequenza attesa nella casella (i, j).

Per poter calcolare la significatività del coefficiente di contingenza, misurato su un campione sperimentale, ai fini della valutazione dell'esistenza di associazione tra le due serie di attributi nella popolazione da cui il campione è stato estratto, occorre conoscere i gradi di libertà del χ^2. Questi sono dati dal prodotto $(r - 1)(c - 1)$; la significatività di C viene calcolata in base alla significatività del χ^2.

I limiti presentati dal coefficiente di contingenza sono costituiti dal fatto che:

1) esso si mantiene sempre inferiore all'unità, anche quando vi è perfetta correlazione fra le variabili considerate; il limite superiore della statistica dipende dalle dimensioni della tabella di contingenza, rendendo quindi non comparabili coefficienti calcolati su tabelle di diversa dimensione;
2) il coefficiente è calcolato in base al χ^2, e quindi fa proprie tutte le limitazioni di questo (frequenze attese almeno superiori a certi valori minimi);

3) il coefficiente non è comparabile con le altre misure di correlazione che si vedranno nel seguito.

I vantaggi presentati dal coefficiente di contingenza sono:

1) applicabilità a variabili misurate secondo la scala nominale;
2) assenza di postulati sulla forma e sulla continuità della distribuzione delle variabili esaminate.

Per questi motivi, è utile adoperare il coefficiente di contingenza come misura di associazione quando non sia applicabile un indice diverso.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione delle caratteristiche principali del test e del suo campo di impiego.

A richiesta dell'utente viene stampata una spiegazione più dettagliata, con l'indicazione del significato di H_0 (ipotesi nulla) e H_1 (ipotesi alternativa) per la prova di ipotesi.

INPUT DATI

Le informazioni richieste sono:

1 - il valore del χ^2 (chi quadrato) determinato dal campione;
2 - il numero di osservazioni complessive su cui è stato calcolato;
---- queste due informazioni sono necessarie per calcolare il coefficiente.
3 - Il numero dei gradi di libertà del χ^2 ; oppure (nel caso in cui tale numero non sia noto) il numero di righe (r) ed il numero di colonne (c) della tabella su cui il χ^2 è stato calcolato.

N.B.: g. d. l. = (r - 1) * (c - 1).

Qualora non si conoscano neppure le dimensioni della tabella, non potrà calcolarsi la significatività del coefficiente, il che rende pressocché inutile la conoscenza del valore del coefficiente.

OUTPUT

Viene stampato il valore del coefficiente di contingenza, e, qualora sia stato possibile il calcolo, il valore del livello di rischio (α) nel rifiutare H_0 quando è vera.

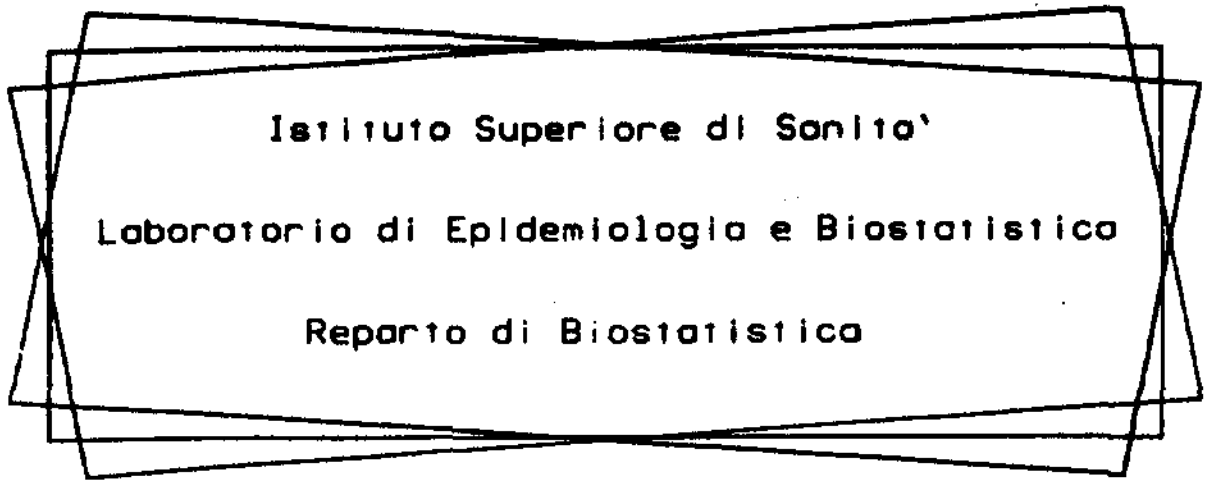
Indicando con α^* il livello di rischio prefissato giudicato accettabile:
se $\alpha^* < \alpha$ non si può rifiutare H_0;
se $\alpha^* > \alpha$ si rifiuta H_0.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. Petrelli et al., The head louse in Italy: an epidemiological study among schoolchildren, in The Royal Society of Health Journal, vol. 100 n. 2, p. 64, 1980.

Infestazione di pidocchi nelle scuole romane per tipo di scuola.

Tipo di scuola	n. bambini infestati	n. bambini non infestati	n. bambini esaminati
Materna	22	301	323
Elementare	90	816	906
Media inferiore	78	681	759
Totale	190	1798	1988



Guida alla Elaborazione Statistica
Nostro G.E.S. 7

*** Programma per il calcolo del coefficiente di contingenza ***

Il coefficiente di contingenza C e' una misura del grado di associazione o relazione fra due serie di attributi. E' utile solo quando si abbia una informazione di tipo nominale di una o di entrambe le serie di attributi.

Vuoi saperne di piu' ? s

****

Spiegazione test

****

Per applicare il coefficiente non e' necessario ordinare le categorie in una maniera particolare ; infatti il coefficiente avra' lo stesso valore in qualsiasi maniera vengono disposte le categorie nelle righe e nelle colonne.

Possiamo valutare se la correlazione osservata fra le 2 serie di attributi nel campione sia o no significativa dell'esistenza di correlazione nella popolazione da cui il campione e' estratto.
Per questo scopo poniamo:

H0: non esiste correlazione nella popolazione
H1: esiste la correlazione nella popolazione.

Il test viene condotto basandosi sul valore del chi quadrato determinato nel campione; e' necessario quindi conoscere i gradi di liberta' del chi quadrato.

Per continuare e cambiare pagina premere il tasto RETURN

Avete una riga o dispaelzione per l'infestazione
Infestazione di pidocchi in scuole romane per tipo di scuola

TEST SUL COEFFICIENTE DI CONTINGENZA (G.E.S. 7)
Infestazione di pidocchi in scuole romane per tipo di scuola

Valore del CHI QUADRATO determinato nel campione: X2 = 3.421

Numero di osservazioni complessive del campione: N = 1988

Conoscete i gradi di liberta` del CHI QUADRATO ? (si,no) s

Inserite i gradi di liberta` del CHI QUADRATO: g.d.l.= 2

****

Il valore del coefficiente di contingenza e` = 0.041447180823 4

****

H0: le 2 serie di attributi non sono correlate nella popolazione.

Il livello di rischio (alfa) nel rifiutare H0 (quando e` vera) e` = 0.18078

****

Volete ripetere il test (si,no) ?

TEST SUL COEFFICIENTE DI CONTINGENZA (G.E.S. 7)
Infestazione di pidocchi in scuole romane per tipo di scuola

Valore del CHI QUADRATO determinato nel campione: X2 = 3.421

Numero di osservazioni complessive del campione: N = 1988

Conoscete i gradi di liberta' del CHI QUADRATO ? (si,no) n

Conoscete il numero di righe e di colonne della tabella su cui e' stato calcolato il CHI QUADRATO ? (si,no) s

Ricordate che i GRADI DI LIBERTA' del chi quadrato calcolato su una tabella RxC sono = (R-1)*(C-1).

Numero righe: R = 3 Numero colonne: C = 2

Gradi di liberta' risultanti: g.d.l. = 2

****

Il valore del coefficiente di contingenza e' = 0.041447180823 4

****

H0: le 2 serie di attributi non sono correlate nella popolazione.

Il livello di rischio (alfa) nel rifiutare H0 (quando e' vera) e' = 0.18078

****

Volete ripetere il test (si,no) ?

TEST SUL COEFFICIENTE DI CONTINGENZA (C.E.S. 7)
Infestazione di pidocchi in scuole romane per tipo di scuola

Valore del CHI QUADRATO determinato nel campione: X2 = 3.421

Numero di osservazioni complessive del campione: N = 1988

Conoscete i gradi di liberta' del CHI QUADRATO ? (si,no) n

Conoscete il numero di righe e di colonne della tabella su cui e' stato calcolato il CHI QUADRATO ? (si,no) n

****

Il valore del coefficiente di contingenza e' = 0.0414471808234

****

Non si puo' calcolare la significativita', poiche' non si conoscono i gradi di liberta' del CHI QUADRATO

****

Volete ripetere il test (si,no) ?

3 - COEFFICIENTE DI CORRELAZIONE PER RANGHI DI SPEARMAN r_s

ASPETTI TEORICI

Il coefficiente di correlazione per ranghi di Spearman r_s può essere utilizzato per valutare il grado di associazione fra due serie di attributi, relative ad N elementi, ciascuna misurata almeno secondo una scala ordinale.

Esso utilizza le informazioni contenute nei ranghi associati alle variabili considerate; qualora vi fosse associazione perfetta, i ranghi di uno stesso elemento in relazione alle due serie di attributi risulterebbero uguali, e la loro differenza sarebbe quindi nulla. Un indicatore del grado di correlazione può quindi essere dato dalla somma dei quadrati delle differenze tra ranghi corrispondenti. Si usano i quadrati per evitare compensazioni fra differenze di segno opposto.

Il coefficiente di correlazione per ranghi di Spearman si calcola in base alla formula:

$$r_s = \frac{\sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} d_i^2}{2\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}}$$

dove N = numero degli elementi considerati

d_i = differenza fra i ranghi relativi all'i-mo elemento

$$\sum_{i=1}^{N} x_i^2 = \frac{N^3 - N}{12} - \sum_{j=1}^{K} T_j \qquad \sum_{i=1}^{N} y_i^2 = \frac{N^3 - N}{12} - \sum_{l=1}^{h} T_l$$

k = numero dei gruppi di osservazioni ex-aequo nella variabile x

$$T_j = \frac{t_j^3 - t_j}{12}$$

t_j = numero di osservazioni ex-aequo in un determinato rango

h = numero dei gruppi di osservazioni ex-aequo nella variabile y

$$T_l = \frac{t_l^3 - t_l}{12}$$

t_i = come sopra, per la variabile y.

Qualora gli elementi classificati costituiscano un campione casuale, è possibile verificare l'ipotesi di non associazione tra le due variabili esaminate nella popolazione di provenienza, calcolando la significatività del coefficiente r_s.

Per campioni di numerosità N compresa fra 4 e 30 sono stati tabulati i valori critici di r_s, per $\alpha = 0.05$ e $\alpha = 0.01$.

Per campioni di numerosità N > 30, si può utilizzare la formula

$$t = r_s \sqrt{\frac{N - 2}{1 - r_s^2}} \quad ;$$

la variabile t si distribuisce approssimativamente come una t di Student con (N - 2) g.d.l.. La approssimazione cresce al crescere della dimensione campionaria.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione delle caratteristiche principali del test e del suo campo di impiego.

A richiesta dell'utente viene stampata una spiegazione più dettagliata del coefficiente e del suo campo di variazione, e del test che su tale coefficiente si basa.

INPUT DATI

L'immissione può avvenire da tastiera o da nastro, a scelta dell'utente.

Le informazioni richieste sono:

1 - numero degli elementi considerati;

2 - opzione nastro: il numero del file in cui i dati sono registrati (RICORDA: occorre inserire il nastro in cui i dati sono registrati);

opzione tastiera: le coppie di dati; i dati vanno inseriti a coppia, separandoli con una virgola, e premendo dopo ogni coppia il tasto RETURN.

Vi è la possibilità di sostituire i dati qualora vi siano stati errori nell'immissione; in questo caso occorre indicare il numero dell'elemento la cui coppia contiene il dato da sostituire, ed inserire poi la coppia di dati corretta.

Vi è la possibilità di registrare i dati, inserendo il nastro per la registrazione ed indicando il numero del file in cui si vuole che i dati vengano conservati.

OUTPUT

Vengono stampate una o più tabelle (a seconda del numero degli elementi classificati) contenenti il numero d'ordine di ciascun elemento, i ranghi relativi ad esso in ciascuna classificazione, la differenza fra i ranghi (con il proprio segno), la differenza al quadrato e la somma delle differenze al quadrato.

Viene poi stampato il valore del coefficiente.

Indichiamo con N il numero degli elementi classificati.

Se $N < 4$ non può essere calcolata la significatività del coefficiente poiché non esistono dati per il confronto.

Se $4 \leq N \leq 30$ il test è unidirezionale: occorre quindi indicare la direzione desiderata per il test. Viene stampata l'ipotesi nulla ed il risultato del test, a 2 livelli di significatività:
$\alpha = 0.05$ e $\alpha = 0.01$.

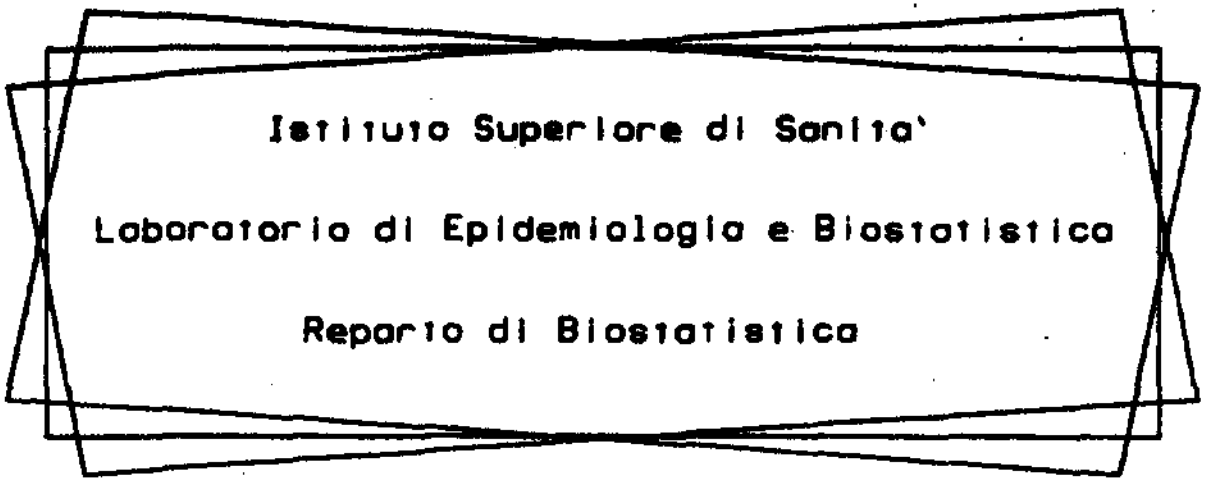
Se $30 < N$ viene stampata l'ipotesi nulla e il livello di rischio (α) che si ha nel rifiutare H_0 quando è vera, sia per il test unidirezionale, sia per il test bidirezionale. Indicando con α^* il livello di rischio prefissato ritenuto accettabile:
se $\alpha^* < \alpha$ non si può rifiutare H_0;
se $\alpha^* > \alpha$ si rifiuta H_0.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. A. Cinotti et al., Relationship between kallikrein and PRA after intravenous furosemide, in J. Endocrinol. Invest. n. 2, p. 147, 1979.

Escrezione di kallikreina urinaria (g/min) contro PRA (ng/ml/h).

Caso TG	UK	PRA
Controllo	40	0.7
15 minuti	114	3.5
30 "	88	3.7
60 "	47	3.5
120 "	26	3.3



Guida alla Elaborazione Statistica
Nastro G.E.S. 7

*** Coefficiente di correlazione per ranghi di Spearman : Rs ***

Questo coefficiente misura la associazione tra 2 variabili, entrambe misurate almeno da una scala ordinale, relative ad uno stesso gruppo di N oggetti o individui.

Vuoi saperne di piu' ? s

****
Spiegazione Test
****

Questo coefficiente si baso sulla discordanza tra le 2 serie di ranghi in cui sono posti gli N oggetti o individui relativamente alle 2 variabili
La discordanza e' misurata dalle differenze fra ranghi corrispondenti, elevate al quadrato per impedire che differenze di segno opposto si compensino.
Maggiori sono tali quadrati, maggiore e' la discordanza tra le 2 serie di ranghi, cioe' tra le 2 variabili.

Il campo di variazione del coefficiente e' -1 , +1 ;
-1= perfetta associazione inverso; +1= perfetta associazione diretta.

Se i punteggi su cui si e' applicato il calcolo del coefficiente Rs appartengono a soggetti estratti a caso dalla stessa popolazione, si puo' stabilire con essi se le 2 variabili sono associate nella popolazione.

Per continuare e cambiare pagina premere il tasto RETURN

COEFFICIENTE DI CORRELAZIONE PER RANGHI DI SPEARMAN: Rs (G.E.S. 7)
Avete una riga per l'intestazione
Urinary kallikrein excretion (g/min) versus PRA (ng/ml/h)

Numero delle unita' statistiche considerate: N= 5

Immissione dati: Tastiera (1) , Nastro (2) ?

Inserire le coppie di dati relativi a ciascuna unita' classificata

UNITA' STAT.		1a CLASSIF. , 2a CLASSIF.
no. 1	--->	40,0.7
no. 2	--->	114,3.5
no. 3	--->	88,3.7
no. 4	--->	47,3.5
no. 5	--->	26,3.3

Volete sostituire dei dati ? (si,no) n

Volete registrare i dati ? (si,no) n

COEFFICIENTE DI CORRELAZIONE PER RANGHI DI SPEARMAN: RS (U.E.S. /)
Urinary kallikrein excretion (g/min) versus PRA (ng/ml/h)

UNIT. STAT.	DATI		RANGHI		d_i	d_i^2
	1a VAR.	2a VAR.	1a VAR.	2a VAR.		
1	40	0.7	2	1	1	1
2	114	3.5	5	3.5	1.5	2.25
3	88	3.7	4	5	-1	1
4	47	3.5	3	3.5	-0.5	0.25
5	26	3.3	1	2	-1	1
						S = 5.5

Per continuare e cambiare pagina premere il tasto RETURN

```
COEFFICIENTE DI CORRELAZIONE PER RANGHI DI SPEARMAN: Rs (G.E.S. 7)
Urinary kallikrein excretion ( g/min) versus PRA (ng/ml/h)

   ***   Il valore del coefficiente e'   0.71818484646

H0: mancanza di associazione fra le 2 variabili

Regione critica nella CODA POSITIVA
Valore critico all'1% = 1
Valore critico al 5% = 0.9

   ***   Non si puo' rifiutare H0 al livello di significativita' dello
         0.01
   ***   Non si puo' rifiutare H0 al livello di significativita' dello
         0.05

Regione critica nella CODA NEGATIVA
Valore critico all'1% = -1
Valore critico al 5% = -0.9

   ***   Non si puo' rifiutare H0 al livello di significativita' dello
         0.01
   ***   Non si puo' rifiutare H0 al livello di significativita' dello
         0.05

Regione critica nel test a DUE CODE
Valori critici al 2% = -1                1
Valori critici al 10% = -0.9           0.9

   ***   Non si puo' rifiutare H0 al livello di significativita' dello
         0.02
   ***   Non si puo' rifiutare H0 al livello di significativita' dello
         0.10

Volete ripetere ?
```

4 - COEFFICIENTE DI CORRELAZIONE PER RANGHI DI KENDALL τ

ASPETTI TEORICI

Il coefficiente di correlazione per ranghi di Kendall τ può essere utilizzato (come r_s) per valutare il grado di associazione tra due serie di attributi, relative ad N elementi, ciascuna misurata almeno da una scala ordinale.

Anche questo coefficiente utilizza le informazioni contenute nei ranghi corrispondenti alle variabili considerate; se vi fosse associazione perfetta tra le due variabili, ordinando gli elementi secondo l'ordine crescente dei ranghi relativi alla prima variabile, anche i ranghi relativi alla seconda variabile dovrebbero trovarsi nell'ordine naturale.

Il coefficiente τ è funzione dell'"ordine" in cui si trovano i ranghi relativi alla seconda classificazione, e può essere calcolato in base alla formula:

$$\tau = \frac{S}{\sqrt{\frac{1}{2} N (N-1) - T_x}\ \sqrt{\frac{1}{2} N (N-1) - T_y}}$$

dove:

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} s_j$$

$$s_j = \begin{cases} -1 & \text{se} \quad r_i > r_j \\ 0 & \text{se} \quad r_i = r_j \\ +1 & \text{se} \quad r_i < r_j \end{cases}$$

r_i = rango associato all'i-mo elemento (nella scala riordinata in base alla 1ª serie di ranghi) nella seconda classificazione

$$T_x = \frac{1}{2} \sum_{i=1}^{K_x} t_i (t_i - 1)$$

$$T_y = \frac{1}{2} \sum_{i=1}^{K_y} t_i (t_i - 1)$$

k_x = numero dei gruppi di osservazioni ex-aequo nella variabile x

k_y = numero dei gruppi di osservazioni ex-aequo nella variabile y

t_j = numero di osservazioni ex-aequo in un determinato rango (var. x o y).

Qualora gli N elementi classificati costituiscano un campione casuale, è possibile calcolare la significatività del coefficiente τ , in modo da analizzare l'ipotesi di non associazione tra le variabili nella popolazione da cui il campione stesso sia stato estratto.

Per campioni di numerosità $N \leq 10$ sono stati tabulati i livelli di significatività associati a tutti i valori che S può assumere (τ è funzione di S e le distribuzioni di campionamento dei due indicatori sono identiche in senso probabilistico).

Per campioni di numerosità $N > 10$ si può utilizzare la formula

$$z = \frac{\tau}{\sqrt{\frac{2(2N+5)}{9N(N-1)}}} \quad ;$$

la variabile z si distribuisce approssimativamente come una VC normale di media 0 e di varianza 1.

La approssimazione cresce al crescere della numerosità campionaria.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione delle caratteristiche principali del test e del suo campo di impiego, ed è indicato il significato dell'ipotesi nulla H_0 e dell'ipotesi alternativa H_1 per la prova di ipotesi.

A richiesta dell'utente viene stampata una spiegazione più ampia e dettagliata.

INPUT DATI

L'immissione può avvenire da tastiera o da nastro, a scelta dell'utente.

Le informazioni richieste sono:

1 - numero degli elementi classificati;

2 - opzione nastro: il numero del file in cui i dati sono registrati
(RICORDA: occorre inserire il nastro in cui i dati sono registrati);
opzione tastiera: le coppie di dati relative a ciascun elemento; i dati vanno inseriti a coppie, separandoli con una virgola, e premendo dopo ogni coppia il tasto RETURN.

Vi è la possibilità di sostituire i dati, qualora vi siano stati errori nell'immissione; in questo caso occorre indicare il numero del dato da sostituire e la classificazione a cui appartiene, ed il valore che il dato deve assumere.

Vi è la possibilità di registrare i dati, inserendo il nastro per la registrazione ed indicando il numero del file in cui si vuole che i dati vengano conservati.

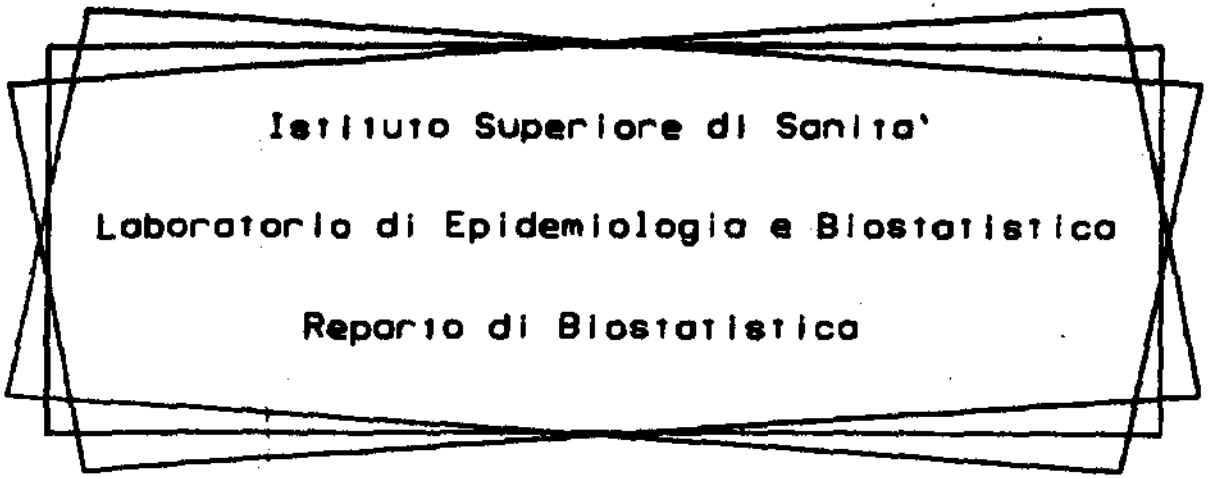
OUTPUT

Vengono stampate una o più tabelle (a seconda del numero degli elementi classificati) con i ranghi relativi a ciascun elemento nelle 2 classificazioni. Gli elementi vengono ordinati secondo il rango ad essi associato nella prima classificazione.

Viene poi stampato il valore del coefficiente τ, ed il livello di rischio (α) che si ha nel rifiutare H_0 quando è vera, sia nel caso di test unidirezionale (a una coda), sia nel caso di test bidirezionale (a due code). Indicando con α^* il livello di rischio prefissato ritenuto accettabile:
se $\alpha^* < \alpha$ non si può rifiutare H_0;
se $\alpha^* > \alpha$ si rifiuta H_0.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato sono gli stessi utilizzati nell'applicazione del programma n. 3 (coefficiente di correlazione per ranghi di Spearman r_s); essi sono tratti da:
G. A. Cinotti et al., Relationship between kallikrein and PRA after intravenous furosemide, in J. Endocrinol. Invest. n. 2, p. 147, 1979.



Guida alla Elaborazione Statistica
Nostro G.E.S. 7

*** Coefficiente di correlazione per ranghi di Kendall : Tau ***

Questo coefficiente puo' essere applicato come misura di correlazione
tra 2 variabili, ciascuna misurata almeno da una scala ordinale.
Tale coefficiente e' soggetto a test di significativita', ponendo:
H0 : assenza di associazione fra le 2 variabili nella popolazione;
H1 : associazione fra le 2 variabili nella popolazione.

Vuoi saperne di piu' ?

****

Spiegazione Test

****

Ad ogni elemento viene attribuito il rango in relazione alle 2 variabili e gli elementi vengono poi ordinati secondo il rango relativo a una delle due variabili.

Si determina poi il numero delle coppie di ranghi, relativi all'altra variabile, disposte nell'ordine corretto (naturale) fra di loro.

Ad ogni coppia che si trova nell'ordine corretto viene assegnato il punteggio +1 ; ad ogni coppia che non vi si trova viene assegnato il punteggio -1.

Il totale di tutti i punteggi e' rapportato al massimo totale possibile, che si avrebbe qualora le 2 variabili variassero concordemente.

Il coefficiente varia quindi fra -1 e +1.

Per continuare e cambiare pagina premere il tasto RETURN

COEFFICIENTE DI CORRELAZIONE PER RANGHI DI KENDALL: Tau (G.E.S. 7)
Avete una riga a disposizione per l'intestazione
Urinary kallikrein excretion (g/min) versus PRA (ng/ml/h)

Numero delle unita' statistiche classificate 5

Immissione dati. Tastiera (1), Nastro (2) 1

UNIT. STAT. CLASS. n.1 , CLASS. n.2

n. 1 ---> 40,0.7
n. 2 ---> 114,3.5
n. 3 ---> 88,3.7
n. 4 ---> 47,3.5
n. 5 ---> 26,3.3

Volete sostituire dei dati ? (si,no) n

Volete registrare i dati ? (si,no) n

Per continuare e cambiare pagina premere il tasto <RETURN>

COEFFICIENTE DI CORRELAZIONE PER RANGHI DI KENDALL TAU (G.E.S. 7)
Urinary kallikrein excretion (g/min) versus PRA (ng/ml/h)

UNITA' STAT.	VAR. 1		VAR. 2	
	DATO	RANGO	DATO	RANGO
5	26	1	3.3	2
1	40	2	0.7	1
4	47	3	3.5	3.5
3	88	4	3.7	5
2	114	5	3.5	3.5

Per continuare e cambiare pagina premere il tasto RETURN

COEFFICIENTE DI CORRELAZIONE PER RANGHI DI KENDALL: Tau (G.E.S. 7)
Urinary kallikrein excretion (g/min) versus PRA (ng/ml/h)

****

Il valore di TAU e` 0.527046276695

*** H0: assenza di associazione fra le 2 variabili nella popolazione

Il livello di rischio (alfa) nel rifiutare H0 (quando e` vera) e`=

Test unidirezionale (a una coda) ---> 0.1795

Test bidirezionale (a due code) ---> 0.359

Volete ripetere il test ? (si,no)

5 - COEFFICIENTE DI CORRELAZIONE PARZIALE PER RANGHI DI KENDALL

$$\tau_{xy.z}$$

ASPETTI TEORICI

Il coefficiente di correlazione per ranghi di Kendall $\tau_{xy.z}$ può essere utilizzato ogni qualvolta si voglia misurare in base ad un gruppo di N elementi l'associazione tra due variabili x e y, mantenendo costante una terza variabile z che si suppone possa influire su tale associazione. Per il calcolo del coefficiente si può utilizzare la formula

$$\tau_{xy.z} = \frac{\tau_{xy} - \tau_{yz} \cdot \tau_{xz}}{\sqrt{(1 - \tau_{xz}^2)(1 - \tau_{yz}^2)}}$$

dove τ_{xy}, τ_{xz}, τ_{yz} sono i coefficienti di correlazione per ranghi di Kendall relativi alle variabili xy, xz, yz.

Non è possibile calcolare la significatività del coefficiente $\tau_{xy.z}$ in quanto non se ne conosce la distribuzione di campionamento.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione delle caratteristiche principali del test e del suo campo di impiego.

INPUT DATI

L'inserimento dei dati può avvenire da tastiera o da nastro, a scelta dell'utente.

N.B. : i dati relativi alla variabile tenuta costante (nel programma è indicata con Z), debbono essere inseriti per ultimi.

Le informazioni richieste sono:

1 - numero degli elementi classificati;

2 - opzione nastro: il numero del file in cui i dati sono registrati

(RICORDA: occorre inserire il nastro in cui i dati sono registrati);

opzione tastiera: le terne di dati relativi a ciascune elemento; i dati vanno inseriti separatamente, premendo dopo ogni dato il tasto RETURN.

Vi è la possibilità di sostituire i dati, qualora vi sia stato errore nell'inserimento: in questo caso occorre indicare il numero dell'elemento e della variabile (X = 1, Y = 2, Z = 3) a cui si riferisce il dato da sostituire, ed il dato corretto.

Vi è la possibilità di registrare i dati, inserendo il nastro per la registrazione ed indicando il numero del file in cui si vuole che i dati vengano conservati.

OUTPUT

Viene stampato il valore del coefficiente di correlazione parziale.

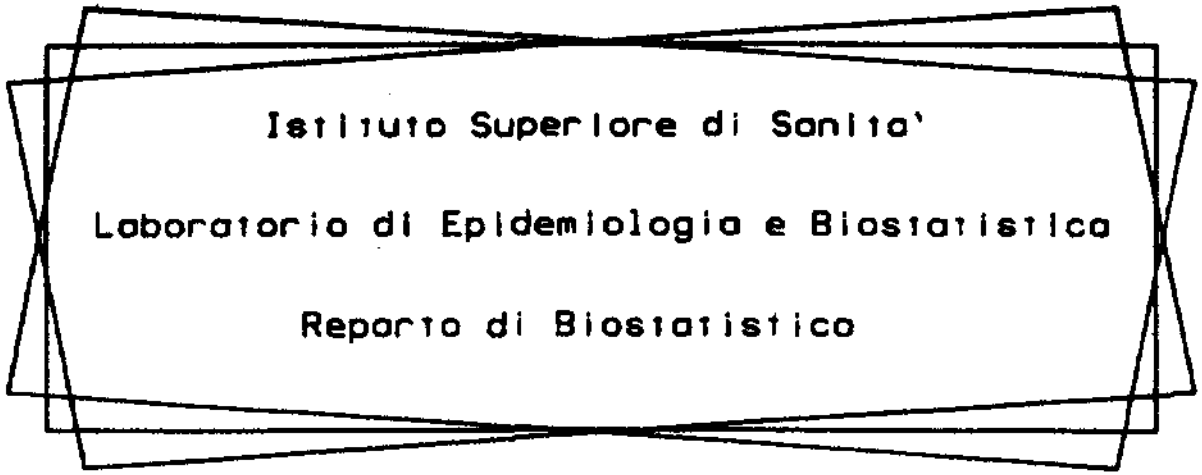
Non ne viene calcolata la significatività, poiché non si conosce ancora la sua distribuzione di campionamento.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. A. Cinotti et al., Relationship between kallikrein and PRA after intravenous furosemide, in J. Endocrinol. Invest. n. 2, p. 147, 1979.

Escrezione di kallikreina urinaria (g/min) contro PRA (ng/ml/h), ferma restando l'escrezione urinaria Na.

Caso TG	UK	PRA	UNa
Controllo	40	0.7	0.20
15 minuti	114	3.5	0.33
30 "	88	3.7	2.07
60 "	47	3.5	1.40
120 "	26	3.3	0.53



Guida alla Elaborazione Statistica
Nostro G.E.S. 7

** Coefficiente di correlazione parziale per ranghi di Kendall:Txy.z **

Questo coefficiente permette di determinare l'entita' della relazione diretta o incontaminata fra 2 variabili, tenendo costante l'effetto di 1 terza variabile che potrebbe influenzare entrambe e il grado dello loro correlazione. Tutte e tre le variabili devono essere misurate almeno da una scala ordinale.

Per continuare e cambiare pagina premere il tasto RETURN

Avete una riga per l'intestazione
X = UK excretion, Y = PRA; Z = Urinary Na excretion (meq/min)

Numero delle unita' statistiche classificate 5

Immissione dati: Tastiera (1), Nastro (2) 1

** INSERIRE PER ULTIMI I DATI RELATIVI ALLA VARIABILE TENUTA COSTANTE **

Per continuare e cambiare pagina premere il tasto <RETURN>

COEFF. DI CORRELAZIONE PARZIALE PER RANGHI DI KENDALL: Txy.z (G.E.B. 7)
X = UK excretion, Y = PRA; Z = Urinary Na excretion (meq/min)

UNIT.STAT.	1	2	3	4	5
VAR. X	40	114	88	47	26
VAR. Y	0.7	3.5	3.7	3.5	3.3
VAR. * Z *	0.20	0.33	2.07	1.40	0.53

Volete sostituire dei dati ? (si,no) n

Volete registrare i dati ? (si,no) n

****

Il valore di Txy.z e' = 0.573819041757

Non e' possibile la prova di ipotesi perche' non si conosce la distribuzione di campionamento dello stimatore Txy.z

Volete ripetere il test ?

6 - COEFFICIENTE DI CONCORDANZA DI KENDALL W

ASPETTI TEORICI

Il coefficiente di concordanza W può essere utilizzato per valutare la associazione fra k serie di attributi, rilevati su N elementi e misurati ciascuno almeno da una scala ordinale.

Tale coefficiente sfrutta le informazioni contenute nei ranghi in cui vengono posti gli N elementi relativamente alle k variabili considerate.

In caso di associazione perfetta, ciascun elemento avrebbe lo stesso rango in ogni classificazione; sommando i ranghi relativi a ciascun elemento si avrebbe la serie k, 2k, ..., Nk.

Qualora invece non vi fosse associazione, le somme dei ranghi dovrebbero essere approssimativamente uguali.

Il coefficiente W è funzione del grado di varianza tra le N somme di ranghi (R_1, R_2, ..., R_N), e può essere calcolato in base alla formula

$$W = \frac{S}{\frac{1}{12} k^2 (N^3 - N) - k \sum_T T} \qquad \text{dove}$$

k = numero delle variabili (e classificazioni) considerate

N = numero degli elementi classificati

$$S = \sum_{j=1}^{N} \left(R_j - \frac{\sum_{j=1}^{N} R_j}{N} \right)^2$$

$$R_j = \sum_{i=1}^{k} r_{ij}$$ (r_{ij} = rango relativo al j-mo elemento nella i-ma classificazione).

$$T = \frac{\sum (t^3 - t)}{12}$$

t = numero delle osservazioni in un gruppo avente lo stesso valore di rango.

$\sum$ è estesa a tutti i gruppi di osservazioni ex-aequo in ognuna delle k classificazioni.

Qualora gli N elementi classificati costitutiscano un campione casuale, è possibile calcolare la significatività di W per valutare l'ipotesi di non associazione tra le k variabili osservate, nella popolazione da cui il campione stesso sia stato estratto.

Nel caso di piccoli campioni ($3 \leq N \leq 7$, $3 \leq k \leq 20$) sono stati tabulati i valori critici di S ai livelli di significatività $\alpha = 0.05$ e $\alpha = 0.01$.

Nel caso di grandi campioni ($N > 7$) si può utilizzare la formula

$$X^2 = k\,(N - 1)\,W \quad .$$

La variabile X^2 si distribuisce approssimativamente come un χ^2 con $(N - 1)$ g.d.l.; la approssimazione cresce al crescere della dimensione campionaria.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione delle caratteristiche principali del test e del suo campo di impiego.

A richiesta dell'utente viene stampata una spiegazione più dettagliata, con l'indicazione del significato di H_0 (ipotesi nulla) e H_1 (ipotesi alternativa), per la prova di ipotesi.

INPUT DATI

L'immissione può avvenire da tastiera o da nastro, a scelta dell'utente.

Le informazioni richieste sono:

1 - numero delle serie di classificazioni;

2 - numero degli elementi classificati;

3 - opzione nastro: il numero del file in cui i dati sono registrati
(RICORDA: occorre inserire il nastro in cui i dati sono registrati);

opzione tastiera: i singoli dati che vanno inseriti separatamente, premendo dopo ogni dato il tasto RETURN.
I dati vanno inseriti in ordine: per ogni elemento, i dati relativi a ciascuna classificazione.

Vi è la possibilità di sostituire i dati quando vi sia stato errore nell'immissione: in questo caso occorre indicare, in ordine, il numero del dato da sostituire, la classificazione a cui appartiene ed il dato corretto.

Vi è la possibilità di registrare i dati, inserendo il nastro per la registrazione ed indicando il numero del file in cui si vogliono conservare i dati.

OUTPUT

Viene stampato il valore del coefficiente di concordanza di Kendall, il significato dell'ipotesi nulla H_0 e, indicando con N il numero di elementi classificati:

se $N \leq 7$ direttamente il risultato del test;

se $7 < N$ il livello di rischio (α) nel rifiutare H_0 quando è vera.

Indicando con α^* il livello di rischio prefissato ritenuto accettabile:

se $\alpha^* < \alpha$ non si può rifiutare H_0;

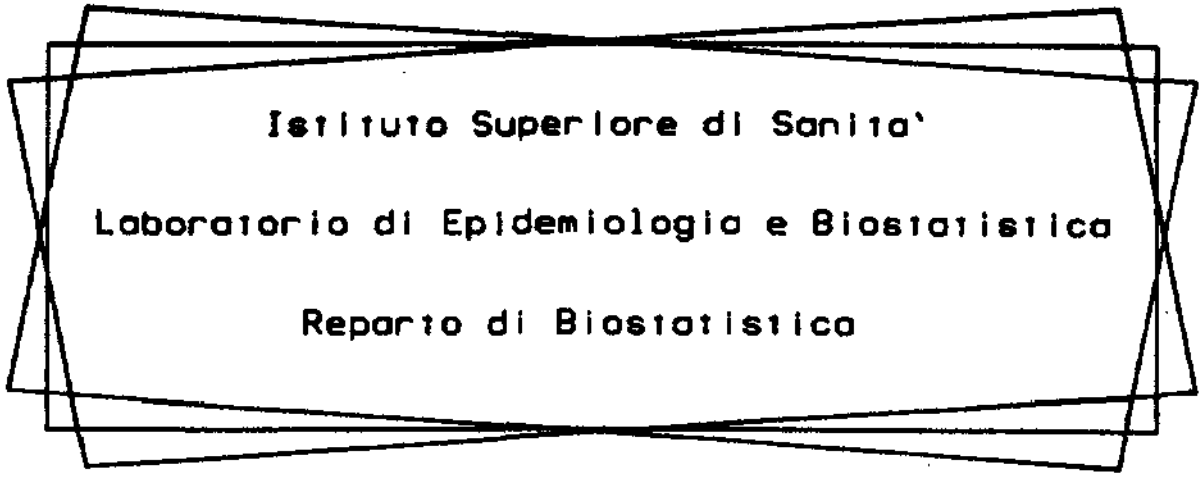
se $\alpha^* > \alpha$ si rifiuta H_0.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. A. Cinotti et al., Relationship between kallikrein and PRA after intravenous furosemide, in J. Endocrinol. Invest. n. 2, p. 147, 1979.

Comportamento della escrezione UK.

	control	15'	30'	60'	120'
TG 1	40	114	88	47	26
ZC 2	22	106	40	10	6
MM 3	14	154	124	31	18
CM 4	14	80	40	10	10
CV 5	40	170	170	60	10
CA 6	18	66	28	14	8
MG 7	7	33	26	8	12
GM 8	22	160	52	60	14



Guida alla Elaborazione Statistica
Nastro G.E.S. 7

*** Coefficiente di concordanza di Kendall : W ***

Esso misura la correlazione tra K (K>2) variabili, ciascuno misurato almeno da una scala ordinale, relative ad uno stesso gruppo di N oggetti o individui.

Vuoi saperne di piu' ? (si,no) s

****

Spiegazione test

****

Questo coefficiente si basa sulla discordanza tra le K serie di ranghi in cui sono posti gli N oggetti o individui relativamente alle K variabili.

Se vi fosse concordanza fra le K variabili, le K serie di ranghi sarebbero coincidenti, e sommando, in relazione a ciascuno entita', i K ranghi, si avrebbe la serie K, 2K, 3K,..., NK, in cui la varianza tra i diversi elementi sarebbe la massima possibile.

Il coefficiente di concordanza e' funzione di questo grado di varianza, e varia tra 0 e +1, assumendo il valore +1 nel caso di massima concordanza tra le variabili.

Se i punteggi su cui si e' applicato il calcolo del coeffic. W appartengono a soggetti estratti a caso dalla stessa popolazione, si puo' stabilire con essi se le K variabili sono associate nella popolazione, ponendo:

H0 : assenza di concordanza fra le K variabili nella popolazione.
H1 : concordanza fra le K variabili nella popolazione.

Per continuare e cambiare pagina premere il tasto RETURN

Avete una riga per intestare il test
Comportamento dell'escrezione UK

Numero delle serie di classificazioni 5

Numero delle unita' statistiche classificate 8

Immissione dati: Tastiero (1); Nastro (2) 1

COEFFICIENTE DI CONCORDANZA DI KENDALL W (G.E.D. 7)
Comportamento dell'escrezione UK

CLASS.	UNITA' STATISTICHE				
	1	2	3	4	5
1	40	114	88	47	26
2	22	106	40	10	6
3	14	154	124	31	18
4	14	80	40	10	10
5	40	170	170	60	10
6	18	66	28	14	8
7	7	33	26	8	12
8	22	160	52	60	14

Per continuare e cambiare pagina premere il tasto RETURN

```
Volete sostituire dei dati ?  (si,no)  n
Volete registrare i dati ?  (si,no)  n
```

TEST SUL COEFFICIENTE DI CONCORDANZA DI KENDALL: W (G.E.S. 7)

Comportamento dell'escrezione UK

****

Il valore del coeffic. di concordanza di Kendall W e`= 0.813291139241

Il valore del numeratore di W e` S = 514

****

H0 = assenza di associazione tra le variabili nella popolazione

****

Valore critico di S al 5% = 183.7

Valore critico di S all'1% = 242.7

Si rifiuta H0 al livello di significativita` del 5%
Si rifiuta H0 al livello di significativita` dell'1%

****

Volete ripetere il test ? (si,no)

7 - TEST SUI TASSI IN PRESENZA DI UN TREND DI ARMITAGE

ASPETTI TEORICI

Ogniqualvolta si abbia una successione di tassi calcolati in diversi gruppi di elementi, gruppi tali da poter essere disposti secondo un loro ordine naturale, si può applicare il χ^2 per il trend di Armitage, ottenuto ricorrendo ad una modificazione nel test del χ^2, per valutare l'ipotesi di esistenza di regressione della variabile "tasso" rispetto alla variabile "numero d'ordine del gruppo" corrispondente.

Il χ^2 per il trend di Armitage può essere calcolato in base alla formula:

$$\chi^2 = \frac{\sum_{j=1}^{k} N_j \left[\sum_{j=1}^{k} N_j \sum_{j=1}^{k} E_j x_j - \sum_{j=1}^{k} E_j \sum_{j=1}^{k} N_j x_j \right]^2}{\sum_{j=1}^{k} E_j \sum_{j=1}^{k} (N_j - E_j) \left[\sum_{j=1}^{k} N_j \sum_{j=1}^{k} N_j x_j^2 - \left(\sum_{j=1}^{k} N_j x_j \right)^2 \right]}$$

dove

k = numero di gruppi considerati

N_j = numero di elementi nel gruppo j-mo

E_j = numero di elementi colpiti nel gruppo j-mo

x_j = numero d'ordine relativo al gruppo j-mo

MEMO: il tasso R_j relativo al gruppo j-mo è dato dal rapporto E_j/N_j.

La variabile così ottenuta si distribuisce come una VC χ^2 con 1 g.d.l..

Su tale statistica si basa il test per la verifica dell'ipotesi di assenza del trend nei tassi, nella popolazione da cui gli elementi complessivamente esaminati siano stati casualmente estratti.

La statistica data dalla differenza fra il χ^2 complessivo (con N - 1 g.d.l.) e il χ^2 per il trend (con 1 g.d.l.) è una VC χ^2 (con N - 2 g.d.l.).

Su di essa si basa il test per la verifica dell'ipotesi di linearità della regressione.

MODALITA' D'USO

Inizialmente vi è una breve spiegazione delle caratteristiche principali del test.

A richiesta dell'utente viene stampata una spiegazione più dettagliata, con l'indicazione del significato di H_0 (ipotesi nulla) e H_1 (ipotesi alternativa), per prova di ipotesi.

INPUT DATI

L'immissione può avvenire da tastiera o da nastro, a scelta dell'utente.
Le informazioni richieste sono:

1 - numero delle classi esaminate;

2 - opzione nastro: il numero del file in cui i dati sono registrati (RICORDA: occorre inserire il nastro in cui i dati sono registrati);

opzione tastiera: le coppie di dati (positivi, esposti; negativi, non esposti) relative a ciascuna classe. I dati vanni inseriti separatamente, premendo dopo ogni dato il tasto RETURN.

N.B.: occorre che le classi siano disposte secondo l'ordine (che si suppone esista) crescente: ciascuna di esse verrà infatti contrassegnata dal rispettivo rango (peso della classe).

Vi è la possibilità di sostituire i dati quando vi siano stati errori nell'immissione; occorre indicare il peso della classe a cui appartiene il dato da sostituire, e la coppia di dati corretta.

Vi è la possibilità di registrare i dati: occorre indicare il numero del file in cui li si vuole conservare, dopo aver inserito il nastro per la registrazione.

OUTPUT

Viene stampato il valore del χ^2 (chi quadrato) e ne viene calcolata la significatività: viene stampato il livello di rischio (α) che si ha nel rifiutare H_0 quando è vera.

Indicando con α^* il livello di rischio prefissato ritenuto accettabile:

se $\alpha^* < \alpha$ non si può rifiutare H_0;

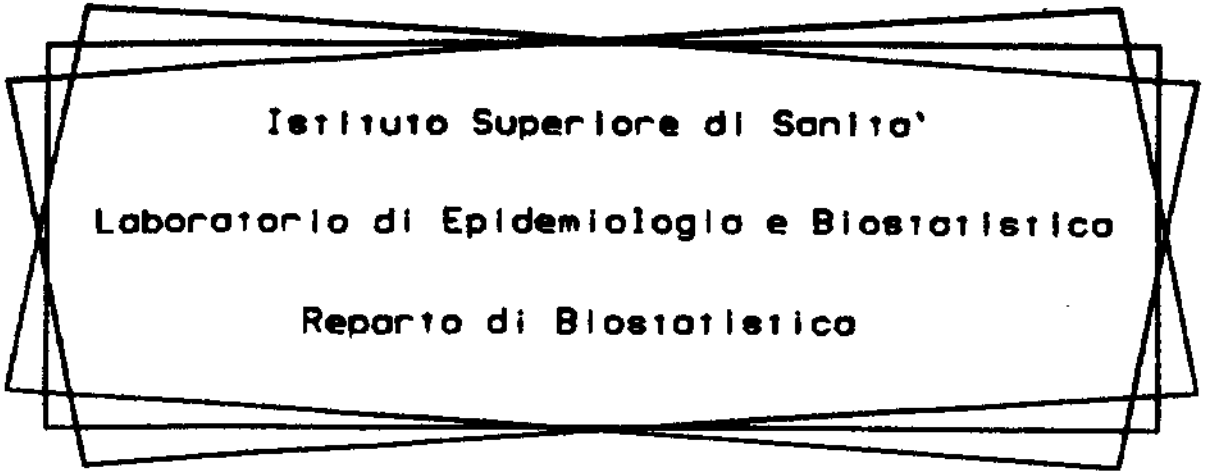
se $\alpha^* > \alpha$ si rifiuta H_0.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. Morisi et al., Programma comunitario sulla sorveglianza biologica della popolazione contro il rischio di saturnismo, Risultati Italiani: Fase I (1978-'79), ISTISAN 1980/35.

Punteggio ISS per le bevande e positività alla piombemia.

TOTALE E	0.0÷0.2	0.2÷0.4	0.4÷0.6	0.6÷0.8	0.8÷1.0
positivi	2	1	11	8	6
negativi	71	39	87	17	28
%positivi	2.7	2.5	11.2	32.0	17.6



Guida alla Elaborazione Statistica
Nostro C.E.S. 7

*** Test sui tassi in presenza di un trend di Armitage ***

Il test viene condotto per verificare se la eventuale differenza fra i tassi sia tale da dar luogo ad un andamento (trend) crescente o decre - scente dei tassi in funzione di una certa variabile.

Vuoi saperne di piu' ? s

****
Spiegazione test
****

Questo test e' applicabile ogni volta che le diverse classi di individui per le quali si confrontano i tassi possono essere disposte secondo un loro ordine naturale: in tal caso e' possibile attribuire ad ognuna di esse un valore numerico progressivo (numero d'ordine).

E' quindi importante ricordare che ogni volta che e' lecito aspettarsi un trend lineare dei tassi in funzione dell'ordine naturale con cui si susseguono i rispettivi gruppi, e' indicato saggiare direttamente la significativita' del trend facendo uso di questo test CHI QUADRO (1 g.d.l)

Per la prova di ipotesi si pone:
H0: assenza di trend nelle proporzioni ;
H1: presenza di trend nelle proporzioni .

La differenza tra le due statistiche (chi quadrato complessivo e chi quadrato con 1 g.d.l.), puo' essere considerata come una statistica CHI QUADRATO (n-2 g.d.l.), che saggia l'ipotesi dello scarto dalla linearita' della regressione delle proporzioni rispetto alle intensita' della variabile considerata.

Per la prova di ipotesi si pone:
H0: linearita' della regressione ;
H1: non linearita' della regressione ;

Per continuare e cambiare pagina premere il tasto RETURN

Avete una riga a disposizione per l'intestazione
Punteggio ISS per le bevande e positivita' alla piombemia

Inserite il numero di classi esaminate 5

Immissione dati: TASTIERA (1) , NASTRO (2) 1

Punteggio ISS per le bevande e positivita' alla piombemia

C L A S S I

PESI	1	2	3	4	5
POSITIVI	2	1	11	8	6
NEGATIVI	71	39	87	17	28
POS.%	2.7	2.5	11.2	32	17.6

Volete sostituire dei dati (si,no) ? n

Volete registrare i dati ? (si,no) n

TEST SUI TASSI IN PRESENZA DI UN TREND: ARMITAGE (G.E.S. 7)
Punteggio ISS per le bevande e positivita' alla piombemia

H0: assenza di trend nelle proporzioni

*** Il valore del CHI QUADRATO e': X2 = 14.1749871491 con 1 g.d.l.

*** Il livello di rischio (alfa) nel rifiutare H0 (quando e' vera) e'= 1.6657E-4

H0: linearita' della regressione

*** Il valore del CHI QUADRATO e': X2 = 7.66067393645 con 3 g.d.l.

*** Il livello di rischio (alfa) nel rifiutare H0 (quando e' vera) e'= 0.053571

Volete ripetere il test (si,no) ?

ASPETTI TEORICI

8, 9, 10 - TESTS SUL COEFFICIENTE DI CORRELAZIONE r DI BRAVAIS-PEARSON

Il coefficiente di correlazione di Bravais-Pearson r (o ρ) misura la correlazione lineare tra due variabili x e y, rilevate su N elementi, assumendo valori compresi fra - 1 e + 1.

r = - 1 perfetta correlazione lineare inversa;

r = 0 assenza di correlazione lineare (assenza di correlazione di alcun tipo o presenza di correlazione non lineare) (*);

r = + 1 perfetta correlazione lineare diretta.

Considerando gli N elementi su cui siano state rilevate le variabili x e y

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$$

il coefficiente di correlazione di Bravais-Pearson può essere calcolato in base alla formula

$$r = \frac{D_{xy}}{\sqrt{D_{xx}\, D_{yy}}} = \frac{\text{Cod } xy}{\sqrt{\text{Dev } x \cdot \text{Dev } y}}$$

dove

$$D_{xy} = \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

$$D_{xx} = \sum_{i=1}^{N} (x_i - \bar{x})^2$$

$$D_{yy} = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

$$x = \sum_{i=1}^{N} x_i / N$$

$$y = \sum_{i=1}^{N} y_i / N \quad .$$

Per quanto concerne la verifica della ipotesi formulata sul valore del coefficiente ρ nella popolazione da cui gli N elementi considerati siano stati estratti casualmente, occore distinguere i 2 casi:

$H_0 : \quad \rho = 0$

$H_0 : \quad \rho = r_0 \ (\neq 0) \quad .$

(*) ciò equivale a dire che $r = 0$ è condizione necessaria, ma non sufficiente, per l'assenza di correlazione tra due variabili.

Occorre poi esaminare a parte il caso in cui si voglia analizzare l'ipotesi

$H_0 : \quad \rho_1 = \rho_2$,

ipotesi di coincidenza dei valori dei coefficienti di correlazione, misurati in base a 2 campioni estratti da popolazioni distinte, nelle popolazioni stesse.

$$H_0 : \rho = 0$$

Per piccoli campioni si utilizza la formula

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$; la variabile t si distribuisce come una t di Student con (N - 2) g.d.l..

Per grandi campioni si utilizza invece la formula

$$Z = r\sqrt{N-1}$$; la variabile Z si distribuisce come una variabile normale standardizzata N (0, 1).

$$H_0 : \rho = r_0 \ (\neq 0)$$

Posto $$z = \frac{1}{2} \lg_e \left[(1 + r)/(1 - r)\right]$$

$$E_{(z)} = \frac{1}{2} \lg_e \left[(1 + r_0)/(1 - r_0)\right]$$

si utilizza la formula

$$Z = \frac{z - E(z)}{1/\sqrt{N-3}}$$; la variabile Z si distribuisce come una variabile normale standardizzata N (0, 1).

$$H_0 : \rho_1 = \rho_2$$

Posto

$$z_1 = \frac{1}{2} \lg_e \left[(1 + r_1)/(1 - r_1)\right]$$

$$z_2 = \frac{1}{2} \lg_e \left[(1 + r_2)/(1 - r_2)\right]$$

si utilizza la formula

$$Z = \frac{z_1 - z_2}{\sqrt{1/(N_1 - 3) + 1/(N_2 - 3)}}$$; la variabile Z si distribuisce come una VC N(0, 1).

MODALITA' D'USO

$$H_0 : \rho = 0$$

Inizialmente vi è una breve spiegazione del test, con l'indicazione del significato dell'ipotesi nulla H_0 e dell'ipotesi alternativa H_1.

INPUT DATI

Le informazioni richieste sono:

1 - valore del coefficiente di correlazione r calcolato sui dati campionari;

2 - numero di coppie su cui il coefficiente di correlazione è stato calcolato.

OUTPUT

Viene stampato il livello di rischio (α) che si ha nel rifiutare H_0 quando è vera, sia nel caso del test unidirezionale (a una coda), sia nel caso del test bidirezionale (a due code).

Indicando con α^* il livello di rischio prefissato ritenuto accettabile:

se $\alpha^* < \alpha$ non si può rifiutare H_0;

se $\alpha^* > \alpha$ si rifiuta H_0.

$$H_0 : \rho = r_0 \ (\neq 0)$$

Inizialmente vi è una breve spiegazione del test, con l'indicazione del significato dell'ipotesi nulla H_0 e dell'ipotesi alternativa H_1.

INPUT DATI

Le informazioni richieste sono:

1 - valore del coefficiente di correlazione r calcolato sui dati campionari;
2 - numero di coppie su cui il coefficiente di correlazione è stato calcolato;
3 - valore atteso del coefficiente di correlazione nella popolazione.

OUTPUT

Analogo al test precedente.

$$H_0 : \rho_1 = \rho_2$$

Inizialmente vi è una breve spiegazione del test, con l'indicazione del significato dell'ipotesi nulla H_0 e dell'ipotesi alternativa H_1.

INPUT DATI

Le informazioni richieste sono:

1 - valore dei coefficienti di correlazione calcolati nei 2 campioni (r1 e r2): i dati possono essere inseriti contemporaneamente, separandoli con una virgola;
2 - numero di coppie su cui i due coefficienti sono stati calcolati (n1 e n2): come sopra.

OUTPUT

Analogo ai tests precedenti.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione dei programmi; i dati a cui essi vengono applicati, riportati nelle tabelle che seguono, sono tratti da: G. Morisi et al., Programma comunitario sulla sorveglianza biologica della popolazione contro il rischio di saturnismo, Risultati Italiani: Fase I (1978-'79), ISTISAN 1980/35.

Correlazione fra Piombemia ed Età - Maschi non esposti ; Femmine non esposte.

$$H_0 : \rho = 0$$

M.N.E.

$r = 0.925$ $\quad n = 13$

$$H_0 : \rho = r_0 \ (\neq 0)$$

M.N.E.

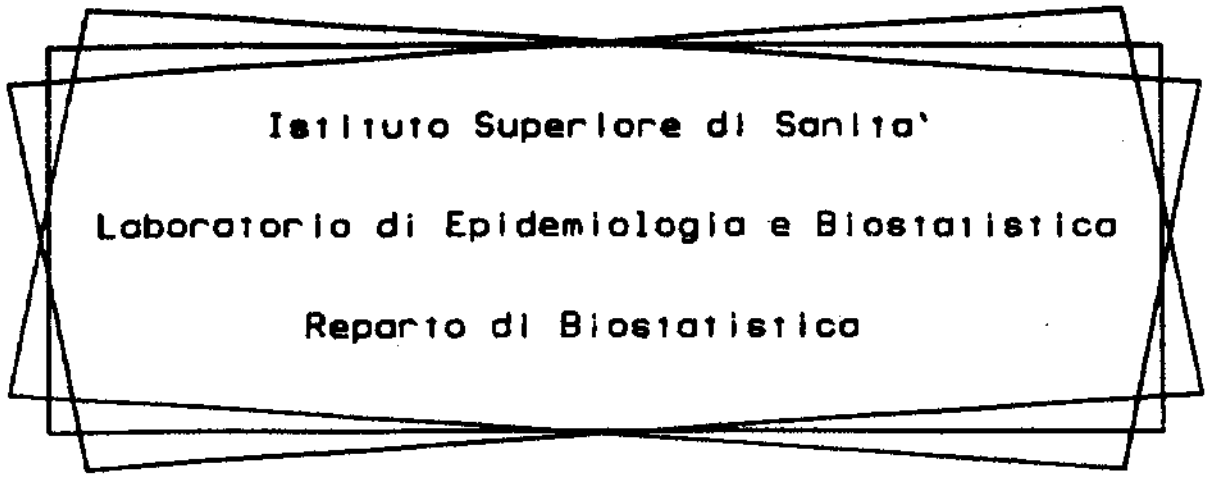
$r = 0.925$ $\quad n = 13$ $\quad r_0 = 0.9$ $\quad r_0 = 0.8$ $\quad r_0 = 0.7$

$$H_0 : \rho_1 = \rho_2$$

M.N.E. F.N.E.

$r_1 = 0.925$ $\quad r_2 = 0.915$

$n_1 = 13$ $\quad n_2 = 15$



Guida alla Elaborazione Statistica
Nastro G.E.S. 7

*** Test sul coefficiente di correlazione R di Pearson ***

Il test viene condotto per verificare le ipotesi :
H0 : coefficiente di correlazione nella popolazione = 0
H1 : coefficiente di correlazione nella popolazione < o > o ≠ 0

Vuoi saperne di piu' ? (si,no) s

****

Spiegazione test

****

Il coefficiente di correlazione R di Bravais-Pearson, che e' dato da Dxy/(sqr(Dxx * Dyy)) = Codev xy /(sqr(Dev x * Dev y)), misura la correlazione lineare fra le variabili X e Y.
Il valore del coefficiente varia fra -1 e +1

- Se le 2 variabili non sono correlate linearmente il valore atteso di R e' = 0;
- se le 2 variabili non sono correlate in alcun modo, il valore atteso di R e' ancora =0;
- se le 2 variabili sono correlate non linearmente, il valore atteso di R puo' ancora essere =0.

ESEMPIO DEL TEST

Coeff. di correlaz. nel campione: r= 0.4
Numero di coppie su cui il coefficiente e' stato calcolato: n= 50

H0: coeff. di correlaz. nella popolazione: R=0

*** Risultati ***

Il livello di rischio (alfa) nel rifiutare H0 (quando e' vera) e':

0.0020003	Test unidirezionale	(H1: R>0)
0.0040006	Test bidirezionale	(H1: R≠0)

Per continuare e cambiare pagina premere il tasto RETURN

TEST SUL COEFFICIENTE DI CORRELAZIONE DI BRAVAIS-PEARSON: R (G.E.S. 7)

Inserite il valore del coeff. di correlazione: r= 0.925

Inserite il numero di coppie su cui il coefficiente di correlazione e' stato calcolato: n= 13

****

*** H0 : R=0

*** H1 : R<0 oppure R>0 (Test a una coda)

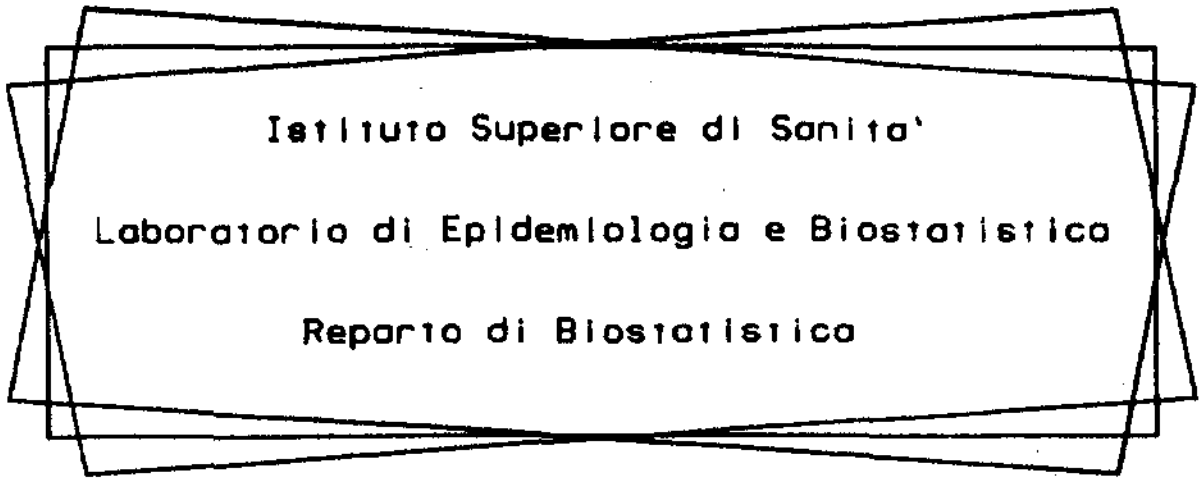
Il livello di rischio (alfa) nel rifiutare H0 e' = 2.9918E-6

*** H1 : R≠0 (Test a due code)

Il livello di rischio (alfa) nel rifiutare H0 e' = 5.9836E-6

****

Volete ripetere il test (si,no)



Guida alla Elaborazione Statistica
Nastro G.E.S. 7

*** Test sul coefficiente di correlazione R di Pearson ***

Il test viene condotto per verificare le ipotesi :
H0 : coefficiente di correlazione nella popolazione = R0
H1 : coefficiente di correlazione nella popolazione < o > o ≠ R0

Vuoi saperne di piu' ? (si,no) s

****

Spiegazione test

****

Il coefficiente di correlazione R di Bravais-Pearson, che e' dato da Dxy/(sqr(Dxx * Dyy)) = Codev xy /(sqr(Dev x * Dev y)), misura la correlazione lineare fra le variabili X e Y.
Il valore del coefficiente varia fra -1 e +1.

- Se le 2 variabili non sono correlate linearmente il valore atteso di R e' = 0;
- se le 2 variabili non sono correlate in alcun modo, il valore atteso di R e' ancora =0;
- se le 2 variabili sono correlate non linearmente, il valore atteso di R puo' ancora essere =0.

ESEMPIO DEL TEST

Coeff. di correlaz. nel campione: r= 0.4
Numero di coppie su cui il coefficiente e' stato calcolato: n= 50
Valore atteso del coefficiente nella popolazione: R= 0.5

H0: coeff. di correlaz. nella popolazione: R=0.5

*** Risultati ***

Il livello di rischio (alfa) nel rifiutare H0 (quando e' vero) e':
0.19449 Test unidirezionale (H1: R<0.5)
0.38898 Test bidirezionale (H1: R≠0.5)

Per continuare e cambiare pagina premere il tasto RETURN

TEST SUL COEFFICIENTE DI CORRELAZIONE DEL BRAVAIS-PEARSON: R (G.E.S. 7)

Inserite il valore del coeff. di correlazione: r= 0.925

Inserite il numero di coppie su cui il coefficiente di correlazione e' stato calcolato: n= 13

Inserite il valore atteso del coefficiente: R= 0.9

****

*** H0 : R=R0

*** H1 : R<R0 oppure R>R0 (Test a una coda)

Il livello di rischio (alfa) nel rifiutare H0 e' = 0.3172

*** H1 : R≠R0 (Test a due code)

Il livello di rischio (alfa) nel rifiutare H0 e' = 0.6344

****

Volete ripetere il test (si,no)

TEST SUL COEFFICIENTE DI CORRELAZIONE DEL BRAVAIS-PEARSON, R (C.E.S. 7)

Inserite il valore del coeff. di correlazione: r= 0.925

Inserite il numero di coppie su cui il coefficiente di correlazione e' stato calcolato: n= 13

Inserite il valore atteso del coefficiente: R= 0.8

****

*** H0 : R=R0

*** H1 : R<R0 oppure R>R0 (Test a una coda)

Il livello di rischio (alfa) nel rifiutare H0 e' = 0.048761

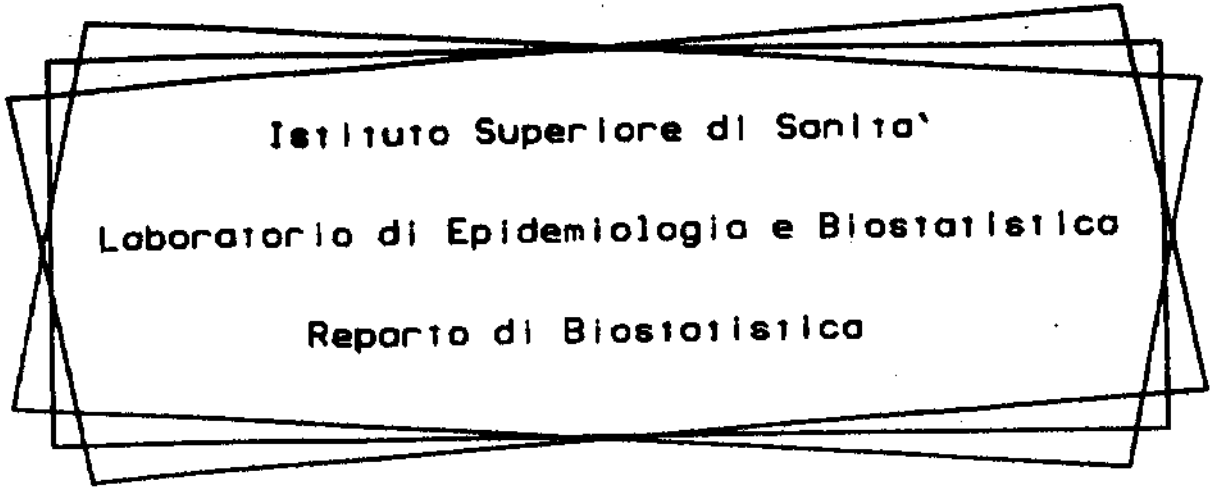
*** H1 : R≠R0 (Test a due code)

Il livello di rischio (alfa) nel rifiutare H0 e' = 0.097522

****

Volete ripetere il test (si,no)

```
TEST SUL COEFFICIENTE DI CORRELAZIONE DEL BRAVAIS-PEARSON: R (C.E.S. 7)

Inserite il valore del coeff. di correlazione:  r= 0.925

Inserite il numero di coppie su cui il coefficiente di correlazione e`
   stato calcolato:  n= 13

Inserite il valore atteso del coefficiente:  R= 0.7

**********************************************************************

   ***     H0 :  R=R0

   ***     H1 :  R<R0   oppure   R>R0   (Test a una coda)

Il livello di rischio (alfa) nel rifiutare H0 e` = 0.0084597

   ***     H1 :  R≠R0                   (Test a due code)

Il livello di rischio (alfa) nel rifiutare H0 e` = 0.0169194

**********************************************************************

Volete ripetere il test   (si,no)
```



Guida alla Elaborazione Statistica
Nostro G.E.S. 7

*** Test sul coefficiente di correlazione R di Pearson ***

Il test viene condotto su 2 campioni diversi per verificare le ipotesi :
H0 : i coeff. di correl. nelle 2 popolazioni di provenienza sono uguali
H1 : i coeff. di correl. nelle 2 popolazioni di provenienza sono diversi
(R1<R2 ; R1>R2 ; R1≠R2)

Vuoi saperne di piu` ? (si,no) s

****

Spiegazione test

****

Il coefficiente di correlazione R di Bravais-Pearson, che e' dato da Dxy/(sqr(Dxx * Dyy)) = Codev xy /(sqr(Dev x * Dev y)), misura la correlazione lineare fra le variabili X e Y.
Il valore del coefficiente varia fra -1 e +1.

- Se le 2 variabili non sono correlate linearmente il valore atteso di R e' = 0;
- se le 2 variabili non sono correlate in alcun modo, il valore atteso di R e' ancora =0;
- se le 2 variabili sono correlate non linearmente, il valore atteso di R puo' ancora essere =0.

ESEMPIO DEL TEST

Coefficienti di correlaz. nei campioni: r1= 0.4 ; r2= 0.7
Numero di coppie su cui i coeff. sono stati calcolati: n1= 40 ; n2= 25

H0: coeff. di correlaz. nella popolazione: R1=R2

*** Risultati ***

Il livello di rischio (alfa) nel rifiutare H0 (quando e' vera) e':

0.049688	Test unidirezionale	(H1: R1<R2)
0.099376	Test bidirezionale	(H1: R1≠R2)

Per continuare e cambiare pagina premere il tasto RETURN

TEST SUL COEFFICIENTE DI CORRELAZIONE DI BRAVAIS-PEARSON: R (G.E.S. 7)

Inserite i valori dei coeff. di correlazione: r1 = 0.925
r2 = 0.915

Inserite il numero di coppie su cui i coefficienti di correlazione sono stati calcolati: n1 = 13
n2 = 15

****

*** H0 : R1=R2

*** H1 : R1<R2 oppure R1>R2 (Test a una coda)

Il livello di rischio (alfa) nel rifiutare H0 e' = 0.4395

*** H1 : R1≠R2 (Test a due code)

Il livello di rischio (alfa) nel rifiutare H0 e' = 0.879

****

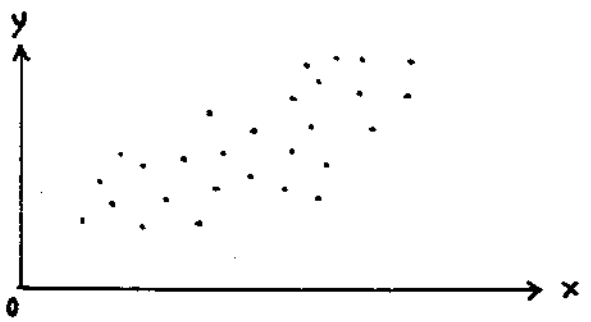
Volete ripetere il test ? (si,no)

11 - REGRESSIONE LINEARE

ASPETTI TEORICI

Si considerino N elementi, sui quali siano state rilevate le intensità delle due variabili x e y: si abbiano quindi le N coppie di valori $(x_1, y_1), \ldots, (x_N, y_N)$.

Rappresentandole su di un piano, secondo un sistema di assi cartesiani ortogonali, si avrà una nuvola (o scatter) di punti.



E' possibile interpolare tale nuvola di punti con una retta di equazione: $y = a + bx$; i parametri a e b (intercetta e coefficiente angolare o coefficiente di regressione della retta), possono essere determinati in base al criterio dei minimi quadrati.

Con questo criterio si individua la retta che minimizza la somma dei quadrati degli scarti fra i singoli punti e la retta stessa.

Supponendo le ascisse dei punti (cioè la variabile x) esatte e le ordinate (la variabile y) affette da errori, si tende a soddisfare la condizione:

$$\sum_{i=1}^{N} \left[y_i - (a + b\, x_i) \right]^2 = \min \quad .$$

La retta che soddisfa questa condizione ha come parametri:

$$a = \overline{y} - b\overline{x}$$

$$b = D_{xy} / D_{xx}$$

dove

$$\overline{y} = \sum_{i=1}^{N} y_i / N \qquad \overline{x} = \sum_{i=1}^{N} x_i / N$$

$$D_{xy} = \sum_{i=1}^{N} (x_i - \overline{x})\,(y_i - \overline{y}) = \sum_{i=1}^{N} x_i\, y_i - N\, \overline{x}\, \overline{y}$$

$$D_{xx} = \sum_{i=1}^{N} (x_i - x)^2 = \sum_{i=1}^{N} x_i - N\, \overline{x}^2$$

E' possibile calcolare l'errore standard del coefficiente di regressione (S_b) e della intercetta (S_a); qualora gli N elementi considerati costituiscano

un campione casuale, è possibile analizzare, in base ai dati campionari, le ipotesi riguardanti i valori che i parametri assumono nella popolazione di provenienza:

$H_0 : a = a_0$ $\qquad$ $H_1 : a \neq a_0$

$H_0 : b = b_0$ $\qquad$ $H_1 : b \neq b_0$

A tale scopo si utilizzano le statistiche:

$$t_a = \frac{a - a_0}{S_a} \qquad t_b = \frac{b - b_0}{S_b}$$

Le due statistiche si distribuiscono come una VC t di Student con (N-2) g.d.l..

Il coefficiente di regressione b è proporzionale al coefficiente di correlazione di Bravais-Pearson r : precisamente

$$r = \frac{D_{xy}}{\sqrt{D_{xx}\, D_{yy}}} = b\sqrt{\frac{D_{xx}}{D_{yy}}} \Longrightarrow b = r\sqrt{\frac{D_{yy}}{D_{xx}}}$$

MODALITA' D'USO

A) ELABORAZIONE NUMERICA

INPUT DATI

L'immissione può avvenire da tastiera o da nastro, a scelta dell'utente. Le informazioni richieste sono:

1 - informazioni sui dati in corso di elaborazione (intestazione del test): al massimo 72 caratteri;

2 - didascalia per l'asse delle ascisse (indicazione delle quantità riportate sull'asse X): al massimo 30 caratteri;

3 - unità di misura per l'asse delle ascisse (per esempio: kg., cm.,...): al massimo 10 caratteri;

4 - didascalia per l'asse delle ordinate: al massimo 25 caratteri;

5 - unità di misura sull'asse delle ordinate: al massimo 10 caratteri.

Se non si desiderano fornire o non si hanno le informazioni richieste, basta premere ad ogni domanda il tasto RETURN.

N.B.: Le coppie di dati da inserire (y, x) al massimo possono essere 500. Se non si conosce il numero delle coppie di dati, ed i dati vengono introdotti da tastiera, alla 500ª coppia l'input si blocca: le 500 coppie già inserite vengono conservate, e l'elaborazione viene eseguita su di esse.

6 - opzione nasto: il numero del file in cui i dati sono registrati
(RICORDA: occorre inserire il nastro in cui i dati sono registrati);
opzione tastiera: le coppie di dati, i cui elementi vanno inseriti separatamente, premendo dopo ognuno di essi il tasto RETURN: per ogni coppia, inserire prima la Y e poi la X.
Dopo l'ultima coppia di dati inserire Z, in modo da bloccare l'input senza dover contare il numero delle coppie.

Vi è la possibilità di sostituire i dati, qualora vi sia stato errore nella immissione; occorre in questo caso indicare il numero della coppia da sostituire, e la coppia di dati corretta.

Vi è la possibilità di registrare i dati, inserendo il nastro per la registrazione ed indicando il numero del file in cui si vuole vengano archiviati.

OUTPUT

Vengono stampati:

- numero di coppie inserite N;
- somma delle x_i = Sx;
- somma delle y_i = Sy;
- somma delle x_i^2 = Sx2;
- somma delle y_i^2 = Sy2;
- somma dei prodotti $x_i y_i$ = Sxy;
- devianza delle x_i = Dxx;
- devianza delle y_i = Dyy;
- codevianza delle $x_i y_i$ = Dxy;
- varianza delle x_i = Vxx;
- varianza delle y_i = Vyy;
- covarianza delle $x_i y_i$ = Cxy;

- X medio;
- Y medio;
- pendenza (coefficiente di regressione = b) con i gradi;
- errore standard della pendenza = Sb;
- intercetta con l'asse delle $\bar{Y}$ = a;
- errore standard dell'intercetta = Sa;
- errore standard della stima = Se;
- varianza della regressione;
- varianza dell'errore;
- coefficiente di correlazione lineare r e r^2;
- alcune statistiche necessarie per i tests di ipotesi;
- coefficiente di regressione forzata = b forzato;
- errore standard di b forzato;
- errore standard della stima forzata;
- statistiche necessarie per il test b forzato.

B) ELABORAZIONE GRAFICA

Vi è la possibilità di visualizzare o il solo scatter (la nuvola di punti), o lo scatter con la retta di regressione, la retta di regressione forzata e la retta a 45° (N.B.: la retta a 45° è tale in relazione alle unità di misura sull'asse X e Y, e non necessariamente rispetto allo schermo).

INPUT

L'utente può, se vuole, indicare o modificare, se si tratta del secondo grafico o dei successivi, il campo di variazione della variabile X e/o Y; inoltre può scegliere il passo, che viene accettato dall'E.E. solo se è compatibile con il campo di variazione della variabile relativa.

N.B.: per passo si intende la distanza fra 2 tacche consecutive sull'asse considerato.

Se l'utente non vuole indicare né il campo di variazione (range) né il passo, queste quantità vengono calcolate dall'E.E..

Se il campo di variazione della X e/o Y è troppo ampio, non è possibile disegnare il grafico.

OUTPUT

Vengono stampati i grafici richiesti, con le intestazioni indicate all'inizio.

C) STIMA PUNTUALE E PER INTERVALLO DELLA X, DATO Y

E' possibile avere la stima puntuale o l'intervallo di confidenza per la X, dato un certo valore della Y, al livello di significatività $\alpha = 0.05$ ($\alpha = 5\%$).

INPUT

Le informazioni richieste sono:

1 - valore della Y, numero di determinazioni della quali è valor medio (se si tratta di un dato singolo, il numero delle determinazioni sarà = 1). Queste due informazioni possono essere inserite contemporaneamente, separate da una virgola o con un segno +.

OUTPUT

Viene stampato:

Stima puntuale (della X); Intervallo di confidenza:

Estr. Inf. , Estr. Sup.

APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. Morisi et al., Programma comunitario sulla sorveglianza biologica della popolazione contro il rischio di saturnismo, Risultati Italiani: Fase I (1978-'79), ISTISAN 1980/35.

Andamento della piombemia con l'età (classi di 5 anni): maschi non esposti.

Elemento n.	X	Y
1	2.5	11.03
2	7.5	15.87
3	12.5	15.87
4	17.5	17.22
5	22.5	19.05
6	27.5	18.73
7	32.5	19.92
8	37.5	21.19
9	42.5	20.95
10	47.5	23.73
11	52.5	23.3
12	57.5	22.06
13	67.5	23.49

****

ISTITUTO SUPERIORE DI SANITA'

LABORATORIO DI EPIDEMIOLOGIA E BIOSTATISTICA

REPARTO DI BIOSTATISTICA

****

***** R E G R E S S I O N E L I N E A R E *****

Inserire le informazioni sui dati in corso di elaborazione
Andamento della piombemia con l'eta' (classi di 5 anni)

Didascalia per l'asse delle X * (max 30 caratteri)
Maschi non esposti
Unita' di misura sull'asse X (max 10 caratteri) anni

Didascalia per l'asse delle Y * (max 25 caratteri)
Piombemia
Unita' di misura sull'asse Y (max 10 caratteri) microg/dl

** Il programma accetta fino a 250 dati -Se utile, cambiate il DIMENSION

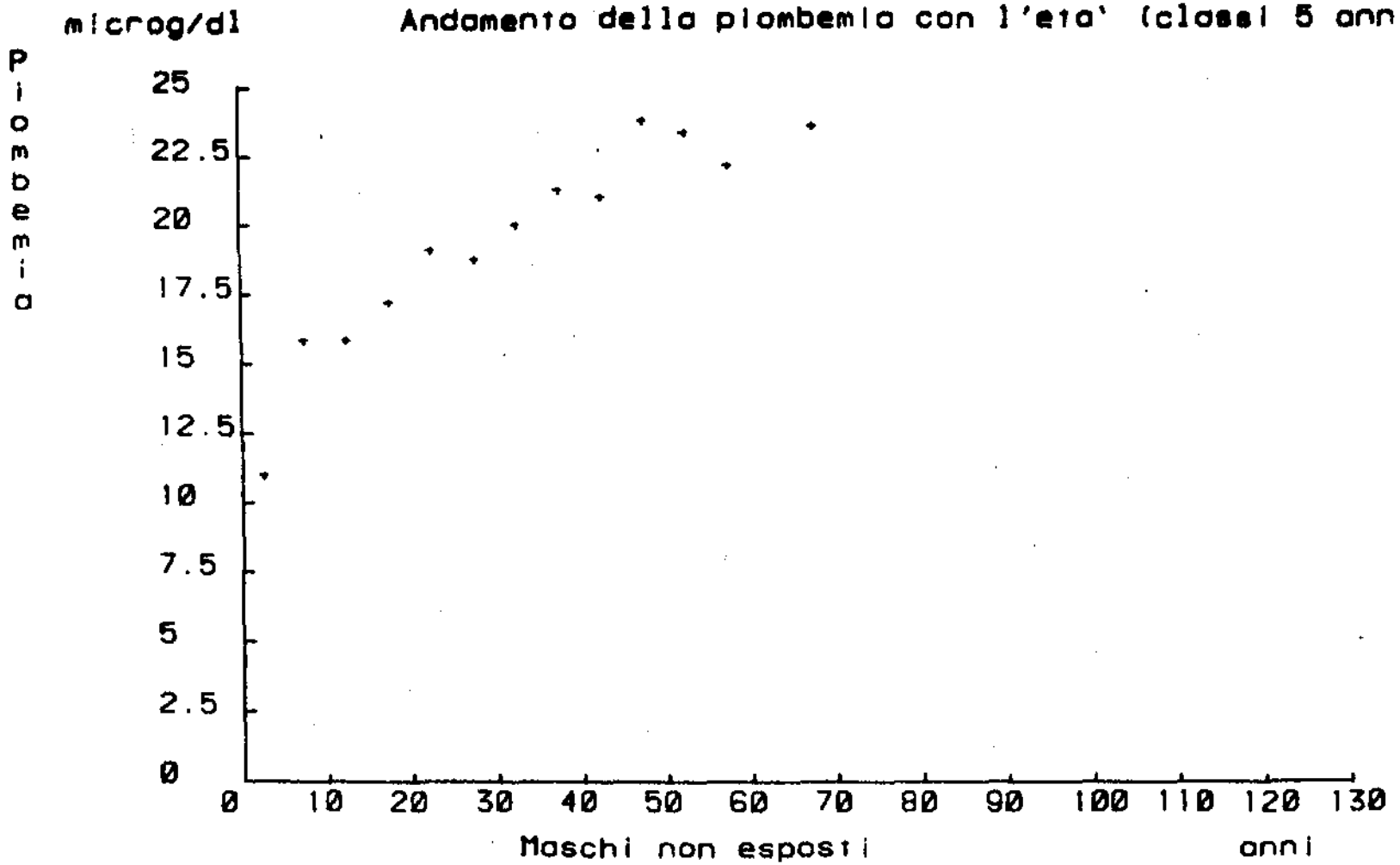
Immissione dati: tastiera (1) , nastro (2) 1

** Dopo l'ultima coppia di dati inserite [z]

```
Coppia no.   1   Y --->   11.03
                 X --->   2.5
Coppia no.   2   Y --->   15.87
                 X --->   7.5
Coppia no.   3   Y --->   15.87
                 X --->   12.5
Coppia no.   4   Y --->   17.22
                 X --->   17.5
Coppia no.   5   Y --->   19.05
                 X --->   22.5
Coppia no.   6   Y --->   18.73
                 X --->   27.5
Coppia no.   7   Y --->   19.92
                 X --->   32.5
Coppia no.   8   Y --->   21.19
                 X --->   37.5
Coppia no.   9   Y --->   20.95
                 X --->   42.5
Coppia no.   10   Y --->   23.73
                 X --->   47.5
Coppia no.   11   Y --->   23.3
                 X --->   52.5
Coppia no.   12   Y --->   22.06
                 X --->   57.5
Coppia no.   13   Y --->   23.49
                 X --->   67.5
Coppia no.   14   Y --->   z

Volete sostituire dati ? (si,no)    n

Volete registrare i dati ? (si,no)    n
```

```
Proseguiamo   ( 1=di seguito;  #1=page )     1

Andamento della piombemia con l'eta' (classi 5 anni)
Numero coppie= 13 Sx= 427.5 Sy= 252.41
Sx2= 18931.25 Sy2= 5064.7701
Sxy= 9126.875
Dxx= 4873.07692308 Dyy= 163.938707692
Dxy= 826.469230769
Vxx= 406.08974359 Vyy= 13.6615589744
Cxy= 68.8724358974
X Medio= 32.8846153846  Y Medio= 19.4161538462
Pendenza= 0.169599052881 (gradi 9.62 )
Sb= 0.0210581122758
Intercetta= 13.8389542226  Sa= 0.803595333411
Errore standard della stima  Se= 1.47001264431
Varianza Regr.= 140.168398774  Varianza Err. (Se)^2= 2.16093717443
Coeff.correl.lin. r= 0.924664751368  (r quadro= 0.855004902422 )
TEST D'IPOTESI
Test a=0:    t= 17.2212973958
Test b=0 (ovvero r=0)    t= 8.05385832592  F= 64.864633934
Test b=1     t= -39.4337790701
b forzato= 0.482106305711   sb forzato= 0.054089748937
se forzato= 7.44225610428
test b forzato=1     t= -9.5747106331

Grafico con i soli punti (1) o con le regressioni (2) 1

Grafico sul video (32) o sul plotter (1)  32

Volete indicare il campo di variazione della X ? (si,no)    n

Volete indicare il campo di variazione della Y ? (si,no)    n
```

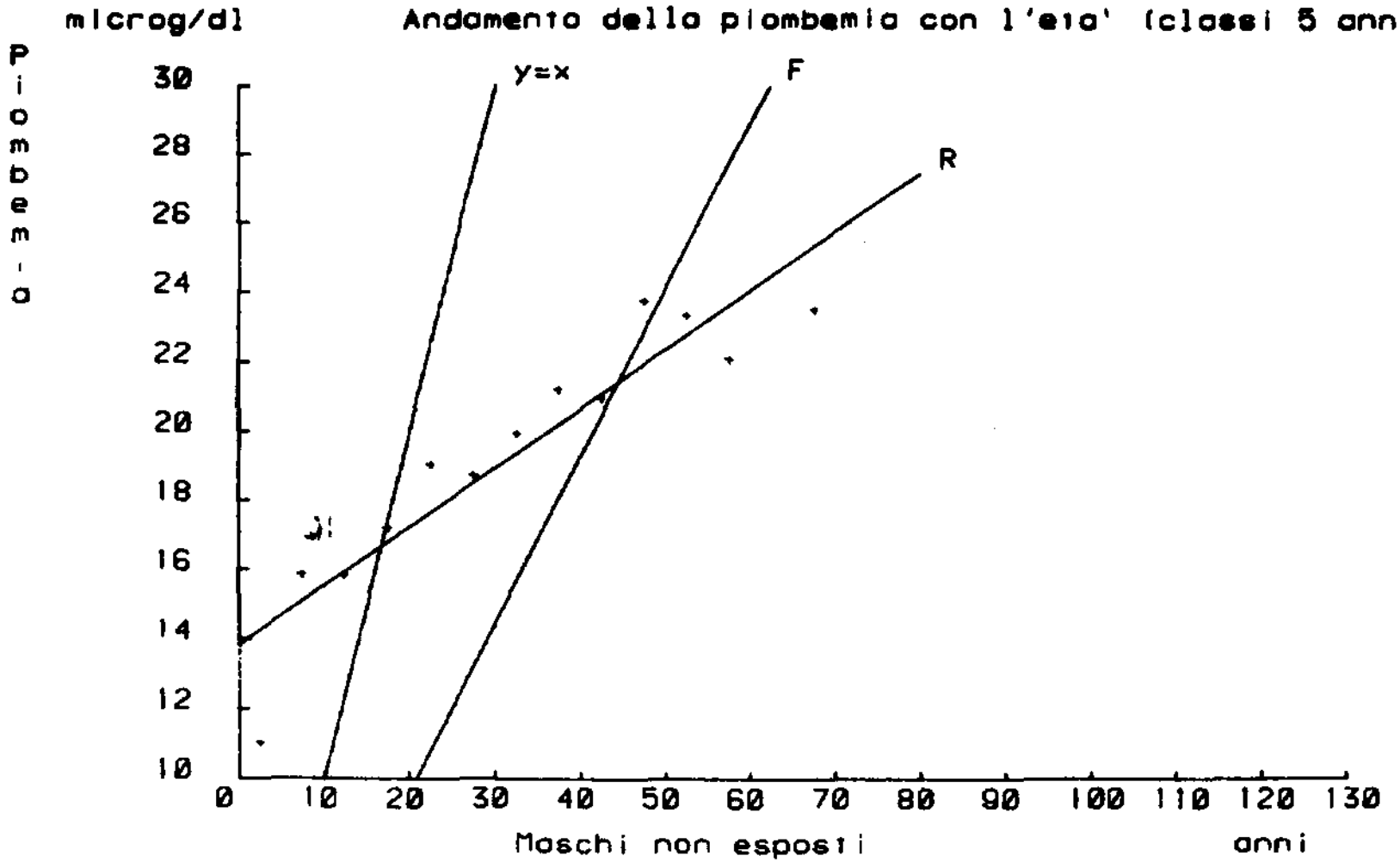
```
Volete indicare il passo della X ? (si,no)    n
Volete indicare il passo della Y ? (si,no)    n
Volete la numerazione degli intervalli sull'asse X ? (s,n) s
```



Volete ripetere il grafico ? (si,no) s

Grafico con i soli punti (1) o con le regressioni (2) 2

```
Grafico sul video (32) o sul plotter (1)  32
Volete modificare il campo di variazione della X (si,no) s
Volete indicare il campo di variazione della X ? (si,no)   s
Indicate in ordine l'estremo inferiore e quello superiore 0,70
Volete modificare il campo di variazione della Y (si,no) s
Volete indicare il campo di variazione della Y ? (si,no)   s
Indicate in ordine l'estremo inferiore e quello superiore 10,25
Volete indicare il passo della X ? (si,no)   n
Volete indicare il passo della Y ? (si,no)   n
Volete la numerazione degli intervalli sull'asse X ? (s,n) s
```



Volete ripetere il grafico ? (si,no) n

```
Volete la stima puntuale o l'intervallo di confidenza per la X, dato Y
(alfa=0.05)  ?   s

I valori di Y che inserite, sono valori esatti (1) o valori medi di de-
terminazioni (2) 1

valore di Y = 19.92

Stima puntuale:    Intervallo:        Estremo inf.,      Estremo sup.
 35.8554229763                        29.8646006387      41.846245314

Volete la stima puntuale o l'intervallo di confidenza per la X, dato Y
(alfa=0.05)  ?   n
```

12 - TEST SUI COEFFICIENTI DI REGRESSIONE b

ASPETTI TEORICI

Qualora si considerino due campioni (di numerosità N_1 e N_2) estratti da due popolazioni distinte, e si indichino con b_1 e b_2 i coefficienti di regressione determinati sui dati campionari, si può analizzare, in base ad essi, l'ipotesi di uguaglianza tra i coefficienti di regressione nelle due popolazioni provenienza.

A tale scopo si usa la statistica:

$$T = \frac{b_1 - b_2}{\sigma \sqrt{\frac{1}{\sum_{i=1}^{N_1} (x_{1i} - \bar{x}_1)^2} + \frac{1}{\sum_{i=1}^{N_2} (x_{2i} - \bar{x}_2)^2}}}$$

dove

$$\sigma = \sqrt{\frac{(N_1 - 2)\, s_1^2 + (N_2 - 2)\, s_2^2}{N_1 + N_2 - 4}}$$

x_{1i}, x_{2i} = intensità della variabile X nell'i-mo elemento del 1° e del 2° campione.

s_1^2, s_2^2 = varianze campionarie della stima del coefficiente di regressione.

$$x_1 = \sum_{i=1}^{N_1} x_{1i}/N_1$$

$$x_2 = \sum_{i=1}^{N_2} x_{2i}/N_2$$

La statistica T si distribuisce come una variabile casuale t di Student con $(N_1 + N_2 - 4)$ g.d.l..

Occorre ricordare che tale statistica è valida solo nel caso in cui si possa ipotizzare l'uguaglianza tra le varianze nelle due popolazioni analizzate (cioè $\sigma_1^2 = \sigma_2^2 = \sigma^2$).

MODALITA' D'USO

Inizialmente vi è una breve illustrazione delle caratteristiche principali del test, con l'indicazione dell'ipotesi nulla H_0 e dell'ipotesi alternativa H_1.

Con B1 si intende il valore del coefficiente di regressione nella popolazione da cui è stato estratto il 1° campione; con B2 si intende il valore del coefficiente di regressione nella popolazione da cui è stato estratto il 2° campione.

INPUT DATI

Le informazioni richieste sono:

1 - numero di coppie utilizzate per il calcolo del coefficiente di regressione nel campione (numerosità campionaria);

2 - valore del coefficiente di regressione osservato nel campione;

3 - devianza delle X ($D_{xx} = \sum (x_i - \bar{x})^2$);

4 - devianza delle Y ($D_{yy} = \sum (y_i - \bar{y})^2$);

5 - codevianza XY ($D_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$).

Queste informazioni vengono richieste separatemente per i due campioni.

L'ordine dei campioni è rilevante solo per il segno della statistica: i due valori che si ottengono inserendo le informazioni con inversione dei campioni, sono infatti simmetrici.

OUTPUT

Viene stampato il valore della statistica e la sua significatività (α), cioè il livello di rischio nel rifiutare H_0 quando è vera, sia nel caso del test unidirezionale (a una coda), sia bidirezionale (a due code).

Indicando con α^* il livello di rischio prefissato ritenuto accettabile:

se $\alpha^* < \alpha$ non si può rifiutare H_0;

se $\alpha^* > \alpha$ si rifiuta H_0.

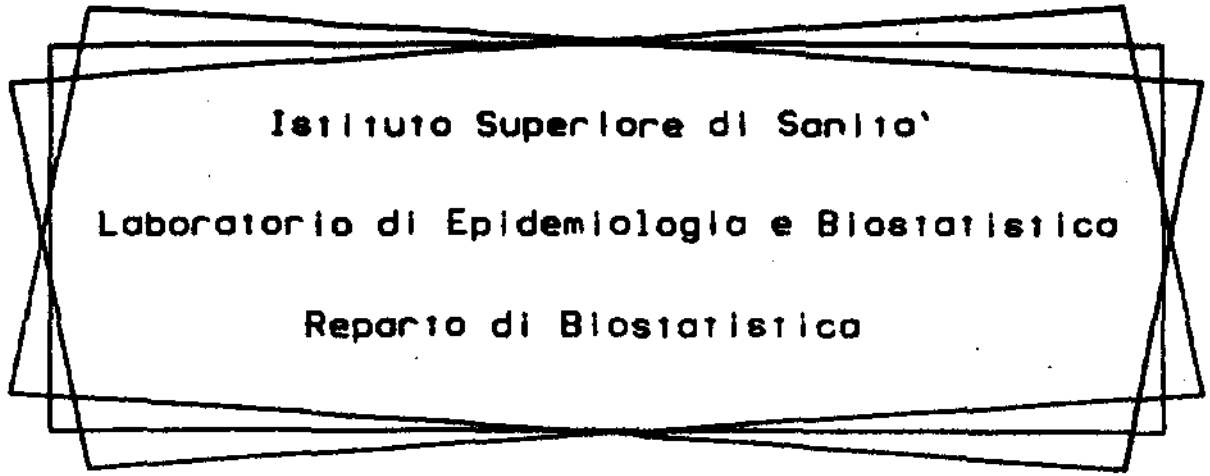
APPLICAZIONE

Illustriamo ora con un esempio l'utilizzazione del programma; i dati a cui

esso viene applicato, riportati nella tabella che segue, sono tratti da:
G. Morisi et al., Programma comunitario sulla sorveglianza biologica della popolazione contro il rischio di saturnismo, Risultati Italiani: Fase I (1978-'79), ISTISAN 1980/35.

Andamento della Piombemia con l'età - Maschi non esposti contro Femmine non esposte.

M.N.E.	F.N.E.
$n_1 = 13$	$n_2 = 15$
$b_1 = 0.17$	$b_2 = 0.08$
Dxx = 4873.08	Dxx = 7373.33
Dyy = 163.94	Dyy = 57.26
Dxy = 826.47	Dxy = 594.79



Guida alla Elaborazione Statistica
Nastro G.E.S. 7

*** Test sui coefficienti di regressione b ***

Il test viene condotto per verificare le ipotesi :
H0 : i coefficienti di regressione nelle 2 popolazioni di provenienza sono uguali (B1=B2);
H1 : i coefficienti di regressione nelle 2 popolazioni di provenienza sono diversi (B1<B2 ; B1>B2 ; B1≠B2).

Per continuare e cambiare pagina premere il tasto RETURN

CONFRONTO TRA DUE COEFFICIENTI DI REGRESSIONE (G.E.S. 7)

PRIMA REGRESSIONE

Numero di punti utilizzati nella regressione n1 = 13
Coefficiente di regressione : b1 = 0.17
Devianza delle X : Dxx = 4873.08
Devianza delle Y : Dyy = 163.94
Codevianza XY : Dxy = 826.47

SECONDA REGRESSIONE

Numero di punti utilizzati nella regressione : n2 = 15
Coefficiente di regressione : b2 = 0.08
Devianza delle X : Dxx = 7373.33
Devianza delle Y : Dyy = 57.26
Codevianza XY : Dxy = 594.79

****

osservato = 8.48711090536 con 24 g.d.l.

*** TEST A UNA CODA - H0: B1=B2 H1: B1<B2 oppure B1>B2

Il livello di rischio nel rifiutare H0 (quando e' vera) e' = 5.4565E-9

*** TEST A DUE CODE - H0: B1=B2 H1: B1≠B2

Il livello di rischio nel rifiutare H0 (quando e' vera) e' = 1.0913E-8

****

Volete ripetere il test ? (si,no)

BIBLIOGRAFIA

1 - G. A. Cinotti, G. Stirati, F. Taggi, R. Ronci, B.M. Simonetti & A. Pierucci, Relationship between kallikrein and PRA after intravenous furosemide, in J. Endocrinol. Invest. n. 2, p. 147, 1979.

2 - G. Morisi, F. Taggi, F. Martini, E. Magliola, G. Mattiello, A. Bortoli, L. Gelosa, E. Fortuna, L. Alessio, G. Vivoli, P. Borella, M. Bergomi, G. Pallotti, A. Consolino, G. Porrozzi, V. Piovano & O. Piombino, Programma comunitario sulla sorveglianza biologica della popolazione contro il rischio di saturnismo, Risultati Italiani: Fase I (1978-'79), ISTISAN 1980/35.

3 - G. Petrelli, G. Majori, M. Maggini, F. Taggi & M. Maroli, The head louse in Italy: an epidemiological study among schoolchildren, in The Royal Society of Health Journal, vol. 100, n. 2, p. 64, 1980.

4 - S. Siegel, Statistica non-parametrica per le scienze del comportamento, OS, Firenze, 1968.

RINGRAZIAMENTI

Gli autori desiderano ringraziare Antonio Giannitelli ed Aldo De Martino per l'assistenza prestata in varie fasi del lavoro negli aspetti connessi alla documentazione; ringraziano, inoltre, Lucilla Di Pasquale e Sara Modigliani per la collaborazione prestata durante la realizzazione del dattiloscritto su macchina IBM (Sistema di Gestione delle Informazioni 6).