

BIBLIOTECA DI SCIENZE STATISTICHE

SERVIZIO BIBLIOTECARIO NAZIONALE

BID PU 0804345 DID

ACQ. 185 / '03 INV. 83087

COLL. _____ CLASS. 5-66.WP.2/2003

**Nuovi sviluppi del
pacchetto software CR-Test**

R. Arboretti, D. Mazzaro, F.
Pesarin, L. Salmaso

2003.2

**Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova**

Febbraio 2003

UNIVERSITA' DI PADOVA
BIBLIOTECA DI SCIENZE STATISTICHE
Via C. Battisti, 241 - 35121 PADOVA

BIBLIOTECA DI SCIENZE STATISTICHE

SEZIONE DI STATISTICA

LIBRERIA

CLASSE 312.501.1

NUOVI SVILUPPI DEL
PACCHETTO SOFTWARE OK TEST

R. ARDENTI, D. MARCONI, F.
PESANTI, L. SERRAVALLO

1983

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 541-240
35121 Padova

Edizione 2003
UNIVERSITÀ DI PADOVA
BIBLIOTECA DI SCIENZE STATISTICHE
Via C. Battisti 541-240 Padova

INDICE

Premessa	pag.	2
Installazione e avvio del programma	“	2
Menù Principale	“	4
Sottomenù di Programma	“	5
Sottomenù Test per due campioni	“	6
- test t direzionale k-variato	“	6
- test di Fisher direzionale k-variato	“	8
- test t-paired non direzionale k-variato	“	10
- test t-paired direzionale k-variato	“	12
Sottomenù MANOVA	“	15
- ANOVA k-variata	“	15
- ANOVA k-variata e dati mancanti	“	18
- MANOVA per dati categoriali	“	20
- MANOVA per dati categoriali e dati mancanti	“	22
- ANOVA a due vie per misure ripetute	“	24
Osservazioni	“	26
Messaggi di errore	“	26

PREMESSA

Il programma **CR-Test** (**Conditional Resampling Test**) nasce come pacchetto per la gestione di routine per il calcolo dei test non parametrici di permutazione. La versione 1.0 per Visual Basic 4 (per sistemi Windows a 16 bit) è una versione preliminare del pacchetto CR-Test, che rende disponibili due gruppi di test di permutazione rispettivamente per due e per più campioni. La versione 1.1 di CR-test è costituita da un programma eseguibile per sistemi Windows a 32 bit. Per i riferimenti teorici delle procedure implementate, si veda "Pesarin F., *Resampling techniques for testing non-parametric hypotheses*, CLEUP, Padova, 1998" (available in 1999).

INSTALLAZIONE E AVVIO DEL PROGRAMMA

CR-TEST versione 1.0

Per l'installazione del programma **CR-TEST** versione 1.0 procedere nel seguente modo:

1. creare una directory **CR-TEST** nel disco fisso **C**;
2. copiare il file **CR-TEST.ZIP** nella directory **CR-TEST** del disco fisso **C**;
3. decomprimere il file **CR-TEST.ZIP** utilizzando PKUNZIP (presente nel dischetto di installazione) o altro decompressore analogo.

A fine installazione risulta creato nella directory **CR-TEST** l'eseguibile **CR-TEST.EXE** e altri file eseguibili .

N.B. Per procedere all'avvio del programma **CR-TEST** versione 1.0 è necessario avere installato nel drive **C** il pacchetto *Visual Basic 4*.

Per la creazione dell'icona **CR-TEST** in ambiente Windows 3.11, procedere nel seguente modo:

1. avviare Windows 3.11;
2. da **Program Manager** scegliere **File**, **Nuovo**, selezionare il bottone **Gruppo di Programmi** e confermare con **OK**;
3. impostare le proprietà del programma specificando in **Descrizione** il nome **CR-TEST** e confermare con **OK**;

4. tornare al **Program Manager**, selezionare nuovamente **File, Nuovo**, premere il bottone **Programma** e confermare con **OK**;

- nel campo **Descrizione** scrivere **CR-TEST**;
- nel campo **Riga di comando** scrivere **C:\CR-TEST\CR-TEST.EXE**;
- nel campo **Directory di Lavoro** scrivere **C:\CR-TEST** e confermare con **OK**.

Alla fine di queste operazioni l'icona di avvio automatico **CR-TEST** viene visualizzata nella finestra **CR-TEST**. Per avviare il programma fare doppio click sull'icona **CR-TEST**.

Per l'avvio del programma **CR-TEST** in ambiente Windows 95 o Windows NT, dalla finestra **Risorse del Computer** selezionare la cartella **CR-TEST** e fare doppio click sull'icona **CR-TEST**.

***N.B.** E' possibile effettuare in ambiente Windows 95 o NT la creazione dell'icona di avvio **CR-TEST** con una procedura analoga a quella sopra descritta in ambiente Windows 3.11.*

Una volta avviato il programma **CR-TEST** viene visualizzato il **Menù Principale**.

CR-TEST versione 1.1

Per l'installazione del programma **CR-TEST** versione 1.1 procedere nel seguente modo:

- selezionare il drive **A** e procedere alla decompressione del file compresso **CR-TEST.EXE** il cui contenuto viene automaticamente posizionato nella directory **C:\CR-TEST**.

A fine installazione risulta creato nella directory **CR-TEST** del disco fisso **C** l'eseguibile **CR-TEST.EXE**, assieme ad altri file di programma.

***N.B.** Per procedere all'avvio del programma **CR-TEST** versione 1.0 non è necessario avere installato nel drive **C** il pacchetto *Visual Basic*.*

Per l'avvio del programma **CR-TEST**, dalla finestra **Risorse del Computer** selezionare la cartella **C:\CR-TEST** e fare doppio click sull'icona **CR-TEST**.

***N.B.** E' possibile effettuare in ambiente Windows 95 o NT la creazione dell'icona di avvio **CR-TEST** con una procedura analoga a quella sopra descritta in ambiente Windows 3.11.*

Una volta avviato il programma **CR-TEST** viene visualizzato il **Menù Principale**.

MENÙ PRINCIPALE

Nel **Menù Principale** vengono visualizzati i due gruppi di test disponibili: **Test per due campioni** e **Manova** ai quali corrispondono i due sottomenù **Test per due campioni** e **Manova**. Per passare ad uno dei due sottomenù fare click sul pulsante corrispondente posto in basso alle caselle di testo relative ai due gruppi.

N.B. Per selezionare un pulsante di comando fare un solo click sul pulsante desiderato.

Per uscire dal programma selezionare **Chiudi** dal **Menù di Controllo**, oppure premere **ALT+F4**. E' possibile uscire dal programma anche da ciascuno dei **Sottomenù** selezionando **Chiudi** dal **Menù di Controllo**, oppure premendo **ALT+F4**.

SOTTOMENÙ DI PROGRAMMA

In ogni **Sottomenù**, sono posizionati in alto a sinistra vicino alle caselle di testo, i pulsanti **Esegui Test** per l'esecuzione dei corrispondenti test di permutazione.

In basso a sinistra sono collocati i seguenti pulsanti:

- **Visualizza Risultati:** premendo tale pulsante dopo la fine dell'esecuzione di un test, vengono visualizzati i risultati nel display posto nella parte destra del **Sottomenù**;
- **CLS - Archivia:** premendo tale pulsante viene ripulita la videata del display e vengono archiviati i risultati del test nel file **archivia.txt** creato automaticamente nella directory **CR-TEST**. Se vengono eseguiti più test, selezionando **CLS-Archivia** al termine dell'esecuzione di ogni singolo test, i risultati corrispondenti vengono copiati di volta in volta nel file **archivia.txt**.
N.B. E' possibile ripulire la videata del display facendo click direttamente sopra il display. In questo caso però, i dati non vengono archiviati nel file **archivia.txt**.
- **Stampa risultati:** premendo tale pulsante vengono messi in coda di stampa i risultati di ogni singolo test eseguito. All'uscita dal programma **CR-TEST** vengono stampati tutti i risultati dei test eseguiti. il programma **CR-TEST** stampa i risultati su stampante generica solo testo, su porta LPT1.
N.B. Non è possibile stampare i risultati senza averli prima visualizzati con il pulsante **Visualizza Risultati**;
- **Esci:** premendo tale pulsante si ritorna al **Menù Principale**.

SOTTOMENÙ TEST PER DUE CAMPIONI

Il sottomenù **Test per due campioni** comprende i seguenti test di permutazione:

TEST T DIREZIONALE K-VARIATO

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema del test t direzionale multivariato per l'ipotesi vettoriale $H_{0i} = \{M_{1i} = M_{2i}, i=1, \dots, k\}$ contro $H_{1i} = \{M_{1i} > M_{2i}, \text{ per almeno un } i=1, \dots, k\}$. Tale test è utilizzato per il confronto tra due campioni quando i dati sono espressi da valori numerici. Il test saggia l'ipotesi nulla di uguaglianza delle medie delle variabili tra i due gruppi contro l'ipotesi alternativa (ad una coda) che almeno una media sia più elevata in uno dei due gruppi posti a confronto. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di $N+2$ record, dove N è il numero di unità del campione; il primo record deve contenere una illustrazione del file di dati (si consiglia riportare la fonte dei dati e la descrizione delle variabili) e il secondo il numero $C=2$ di campioni e le numerosità N_j dei due campioni, $j=1,2$. Gli N record successivi devono essere ciascuno di q colonne (dove q è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati numerici relativi alle variabili osservate; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file può arrivare a $q \cdot N \leq 16000$.

Es. dati1.dat: **☛ Pallini-Pesarin 93, job satisfaction**

2 12 8

66

57

81

61

62

60

73

59

80

55

67

70

64

58

45

43

37

56

44

42

Nel file dati1.dat, il primo record contiene una breve descrizione del file, il secondo contiene il numero C=2 di campioni e le numerosità dei due campioni pari rispettivamente a 12 e 8 unità (individui). I record successivi contengono i dati della variabile 'grado di soddisfazione per il lavoro'.

Premendo il pulsante **Esegui Test** relativo al test **t direzionale k-variato**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati :

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza la descrizione del file, il numero di campioni, il numero di record e il numero di variabili contenute nel file:

*Nome file F\$ = c:\dati1.dat

V Dati: Pallini-Pesarin 93, job satisfaction

V N° campioni C=2 ; N° records N = 20 ; N° var./rec. q = 1

Il programma inoltre chiede il numero **k** di test di primo ordine, il numero **B** di replicazioni di simulazione condizionata e il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.), e se l'utente desidera o meno l'esecuzione del test combinato:

> N° test l° k = 1

> N° simulaz. B = 1000

> 1° variabile da testare : 1

> Vuoi la combinazione ? (S/N) : n

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù Test per due campioni**. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

V. Liv. Sign. Test 1 =

V. Liv. Sign. Test - Comb. =

Tempo s. =

TEST DI FISHER DIREZIONALE K-VARIATO

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema del test di Fisher direzionale k-variato per l'ipotesi vettoriale $H_{0i} = \{F_{1i} = F_{2i}, i=1, \dots, k\}$ contro $H_{1i} = \{F_{1i} > F_{2i}, \text{ per almeno un } i=1, \dots, k\}$. Tale test è utilizzato per il confronto tra due campioni rispetto a due o più variabili categoriche binarie a valori 0,1. Il test saggia l'ipotesi nulla di uguaglianza delle frequenze delle variabili tra i due gruppi contro l'ipotesi alternativa (ad una coda) che almeno una frequenza sia più elevata in uno dei due gruppi posti a confronto. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di $N+2$ record, dove N è il numero di unità del campione; il primo record deve contenere una illustrazione del file di dati (si consiglia riportare la fonte dei dati e la descrizione delle variabili) e il secondo il numero $C=2$ di campioni e le numerosità N_j dei due campioni, $j=1,2$. Gli N record successivi devono essere ciascuno di q colonne (dove q è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati binari (0,1) relativi alle variabili osservate; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file può arrivare a $q \cdot N \leq 16000$.

Es. dati2.dat:

```
Dati reazioni avverse 93, mal di testa dolori articolari (0=assenza, 1=presenza)
2 10 7
1 0
1 1
0 0
1 0
1 1
1 1
0 1
0 1
0 0
0 0
1 1
0 1
0 1
0 0
1 1
1 1
1 0
0 0
```

Nel file dati2.dat, il primo record contiene una breve descrizione del file, il secondo contiene il numero $C=2$ di campioni e le numerosità dei due campioni pari rispettivamente a 10 e 7 unità (individui). I record successivi contengono i dati delle variabili 'presenza di mal di testa (prima colonna; 0=no, 1=sì) e 'presenza di dolori articolari' (seconda colonna).

Premendo il pulsante **Esegui Test** relativo al test di **Fisher direzionale k-variato**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati :

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza la descrizione del file, il numero di campioni, il numero di record e il numero di variabili contenute nel file:

> Nome file F\$ = c:\dati2.dat

V Dati: Dati reazioni avverse 93, mal di testa dolori articolari (0=assenza, 1=presenza)

V N° campioni C=2 ; N° records N = 17 ; N° var.li/rec. q = 2

Il programma inoltre chiede il numero **k** di test di primo ordine, il numero **B** di replicazioni di simulazione condizionata e il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.), e se l'utente desidera o meno l'esecuzione del test combinato:

> N° test l° k = 2

> N° simulaz. B = 1000

> 1° variabile da testare : 1

> 2° variabile da testare : 2

> Vuoi la combinazione ? (S/N) : s

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù Test per due campioni**. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

V. Liv. Sign. Test 1 =

V. Liv. Sign. Test 2 =

V. Liv. Sign. Test - Comb. =

Tempo s. =

TEST T-PAIRED NON DIREZIONALE K-VARIATO

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema del test t-paired non direzionale multivariato per il sistema di ipotesi vettoriale $H_0=\{X=Y\}$ contro $H_1=\{X\neq Y\}$. Tale test è utilizzato per il confronto tra due campioni appaiati rispetto ad una o più variabili quando i dati sono espressi da valori numerici. Il programma esegue sia i test di primo ordine univariati per ogni coppia di variabili specificate, sia il test combinato relativo a tutte le coppie di variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di N record, dove N è il numero di unità del campione. Gli N record devono essere ciascuno di $2 \cdot q$ colonne (dove q è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati numerici relativi alle variabili osservate, per i quali abbia senso la differenza; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file deve soddisfare: $q \cdot N \leq 16000$.

Es. dati3.dat:

```
320 105.03 278 101.28
478  94.28 513  95.92
921 153.58 701 124.89
213 102.35 230 101.29
273 119.23 338 125.56
392 106.79 302  93.60
469  97.87 443  95.89
422 139.78 389 132.58
613  84.09 649  86.04
```

Il file dati3.dat contiene 9 record relativi a 9 unità sulle quali sono state osservate due variabili (testosterone e solitolo) in due istanti di tempo diversi (Landenna, Marasini, 1990). La prima e terza colonna contengono i dati della variabile 'testosterone' nei due istanti di tempo, la seconda e la quarta i dati della variabile 'solitolo' nei due istanti di tempo.

Premendo il pulsante **Esegui Test** relativo al test **t-paired non direzionale k-variato**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati:

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza il numero di record e il numero di variabili contenute nel file:

> Nome file F\$ = c:\dati3.dat

V N° records nel file n = 9 ; N° var.li/rec. q = 2

Il programma inoltre chiede il numero **k** di test di primo ordine, il numero **S** di replicazioni di simulazione condizionata e le coppie di variabili (numero d'ordine) relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.), e se l'utente desidera o meno l'esecuzione del test combinato:

* N° test 1° k = 2

* N° simulaz. condiz. S = 1000

> 1° coppia di variabili (x,y) : 1,3

> 2° coppia di variabili (x,y) : 2,4

> Vuoi la combinazione ? (S/N) : s

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù Test per due campioni**. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività e i valori dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

V. Liv. Sign. t-paired 1 =

t-paired 1 =

V. Liv. Sign. t-paired 2 =

t-paired 2 =

V. Liv. Sign. Test - Comb. =

Tempo s. =

TEST T-PAIRED DIREZIONALE K-VARIATO

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema del test t-paired direzionale multivariato per il sistema di ipotesi vettoriale $H_0=\{X=Y\}$ contro $H_1=\{X<Y\}$. Tale tipo di test è indicato per il confronto di due campioni appaiati rispetto ad una o più variabili quando i dati sono espressi da valori numerici, nelle situazioni in cui si vuole saggiare un sistema di ipotesi con alternative ristrette. Il programma esegue sia i test di primo ordine univariati per ogni coppia di variabili specificate, sia il test combinato relativo a tutte le coppie di variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di N+1 record, dove N è il numero di unità del campione; il primo record deve contenere una illustrazione del file di dati (si consiglia riportare la fonte dei dati e la descrizione delle variabili). Gli N record successivi devono essere ciascuno di 2*q colonne, dove q è il numero di variabili,

separate da uno spazio nelle quali vengono riportati i dati numerici relativi alle variabili osservate per i quali abbia senso la differenza; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file deve soddisfare: $q \cdot N \leq 16000$.

Es. dati4.dat:

Dati '97 , trigliceridi, colesterolo

185 230 190 240

177 267 183 254

169 195 177 276

180 243 180 231

176 210 184 212

173 223 187 245

Nel file dati4.dat, il primo record contiene una breve descrizione del file di dati. I record successivi relativi a 6 soggetti osservati in due istanti di tempo, contengono i dati delle variabili 'trigliceridi' (prima e terza colonna) e 'colesterolo' (seconda e quarta colonna). Si decide di adottare un'analisi di dominanza (ovvero saggiare un sistema di ipotesi con alternative ristrette) in quanto si è interessati a valutare se in uno dei due istanti di tempo i valori del colesterolo e della triglicerici risultino più bassi.

Premendo il pulsante **Esegui Test** relativo al test **t-paired direzionale k-variato**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' inoltre necessario specificare il path del file di dati:

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza la descrizione del file, il numero di record e il numero di variabili contenute nel file:

> Nome file F\$ = c:\dati4.dat

V File Dati: Dati '97 , trigliceridi, colesterolo

V N° records nel file n = 6 ; N° var.li/rec. q = 2

Il programma inoltre chiede il numero **k** di test di primo ordine, il numero **B** di replicazioni di simulazione condizionata e le coppie delle variabili (numero d'ordine) relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.), e se l'utente desidera o meno l'esecuzione del test combinato:

✚ N° test l° k = 2

✚ N° simulaz. condiz. B = 1000

> 1° coppia di variabili (x,y) : 1,3

> 2° coppia di variabili (x,y) : 2,4

> Vuoi la combinazione ? (S/N) : s

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù Test per due campioni**. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività e i valori dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

V. Liv. Sign. t-paired 1 =

t-paired 1 =

V. Liv. Sign. t-paired 2 =

t-paired 2 =

V. Liv. Sign. Test - Comb. =

Tempo s. =

SOTTOMENÙ MANOVA

Il sottomenù **MANOVA** comprende i seguenti test di permutazione:

ANOVA K-VARIATA

Questo programma risolve, via simulazione condizionata e combinazione non parametrica diretta, il problema dell'ANOVA multivariata per il sistema di ipotesi vettoriale $H_0 = \{M_{1i} = \dots = M_{ci}, i=1, \dots, k\}$ contro $H_1 = H_0^c$. Tale test è utilizzato per il confronto tra più campioni rispetto ad una o più variabili quando i dati sono espressi da valori numerici. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di $N+2$ record, dove N è il numero di unità del campione; il primo record deve contenere una illustrazione del file di dati (si consiglia riportare la fonte dei dati e la descrizione delle variabili). Il secondo record contiene il numero C di campioni e le numerosità N_j dei vari campioni, dove N_1 corrisponde al campione più numeroso. Gli N record successivi devono essere ciascuno di q colonne (dove q è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati numerici relativi alle variabili osservate; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file deve soddisfare: $q \cdot N \leq 16000$, $C \leq 50$.

Es. dati5.dat:

Dati '97, diametro cellulare

3 95 85 70

0.7

0.9

0.7

1.3

0.9

1.2

0.8

0.9

...

...

Nel file dati5.dat, il primo record contiene la descrizione del file, il secondo contiene il numero $C=3$ di campioni e le numerosità dei tre campioni pari rispettivamente a 95, 85 e 70 unità. I record successivi (non riportati tutti nell'esempio) contengono i dati della variabile 'diametro cellulare'.

N.B. Poichè si utilizza la procedura di combinazione non parametrica diretta, affinché il risultato del test combinato sia affidabile, è necessario che le numerosità campionarie siano sufficientemente elevate.

Premendo il pulsante **Esegui Test** relativo al test **ANOVA k-variata**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati:

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza la descrizione del file, il numero di campioni, il numero di record e il numero di variabili contenute nel file. Il programma prevede inoltre la possibilità di trasformare i dati; vengono difatti elencate alcune possibili trasformazioni indicate da un codice da 1 a 5 (il codice 1 non opera alcuna trasformazione sui dati):

> Nome file F\$ = c:\dati5.dat

V File dati: Dati '97, diametro cellulare

V N° campioni C=3 ; N° records N = 250 ; N° var.li/rec. q = 1

V Cod. Trasf. Dati 1 : $X(t)=X(t)$

V " " " 2 : $X(t)=X(t)-X(t-1)$

V " " " 3 : $X(t)=X(t)/X(t-1)$

V " " " 4 : $X(t)=X(t)-X(1)$

V " " " 5 : $X(t)=X(t)/X(1)$

Il programma chiede quindi il codice di trasformazione dei dati, il numero **k** di test di primo ordine, il numero **B** di replicazioni di simulazione condizionata e il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.); il test combinato viene eseguito automaticamente:

> Cod. Trasf. Dati = 1

> N° test I° k = 1

> N° ricamp. B = 1000

> 1° variabile da testare : 1

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù MANOVA**. Premendo il pulsante **Visualizza Risultati** vengono

visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività e i valori dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

Test No. 1 Liv. Sign. =

F Test No. 1 = ; g.l. = ()

Test-Comb Liv. Sign. = Tempo s. =

ANOVA K-VARIATA E DATI MANCANTI

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema dell'ANOVA multivariata per il sistema di ipotesi vettoriale $H_0 = \{M_{1i} = \dots = M_{ci}, i=1, \dots, k\}$ contro $H_1 = H_0$ e dati mancanti. Tale test è utilizzato per il confronto tra più campioni rispetto ad una o più variabili in presenza di dati mancanti quando i dati sono espressi da valori numerici. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di N+1 record, dove N è il numero di unità del campione; il primo record contiene il numero C di campioni e le numerosità N_j dei vari campioni, dove N_1 corrisponde al campione più numeroso. Gli N record successivi devono essere ciascuno di q colonne (dove q è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati numerici relativi alle variabili osservate; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. I dati mancanti possono essere indicati con uno dei caratteri seguenti: #, /, *, _, =. Il numero di dati mancanti deve essere tale da permettere un numero sufficiente di permutazioni (in generale almeno 5 permutazioni per ogni test di primo ordine). La dimensione del file deve soddisfare: $q \cdot N \leq 16000$, $C \leq 50$.

Es. dati6.dat:

Dati '98, diametro cellulare

3 8 5 5

0.7

#

0.7

1.3

0.9

1.2

0.8

0.9

0.7

/

1.4

1.4

0.9

*

0.9

1.6

1.4

1.4

Nel file dati6.dat, il primo record contiene il numero $C=3$ di campioni e le numerosità dei tre campioni pari rispettivamente a 8, 5 e 5 unità. I record successivi contengono i dati della variabile 'diametro cellulare'.

Premendo il pulsante **Esegui Test** relativo al test **ANOVA k-variata e dati mancanti**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati:

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza il numero di campioni, il numero di record e il numero di variabili contenute nel file:

> Nome file F\$ = c:\dati6.dat

V N° campioni C = 3 ; N° records N = 18 ; N° var.li/rec. q = 1

Il programma chiede il numero **k** di test di primo ordine, il numero **S** di replicazioni di simulazione condizionata, il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.) e se l'utente desidera o meno l'esecuzione del test combinato:

✓ N° test 1° k = 1

✓ N° ricamp. S = 1000

> 1° variabile da testare : 1

> Vuoi la combinazione ? (S/N) : n

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al Sottomenù MANOVA. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività e i valori dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

Var. No. 1

V. Liv. Sign. Test No. 1=

V. Liv. Sign. Test-comb. =

Tempo s. =

MANOVA PER DATI CATEGORIALI

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema della MANOVA con dati categoriali per il sistema di ipotesi vettoriale $H_0 = \{F_{1i} = F_{ci}, i=1, \dots, k\}$ contro $H_1 = H_0^c$. Tale test è utilizzato per il confronto tra le frequenze di una o più variabili categoriche tra due o più campioni. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di $N+1$ record, dove N è il numero di unità del campione; il primo record contiene il numero C di campioni e le numerosità N_j dei vari campioni, dove N_1 corrisponde al campione più numeroso. Gli N record successivi devono essere ciascuno di p colonne (dove p è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati espressi da numeri indice di classe relativi alle variabili osservate; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file deve soddisfare: $p \cdot N \leq 16000$, $C \leq 50$.

Es. dati7.dat:

```

2 8 4
1 1
1 1
2 1
2 2
1 1
3 1
3 2
2 1
2 2
1 1
1 2
1 1

```

Nel file dati7.dat, il primo record contiene il numero $C=2$ di campioni e le numerosità dei due campioni pari rispettivamente a 8 e 4 unità. I record successivi contengono i dati delle variabili 'gravità effetto collaterale A' (prima colonna) e 'gravità effetto collaterale B' (seconda colonna), dove 1=lieve, 2=moderato e 3=grave.

Premendo il pulsante **Esegui Test** relativo al test **MANOVA per dati categoriali**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati:

> **Nome file F\$ =**

Inserendo il nome del file, il programma visualizza il numero di campioni, il numero di record e il numero di variabili contenute nel file:

> **Nome file F\$ = c:\dati7.dat**

V N° campioni C = 2 ; N° records N = 12; N° var.li/rec. p = 2

Il programma chiede il numero **k** di test di primo ordine, il numero **S** di replicazioni di simulazione condizionata, il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.) e se l'utente desidera o meno l'esecuzione del test combinato:

> **N° test I° k = 2**

> **N° simulaz. S = 1000**

✕ **1° variabile da testare : 1**

✕ **2° variabile da testare : 2**

> **Vuol la combinazione ? (S/N) : s**

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù MANOVA**. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività e i valori dei **k** test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

V. Liv. Sign. Test 1=

Test Chi-square No. 1=

V. Liv. Sign. Test 2=

Test Chi-square No. 2=

V. Liv. Sign. Test-Comb. =

Tempo s. =

MANOVA PER DATI CATEGORIALI E DATI MANCANTI

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema della MANOVA con dati categoriali per il sistema di ipotesi vettoriale $H_0 = \{F_{1i} = F_{ci}, i=1, \dots, k\}$ contro $H_1 = H_{0c}$ e dati mancanti. Tale test è utilizzato per il confronto delle frequenze di una o più variabili categoriche tra due o più campioni in presenza di dati mancanti. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di $N+1$ record, dove N è il numero di unità del campione; il primo record contiene il numero C di campioni e le numerosità N_j dei vari campioni, dove N_1 corrisponde al campione più numeroso. Gli N record successivi devono essere ciascuno di p colonne (dove p è il numero di variabili) separate da uno spazio, nelle quali vengono riportati i dati espressi da numeri indice di classe relativi alle variabili osservate; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. I dati mancanti possono essere indicati con uno dei caratteri seguenti: #, /, *, _, =. Il numero di dati mancanti deve essere tale da permettere un numero sufficiente di permutazioni (in generale almeno 5 permutazioni per ogni test di primo ordine). La dimensione del file deve soddisfare: $p*N \leq 16000$, $C \leq 50$.

Es. dati8.dat:

```
2 9 5
1 1
1 1
#2
2 1
2 2
1 1
3 1
3 2
2 1
2 2
2 #
1 1
1 2
1 1
```

Nel file dati8.dat, il primo record contiene il numero $C=2$ di campioni e le numerosità dei due campioni pari rispettivamente a 9 e 5 unità. I record successivi contengono i dati delle variabili 'gravità effetto collaterale A' (prima colonna) e 'gravità effetto collaterale B' (seconda colonna), dove 1=lieve, 2=moderato e 3=grave.

Premendo il pulsante **Esegui Test** relativo al test **MANOVA** per dati categoriali e dati mancanti, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati:

> Nome file F\$ =

Inserendo il nome del file, il programma visualizza il numero di campioni, il numero di record e il numero di variabili contenute nel file:

> Nome file F\$ = c:\dati8.dat

V N° campioni C = 2 ; N° records N = 14 ; N° var.li/rec. p = 2

Il programma chiede il numero **k** di test di primo ordine, il numero **S** di replicazioni di simulazione condizionata, il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.) e se l'utente desidera o meno l'esecuzione del test combinato:

> N° test I° k = 2

> N° simulaz. S = 1000

> 1° variabile da testare : 1

> 2° variabile da testare : 2

> Vuoi la combinazione ? (S/N) : s

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al **Sottomenù MANOVA**. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività e i valori dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

V. Liv. Sign. Test 1=

Test Chi-square No. 1=

V. Liv. Sign. Test 2=

Test Chi-square No. 2=

V. Liv. Sign. Test-Comb. =

Tempo s. =

ANOVA A DUE VIE PER MISURE RIPETUTE

Questo programma risolve, via simulazione condizionata e combinazione non parametrica, il problema della ANOVA a due vie per misure ripetute per il sistema di ipotesi vettoriale $H_0 = \{M_{1i} = M_{ci}, i=1, \dots, q\}$ contro $H_1 = H_0^c$. Il programma esegue sia i test di primo ordine univariati per ogni singola variabile specificata, sia il test combinato relativo a tutte le variabili o ad un sottoinsieme di queste.

I dati vanno scritti in un file ASCII (o testo formattato delimitato da spazio) di $N+2$ record, dove N è il numero di unità del campione; il primo record deve contenere una illustrazione del file di dati (si consiglia riportare la fonte dei dati e la descrizione delle variabili). Il secondo record contiene nell'ordine: 3 (tre), il numero di unità N , il numero C di misure ripetute, e il numero q di variabili. Gli N record successivi devono essere ciascuno di $q \cdot c$ colonne (dove q è il numero di variabili e c il numero di misure ripetute) separate da uno spazio, nelle quali vengono riportati i dati espressi da valori numerici; i record vanno separati da RETURN. Non vi devono essere righe vuote alla fine del file di dati. La dimensione del file deve soddisfare: $q \cdot c \cdot N \leq 16000$.

Es. dati9.dat:

```
Dati 1990, testosterone e solitolo
3 6 3 2
236 98.42 222 105.56 232 94.16
415 103.77 359 99.61 292 84.05
645 119.47 526 139.51 458 122.57
261 114.50 253 101.03 199 92.30
323 133.99 332 136.38 222 103.07
289 94.14 305 89.24 172 66.80
```

Nel file dati9.dat, il primo record contiene il numero 3, il numero di unità $N=6$, il numero di misure ripetute $C=3$ e il numero di variabili $q=2$. I record successivi contengono i dati delle variabili 'testosterone' (prima, terza e quinta colonna) e 'solitolo' (seconda, quarta e sesta colonna) (Landenna, Marasini, 1990).

Premendo il pulsante **Esegui Test** relativo al test **ANOVA a due vie per misure ripetute**, compare una videata che riporta una descrizione del test e le modalità per la predisposizione del file di dati. Nella seconda videata il programma chiede il nome del file di dati, completo di estensione. E' necessario inoltre specificare il path del file di dati:

>> Nome file F\$ =

Inserendo il nome del file, il programma visualizza la descrizione del file, il numero di N di unità, il numero C di misure ripetute e il numero di variabili contenute nel file:

```
>> Nome file F$ = c:\dati9.dat
<> Dati 1990, testosterone e solitolo
<> # individui n = 6; # trattam. c = 3 ; # var. q = 2
```

Il programma chiede il numero k di test di primo ordine, il numero B di replicazioni di simulazione condizionata, il numero d'ordine delle variabili relative a ciascun test (1 per la variabile nella prima colonna, 2 per la variabile nella seconda colonna, ecc.) e se l'utente desidera o meno l'esecuzione del test combinato:

```
>> # Test l^tipo K<=q = 2
>> # ricampionam.. B = 1000

>> 1°  variabile da testare      # : 1
>> 2°  variabile da testare      # : 2
>> Combinazione test trattam. ? (S/N) : s
```

Al termine dell'inserimento dei valori dei parametri richiesti, il programma lancia l'esecuzione del test conservando l'ultima videata per il tempo necessario al completamento dell'esecuzione del test. Al termine dell'esecuzione del test, il programma ritorna al Sottomenù MANOVA. Premendo il pulsante **Visualizza Risultati** vengono visualizzati nel display i risultati dell'esecuzione del test, ovvero i livelli di significatività dei k test di primo ordine, il livello di significatività del test combinato e il tempo dell'esecuzione in secondi:

```
<> Test trattamenti # 1 , Liv. Sig. =
<> Test trattamenti # 2 , Liv. Sig. =
<> Test combinato trattam., Liv. Sig.=
<> Tempo impiegato sec. =
```

OSSERVAZIONI

1. In generale, il numero di replicazioni di simulazione condizionata può essere posto pari a 1000. Una buona regola di arresto consiste nel ripetere il test con diverse numerosità per le replicazioni di simulazione condizionata finchè i p-value dei test non si stabilizzano attorno ad un determinato valore.

Si noti che il numero di replicazioni di simulazione condizionata non è in alcun modo correlato con l'ampiezza dei campioni in esame.

2. Nel caso di confronti multipli è possibile ricorrere alle procedure di correzione (aggiustamento del p-value) come, ad esempio, la procedura di aggiustamento di Bonferroni o di Sidak.

MESSAGGI DI ERRORE

Durante l'esecuzione dei test di permutazione, possono comparire i seguenti messaggi di errore:

- 1. Incompatibilità nelle numerosità campionarie:**

questo messaggio viene visualizzato se esiste incompatibilità tra il valore dei parametri dichiarati e il numero e/o le numerosità effettive dei campioni;

- 2. Errore nell'esecuzione del test prescelto. Ripetere l'esecuzione controllando che il file di dati sia nel formato richiesto e che le impostazioni del test siano corrette.**

Modulo di Statistica Genetica

Questo software è stato aggiunto per poter risolvere specifici problemi di inferenza isotonica che trovano diverse applicazioni negli studi di associazione allelica del tipo caso-controllo in genetica. Il test opera su tabelle di associazione di dimensioni ridotte, pertanto i dati devono essere imputati manualmente. Tutto il processo è comunque molto semplice, poiché, ad ogni richiesta di inserimento dati, viene anche associata una frase di spiegazione. Infine, i risultati del test vengono visualizzati a video ed inoltre sono memorizzati nel file testo "results.dat" insieme alla descrizione dei dati e al numero di iterazioni Monte Carlo considerate.

Si mandi in esecuzione il programma MS-DOS eseguibile "**ModuleG.exe**". Con questo programma si possono richiamare 3 sotto-programmi:

- digitando il numero **1** e dando **invio**, viene mandata in esecuzione la routine **Rout1**;
- digitando il numero **2** e dando **invio**, viene mandata in esecuzione la routine **Rout2**;
- digitando il numero **3** e dando **invio**, viene mandata in esecuzione la routine **Rout3**;
- digitando il numero **4** e dando **invio**, il programma MS-DOS terminerà (se l'utente aveva lanciato il programma da ambiente Windows, è poi necessario che chiuda anche la relativa finestra);
- digitando qualsiasi altro valore e dando **invio**, il programma ripeterà le opzioni di scelta disponibili.

Il tempo di esecuzione dei sotto-programmi può variare parecchio in funzione del processore a disposizione e del numero di dati (in particolare il numero di iterazioni Monte Carlo).

ROUTine1, analisi di associazione allelica sul singolo locus

Il sotto-programma **Rout1** esegue una verifica d'ipotesi per l'analisi di associazione tra il genotipo di un locus (che potrà contenere 2 o 3 forme alleliche distinte) ed una specifica patologia, dove i dati sono raggruppati secondo un tipico studio caso-controllo:

Genotipo	Casi (malati)	Controlli (sani)
raro1/raro1	X1 (es. 3)	Y1 (es. 0)
raro1/comune	X2 (es. 6)	Y2 (es. 3)
comune/comun e	X3 (es. 33)	Y3 (es. 51)
raro2/raro2	X4 (es. 1)	Y4 (es. 1)
raro2/comune	X5 (es. 8)	Y5 (es. 15)
raro1/raro2	X6 (es. 1)	Y6 (es. 0)

Con questi dati, si costruiscono le stime del rischio relativo associato ai genotipi. Consideriamo un locus con due soli alleli, uno comune ed uno più raro (del resto la procedura si ripete allo stesso modo anche per un secondo allele raro). Si chiami con 'A' l'allele più raro e con 'a' quello comune: si possono presentare quindi tre diversi genotipi, l'omozigote raro 'AA', l'eterozigote 'Aa' e l'omozigote comune 'aa'. Le stime dei rischi relativi si ottengono come $R_{AA} = (\text{Numero di casi 'AA'})/(\text{Numero di controlli 'AA'})$, $R_{Aa} = (\text{Numero di casi 'Aa'})/(\text{Numero di controlli 'Aa'})$ e $R_{aa} = (\text{Numero di casi 'aa'})/(\text{Numero di controlli 'aa'})$. Per verificare l'ipotesi che l'allele raro sia di rischio o di protezione verso la patologia, è quindi necessario considerare gli Odds che sono ottenuti come rapporti tra i rischi: $\theta_{AA} = R_{AA}/R_{Aa}$, $\theta_{Aa} = R_{Aa}/R_{aa}$, $\theta_{aa} = R_{aa}/R_{aa} = 1$.

Come è noto a livello biologico, l'effetto dell'allele raro 'A' si può presentare in tre forme distinte: codominante ($\theta_{AA} \neq 1$, $\theta_{Aa} \neq 1$), dominante ($\theta_{AA} = 1$, $\theta_{Aa} \neq 1$) oppure recessivo ($\theta_{AA} \neq 1$, $\theta_{Aa} = 1$). Naturalmente, a seconda che l'allele raro 'A' risulti di rischio o protezione, si possono presentare soltanto i seguenti casi generali: $\theta_{AA} \geq \theta_{Aa} \geq 1$ (rischio) oppure $\theta_{AA} \leq \theta_{Aa} \leq 1$ (protezione).

Pertanto il sistema d'ipotesi che si vuole testare presenta la seguente alternativa isotonica:

$$H_0: \{ \theta_{AA} = \theta_{Aa} = 1 \} \text{ contro}$$

$$H_1: \{ (\theta_{AA} \geq \theta_{Aa} \geq 1) \text{ XOR } (\theta_{AA} \leq \theta_{Aa} \leq 1) \},$$

dove *XOR* significa “OR esclusivo” ed almeno una delle disuguaglianze deve valere in senso stretto.

A seconda che il locus abbia due o tre alleli, l'utente inserirà o meno degli zeri per i valori superflui. Tutta la procedura del test è molto semplice e viene descritta visivamente, inoltre viene specificato il sistema d'ipotesi che si sta testando.

Inserimento dei dati:

N_casi (numero totale di casi) > 52 invio;

N_controlli (numero totale di controlli) > 70 invio;

X1 > 3 invio;

X2 > 6 invio;

X3 > 33 invio;

X4 (inserire sempre zero se sono presenti solo 2 forme alleliche, 1 comune ed 1 rara) > 1 invio;

X5 (inserire sempre zero se sono presenti solo 2 forme alleliche, 1 comune ed 1 rara) > 8 invio;

X6 (inserire sempre zero se sono presenti solo 2 forme alleliche, 1 comune ed 1 rara) > 1 invio;

Y1 > 0 invio;

Y2 > 3 invio;

Y3 > 51 invio;

Y4 (inserire sempre zero se sono presenti solo 2 forme alleliche, 1 comune ed 1 rara) > 1 invio;

Y5 (inserire sempre zero se sono presenti solo 2 forme alleliche, 1 comune ed 1 rara) > 15 invio;

Y6 (inserire sempre zero se sono presenti solo 2 forme alleliche, 1 comune ed 1 rara) > 0 invio;

CMC-iterations (numero di ricampionamenti condizionati Monte Carlo) > 3000 invio.

Il test restituisce 2 *p*-value, ciascuno per la verifica d'ipotesi sul singolo allele raro, inoltre il test è ripetuto sotto le ipotesi che i dati del modello siano generati secondo l'equilibrio di Hardy-Weinberg.

Fissato uno specifico livello di significatività (generalmente 0.05), se il *p*-value risulta inferiore ad esso si accetterà l'ipotesi alternativa che la variante allelica rara sia di rischio ($\theta_{AA} \geq \theta_{Aa} \geq 1$) oppure di protezione ($\theta_{AA} \leq \theta_{Aa} \leq 1$), verso la patologia in questione.

Con i dati precedentemente inseriti, si ottengono il *p*-value 0.006 per l'allele raro 1 e 0.743 per l'allele raro 2. Ciò significa che il primo risulta altamente significativo per un effetto di rischio (infatti, se si calcolano gli odds, si ottiene $\theta_{AA} = \infty$ e $\theta_{Aa} = 3.0909$), mentre il secondo non risulta evidenziare alcun effetto rilevante. Al contrario, assumendo il modello sia generato secondo l'Hardy-Weinberg equilibrium, si ha rispettivamente 0.008 e 0.739. Il risultato dei due test in questo caso è praticamente lo stesso, tuttavia non sempre si è in questa situazione, in tal caso sta' al ricercatore decidere se i suoi dati rispettino il principio

di H.-W. o meno (generalmente si esegue prima un test specifico, che può essere anche un test chi-quadro per vedere se questo avviene).

Naturalmente ogni considerazione sui risultati va' letta anche in funzione della numerosità campionaria dei dati considerati.

Alla fine delle operazioni il programma MS-DOS termina automaticamente, tuttavia, qualora si abbia eseguito il programma da ambiente Windows, è sempre necessario chiudere la finestra. I risultati del test e tutti i dati inseriti sono salvati automaticamente nel file "results.dat".

Se l'utente imputa scorrettamente i dati, cioè, ad esempio, se la somma di tutti gli Xi risulta diversa dal totale dei casi, il programma fornirà un messaggio appropriato. Al contrario, se l'utente va' oltre la dimensione massima degli array supportata dal programma, verrà visualizzata la scritta "Error 7 at pgm-ctr: xxx" ed il programma terminerà. La dimensione massima di dati inseribile non è chiaramente definibile e dipende sia dal numero totale di casi e controlli che dalle iterazioni condizionate Monte Carlo.

ROUTine2, analisi di associazione su due loci biallelici

Il sotto-programma **Rout2** esegue una verifica d'ipotesi per l'analisi di associazione allelica congiunta su due loci biallelici (in ciascun locus si presentano due varianti alleliche: una comune e l'altra più rara). Supponiamo di avere i loci 1 e 2 con alleli (a = comune) e (A = raro) per 1 e (b = comune) e (B = raro) per 2. Potremmo, naturalmente, eseguire separatamente con i dati dei genotipi dei due loci l'analisi precedente, tuttavia, se si ritiene che tali loci possano avere anche un effetto d'interazione verso la patologia, sembra più opportuno analizzarli contemporaneamente. Quello a cui il ricercatore è interessato, potrebbe essere visto più semplicemente come un ordinamento tra gli effetti dei possibili genotipi verso la patologia.

Continuando il nostro esempio, i due loci possono dare origine separatamente a questi genotipi:

- locus 1, "AA = omozigote raro", "Aa = eterozigote" e "aa = omozigote comune";
- locus 2, "BB = omozigote raro", "Bb = eterozigote" e "bb = omozigote comune".

Pertanto, considerati congiuntamente, essi danno origine ai seguenti aplotipi:

Genotipi	BB	Bb	bb
AA	AABB	AABb	AAbb
Aa	AaBB	AaBb	Aabb
aa	aaBB	aaBb	aabb

In questo caso il ricercatore potrebbe ritenere interessante conoscere l'effetto dell'allele raro 'A' nel primo locus condizionatamente al genotipo presente nel locus 'B', in modo da capire se tale effetto venga incrementato, diminuito o risulti indifferente alla presenza dell'allele raro 'B'.

Per fare ciò, viene eseguito il test "Rout2" che restituisce 6 p -value, ciascuno dei quali rappresenta l'effetto dell'analisi di associazione in un locus, condizionatamente al genotipo presente nell'altro locus. Pertanto, nel caso precedente, si avranno i seguenti test:

- test sul locus 2 condizionatamente alla presenza del genotipo "aa" nel locus 1;
- test sul locus 2 condizionatamente alla presenza del genotipo "Aa" nel locus 1;
- test sul locus 2 condizionatamente alla presenza del genotipo "AA" nel locus 1;
- test sul locus 1 condizionatamente alla presenza del genotipo "bb" nel locus 2;
- test sul locus 1 condizionatamente alla presenza del genotipo "Bb" nel locus 2;

- test sul locus 1 condizionatamente alla presenza del genotipo "BB" nel locus 2.

Con i primi tre risultati è possibile stabilire in quale situazione risulti maggiormente significativo l'effetto (se vi è un qualche effetto!) portato dall'allele 'B' nel locus 2, mentre con gli ultimi tre è possibile stabilire quale sia il genotipo del locus 2 che rende più incisivo l'effetto dell'allele 'A' in 1.

Si consideri ora un data set reale ottenuto grazie alla collaborazione del laboratorio 'Angela Valenti' del Consorzio Mario Negri Sud. Nello studio dei fattori genetici associati all'infarto miocardico sono stati presi in esame, in particolare, due polimorfismi nella regione genetica del fattore VII: la regione 4 Ipervariabile (6 = allele comune e 7 = allele raro) ed il polimorfismo R353Q (R = allele comune e Q = allele raro). Pertanto, si è interessati a conoscere l'associazione tra le varianti alleliche rare e l'infarto al miocardio.

I dati vengono imputati in questo modo:

N_casi (numero totale di casi) > 155 invio;

N_controlli (numero totale di controlli) > 221 invio;

N_66RR_casi > 77 invio;

N_66RQ_casi > 6 invio;

N_66QQ_casi > 0 invio;

N_67RR_casi > 28 invio;

N_67RQ_casi > 32 invio;

N_67QQ_casi > 0 invio;

N_77RR_casi > 3 invio;

N_77RQ_casi > 8 invio;

N_77QQ_casi > 1 invio;

N_66RR_controlli > 78 invio;

N_66RQ_controlli > 15 invio;

N_66QQ_controlli > 1 invio;

N_67RR_controlli > 50 invio;

N_67RQ_controlli > 45 invio;

N_67QQ_controlli > 2 invio;

N_77RR_controlli > 7 invio;

N_77RQ_controlli > 16 invio;

N_77QQ_controlli > 7 invio;

CMC-iterations (numero di ricampionamenti condizionati Monte Carlo) > 10000 invio.

Consideriamo adesso i risultati forniti dal programma:

- L'analisi di associazione sul polimorfismo R353Q condizionatamente al genotipo '66' nella regione 4 ipervariabile del fattore VII ha fornito un p -value di 0.049. Ciò significa che abbiamo un risultato significativo al livello $\alpha = 0.05$ che l'allele Q nel polimorfismo R353Q presenta un effetto codominante protettivo (odds di 0 e 0.40) contro l'infarto al miocardio, se esso risulta associato alla forma '66' della regione ipervariabile 4.
- L'analisi di associazione sul polimorfismo R353Q condizionatamente al genotipo '67' nella regione 4 ipervariabile del fattore VII ha fornito un p -value di 0.877. Quindi, non c'è alcuna variazione di effetto, rispetto al verificarsi dell'infarto miocardico, dovuta alla presenza dell'allele raro 'Q' nel polimorfismo R353Q condizionatamente alla forma '67' nella regione ipervariabile 4.
- L'analisi di associazione sul polimorfismo R353Q condizionatamente al genotipo '77' nella regione 4 ipervariabile del fattore VII ha fornito un p -value di 0.383. Quindi, non c'è alcuna variazione di effetto, rispetto al verificarsi dell'infarto miocardico, dovuta alla presenza dell'allele raro 'Q' nel polimorfismo R353Q condizionatamente alla forma '77' nella regione ipervariabile 4.
- L'analisi di associazione allelica sulla regione 4 ipervariabile condizionatamente al genotipo 'RR' nel polimorfismo R353Q del fattore VII ha fornito un p -value di 0.043. Ciò significa che abbiamo un risultato significativo al livello $\alpha = 0.05$ che l'allele 7 nella regione 4 presenta un effetto protettivo codominante (odds di 0.76 e 0.56) contro l'infarto al miocardio, quando esso risulta associato alla forma 'RR' del polimorfismo R353Q.
- L'analisi di associazione allelica sulla regione 4 ipervariabile condizionatamente al genotipo 'RQ' nel polimorfismo R353Q del fattore VII ha fornito un p -value di 0.841. Ciò significa che la presenza dell'allele 7 nella regione 4 non presenta alcun effetto riguardo l'infarto al miocardio, quando è associata alla forma 'RQ' del polimorfismo R353Q.
- L'analisi di associazione allelica sulla regione 4 ipervariabile condizionatamente alla forma 'QQ' nel polimorfismo R353Q del fattore VII ha fornito un p -value di 0.734. Pertanto, non vi è nessuna significatività per un effetto dell'allele 7 nella regione 4 verso l'infarto al miocardio, qualora esso risulti associato alla forma 'QQ' del polimorfismo R353Q.

Il test è pertanto riuscito a determinare che entrambi le varianti rare '7' e 'Q' dei due polimorfismi presentano un effetto di protezione, il quale però si rivela significativo soltanto quando i due alleli non sono presenti contemporaneamente: in particolare, è risultata significativa la protezione dall'infarto al miocardio sviluppata dall'allele '7' della regione 4 ipervariabile, quando si considerino soltanto i soggetti con genotipo 'RR' nel polimorfismo R353Q, mentre, a sua volta, l'allele 'R' del polimorfismo R353Q si è rivelato di protezione quando è associato alla forma omozigote comune '66' nella regione 4 ipervariabile.

L'analisi viene anche ripetuta sotto l'assunzione che si utilizzi un modello di popolazione in Hardy-Weinberg equilibrium. Naturalmente ogni considerazione sui risultati va' letta anche in funzione della numerosità campionaria dei dati considerati.

Alla fine delle operazioni il programma MS-DOS termina automaticamente, tuttavia è sempre necessario chiudere la finestra. I risultati del test e tutti i dati inseriti sono salvati automaticamente nel file "results.dat".

Se l'utente imputa scorrettamente i dati, cioè, ad esempio, se la somma di tutti gli Xi risulta diversa dal totale dei casi, il programma fornirà un messaggio appropriato. Al contrario, se l'utente va' oltre la dimensione massima degli array supportata dal programma, verrà visualizzata la scritta "Error 7 at pgm-ctr: xxx" ed il programma terminerà. La dimensione massima di dati inseribile non è chiaramente definibile e dipende sia dal numero totale di casi e controlli che dalle iterazioni condizionate Monte Carlo.

ROUTINE 3, analisi di associazione stratificata su un locus biallelico

Il sotto-programma **Rout3** esegue una verifica d'ipotesi per l'analisi di associazione allelica stratificata su un locus biallelico.

Si consideri la seguente tabella di dati, dove 'a' è l'allele comune ed 'A' quello raro:

Genotipo	Fattore di confondimento				
	Esposizione (-)		Esposizione (+)		
	Caso	Controllo	Caso	Controllo	
aa	F1	F2	F3	F4	n_{aa}
Aa	F5	F6	F7	F8	n_{Aa}
AA	F9	F10	F11	F12	n_{AA}
	n_{CA-}	n_{CO-}	n_{CA+}	n_{CO+}	

Utilizzando i consueti odds, si ottengono $\theta_{aa} = (F3F2)/(F4F1)$, il valore che usiamo come riferimento, $\theta_{Aa} = (F7F6)/(F8F5)$ e $\theta_{AA} = (F11F10)/(F12F9)$. Tali odds non sono altro che stime puntuali del livello di associazione che è presente in qualsiasi genotipo tra "esposti" e "non esposti".

Innanzitutto, è possibile verificare se le singole proporzioni sono diverse da 1, cioè se $\theta_x \neq 1$, dove $x = aa, Aa, AA$. Sui primi odds calcolati, si procede adesso a calcolare i secondi, in modo da poter costruire ancora un sistema d'ipotesi con un'alternativa isotonica che ci permetta un'analisi più dettagliata dei dati:

$$K_{aa} = \theta_{aa}/\theta_{aa} = 1, K_{Aa} = \theta_{Aa}/\theta_{aa} \text{ e } K_{AA} = \theta_{AA}/\theta_{aa}$$

con cui si costruiscono le seguenti ipotesi

$$H_0: \{K_{AA} = K_{Aa} = 1\} \text{ contro}$$

$$H_1: \{(K_{AA} \geq K_{Aa} \geq 1) \text{ XOR } (K_{AA} \leq K_{Aa} \leq 1)\},$$

dove XOR significa "OR esclusivo" ed almeno una delle disuguaglianze deve valere in senso stretto. L'ipotesi alternativa è costruita in modo tale da corrispondere ai possibili effetti che la variabile di stratificazione può presentare congiuntamente all'allele putativo sui casi rispetto che sui controlli. Infatti, se un maggior livello di "esposizione" è tale da incrementare gli effetti dell'allele raro (o se presenta una qualche influenza su altri fattori correlati ad esso) il rischio dovrebbe comunque aumentare negli omozigoti 'AA' esposti al fattore rispetto agli eterozigoti 'Aa' esposti.

Pertanto, ad esempio, non sono ammessi punti nello spazio formato dall'insieme $\{K_{AA} > K_{Aa} < 1\}$ oppure $\{K_{AA} < K_{Aa} > 1\}$.

Si consideri il caso studio riassunto nella tabella seguente (dati forniti dal Consorzio Mario Negri Sud), in cui si analizza il polimorfismo R353Q (allele comune 'R' ed allele raro 'Q') del gene del fattore VII per un'eventuale associazione con l'infarto al miocardio, suddividendo i soggetti tra 'fumatori' e 'non fumatori':

	Fumatore = NO	Fumatore = NO	Fumatore = SI	Fumatore = SI
R353Q	Casi	Controlli	Casi	Controlli
RR	27	60	87	53
RQ	17	42	32	25
QQ	0	1	1	6
Totale	44	103	130	84

Questi dati sono imputati nella maniera seguente:

N_casi_con_strato_1 (NO) > 44 invio;

N_casi_con_strato_2 (SI) > 130 invio;

N_controlli_con_strato_1 (NO) > 103 invio;

N_controlli_con_strato_2 (SI) > 84 invio;

N_RR_casi_con_strato_1 (NO) > 27 invio;

N_RQ_casi_con_strato_1 (NO) > 17 invio;

N_QQ_casi_con_strato_1 (NO) > 0 invio;

N_RR_casi_con_strato_2 (SI) > 87 invio;

N_RQ_casi_con_strato_2 (SI) > 32 invio;

N_QQ_casi_con_strato_2 (SI) > 1 invio;

N_RR_controlli_con_strato_1 (NO) > 60 invio;

N_RQ_controlli_con_strato_1 (NO) > 42 invio;

N_QQ_controlli_con_strato_1 (NO) > 1 invio;

N_RR_controlli_con_strato_2 (SI) > 53 invio;

N_RQ_controlli_con_strato_2 (SI) > 25 invio;

N_QQ_controlli_con_strato_2 (SI) > 6 invio;

CMC-iterations (numero di ricampionamenti condizionati Monte Carlo) > 10000 invio.

Il p -value risultante è stato di 0.000, pertanto la variante allelica rara 'Q' del polimorfismo R353Q del fattore VII ha un effetto protettivo contro l'infarto al miocardio che diventa più accentuato nei fumatori rispetto ai non fumatori.

L'analisi viene anche ripetuta sotto l'assunzione che si utilizzi un modello di popolazione in Hardy-Weinberg equilibrium. Naturalmente ogni considerazione sui risultati va' letta anche in funzione della numerosità campionaria dei dati considerati.

Alla fine delle operazioni il programma MS-DOS termina automaticamente, tuttavia è sempre necessario chiudere la finestra. I risultati del test e tutti i dati inseriti sono salvati automaticamente nel file "results.dat".

Se l'utente imputa scorrettamente i dati, cioè, ad esempio, se la somma di tutti gli X_i risulta diversa dal totale dei casi, il programma fornirà un messaggio appropriato. Al contrario, se l'utente va' oltre la dimensione massima degli array supportata dal programma, verrà visualizzata la scritta "Error 7 at pgm-ctr: xxx" ed il programma terminerà. La dimensione massima di dati inseribile non è chiaramente definibile e dipende sia dal numero totale di casi e controlli che dalle iterazioni condizionate Monte Carlo.

Appendice (Ilistato dei programmi di statistica genetica)

ModuleG.bas

```
1 CLS:?  
  randomize timer  
  $stack &H1000  
  CLS  
  
2 print "This program can perform four different types of allelic "  
  print "association analysis (the data are always divided between "  
  print "cases and controls): "  
  print " 1 - the simple association test for 1 gene locus with 2 "  
  print "    or 3 alleles, where you would like to know if the "  
  print "    rarest alleles are (both or not) deleterious or "  
  print "    protective for the pathology; "  
  print " 2 - the association test for two gene loci, both with 2 "  
  print "    allele forms, where you would like to know what type "  
  print "    of effect there is with the interaction of two rarest "  
  print "    allele forms; "  
  print " 3 - the association test for 1 gene locus with 2 allele "  
  print "    forms stratified according to a confounding factor "  
  print "    with two levels (yes or not, high or low, etc.); "  
  print " 4 - to exit from the program. "  
3 input "Type the number of analysis you want performing to: ",TYPE  
  select case TYPE  
    case 1  
      CHAIN "rout1.tbc"  
    case 2  
      CHAIN "rout2.tbc"  
    case 3  
      CHAIN "rout3.tbc"  
    case 4  
      goto 5  
    case else  
      goto 2  
  end select  
  
5 END
```

Routine1.bas

```

10 CLS:?
randomize timer
$stack &H1000
CLS

print "Nonparametric Permutation Test with monotone restriction on the odds."
print "The test is univariate, but there could be 3 alleles, and the system "
print "of hypotheses is: "
print "  H0          ODD-RATIO(AA/Aa) = 1   and  ODD-RATIO(Aa/aa) = 1   "
print " "
print "          ODD-RATIO(AA/Aa) >= 1 and  ODD-RATIO(Aa/aa) >= 1   "
print "  H1          OR "
print "          ODD-RATIO(AA/Aa) <= 1 and  ODD-RATIO(Aa/aa) <= 1   "
print "The test is performed twice, the first one by comparing the odds "
print "obtained with two alleles, one of which is the commonest, the second "
print "time by comparing the odds obtained with the commonest allele and the "
print "other rare."

15 input "Type the total number of CASES:>",n1
input "Type the total number of CONTROLS:>",n2
input "Type the number of CASES with genotype (rare1) (rare1):>      ",X1%
input "Type the number of CASES with genotype (rare1) (common):>    ",X2%
input "Type the number of CASES with genotype (common) (common):>   ",X3%
input "Type the number of CASES with genotype (rare2) (rare2):>     ",X4%
input "Type the number of CASES with genotype (rare2) (common):>    ",X5%
input "Type the number of CASES with genotype (rare1) (rare2):>     ",X6%
input "Type the number of CONTROLS with genotype (rare1) (rare1):>  ",Y1%
input "Type the number of CONTROLS with genotype (rare1) (common):> ",Y2%
input "Type the number of CONTROLS with genotype (common) (common):> ",Y3%
input "Type the number of CONTROLS with genotype (rare2) (rare2):>  ",Y4%
input "Type the number of CONTROLS with genotype (rare2) (common):> ",Y5%
input "Type the number of CONTROLS with genotype (rare1) (rare2):>  ",Y6%
print "Type the number of Conditional Monte Carlo Iterations, the default"
input "is 3000:>",niter%
if (n1<>X1%+X2%+X3%+X4%+X5%+X6%)or(n2<>Y1%+Y2%+Y3%+Y4%+Y5%+Y6%) then
  print "Error in the inserted data. Repeat, please!":goto 15
end if

fla = 0

N=n1+n2
DIM CONN1!(niter%),PPV1!(niter%),PPV2!(niter%),CONN2!(niter%)
DIM INDEX%(n1),IX%(n2),INDICE%(N),TEST%(10),MATRIOSS%(N),CONN4!(niter%)
DIM MATRIX%(N),MAT1%(n1),MAT2%(n2),PAZ!(4),CONN3!(niter%)

20 SLOB%=0:cont1%=0:cont2%=0:LOB%=0:cont3%=0:cont4%=0

for j=1 to X1%
  MAT1%(j)=1
next
for j=X1%+1 to X1%+X2%
  MAT1%(j)=2
next
for j=X1%+X2%+1 to X1%+X2%+X3%
  MAT1%(j)=3
next
for j=X1%+X2%+X3%+1 to X1%+X2%+X3%+X4%
  MAT1%(j)=4
next
for j=X1%+X2%+X3%+X4%+1 to X1%+X2%+X3%+X4%+X5%
  MAT1%(j)=5
next
for j=X1%+X2%+X3%+X4%+X5%+1 to n1
  MAT1%(j)=6

```

```

next

for j=1 to Y1%
  MAT2%(j)=1
next
for j=Y1%+1 to Y1%+Y2%
  MAT2%(j)=2
next
for j=Y1%+Y2%+1 to Y1%+Y2%+Y3%
  MAT2%(j)=3
next
for j=Y1%+Y2%+Y3%+1 to Y1%+Y2%+Y3%+Y4%
  MAT2%(j)=4
next
for j=Y1%+Y2%+Y3%+Y4%+1 to Y1%+Y2%+Y3%+Y4%+Y5%
  MAT2%(j)=5
next
for j=Y1%+Y2%+Y3%+Y4%+Y5%+1 to n2
  MAT2%(j)=6
next

CALL GENET(INDEX%(),n1):CALL GENET(IX%(),n2)
CALL PERMET(INDEX%(),n1,1):CALL PERMET(IX%(),n2,1)
for j=1 to n1
  MATRIOSS%(j)=MAT1%(INDEX%(j))
next
for j=1 to n2
  MATRIOSS%(n1+j)=MAT2%(IX%(j))
next
MAX!=(X1%+Y1%)*(X2%+Y2%)
if ((X2%=0)or(Y1%=0)) then
  PAZ!(1)=MAX!
else
  PAZ!(1)=X1%*Y2%/(X2%*Y1%)
end if
MAX!=(X2%+Y2%)*(X3%+Y3%)
if ((X3%=0)or(Y2%=0)) then
  PAZ!(2)=MAX!
else
  PAZ!(2)=X2%*Y3%/(X3%*Y2%)
end if
MAX!=(X4%+Y4%)*(X5%+Y5%)
if ((X5%=0)or(Y4%=0)) then
  PAZ!(3)=MAX!
else
  PAZ!(3)=X4%*Y5%/(X5%*Y4%)
end if
MAX!=(X5%+Y5%)*(X3%+Y3%)
if ((X3%=0)or(Y5%=0)) then
  PAZ!(4)=MAX!
else
  PAZ!(4)=X5%*Y3%/(X3%*Y5%)
end if

CALL GENET(INDICE%(),N)
for i=1 to niter%
  CALL PERMET(INDICE%(),N,1)
  TEST%(1)=0:TEST%(2)=0:TEST%(3)=0:TEST%(7)=0:TEST%(8)=0
  TEST%(4)=0:TEST%(5)=0:TEST%(6)=0:TEST%(9)=0:TEST%(10)=0
  for j=1 to N
    MATRIX%(j)=MATRIOSS%(INDICE%(j))
  next
  for j=1 to n1
    if MATRIX%(j)=1 then TEST%(1)=TEST%(1)+1
    if MATRIX%(j)=2 then TEST%(2)=TEST%(2)+1
    if MATRIX%(j)=3 then TEST%(3)=TEST%(3)+1

```

```

    if MATRIX%(j)=4 then TEST%(7)=TEST%(7)+1
    if MATRIX%(j)=5 then TEST%(8)=TEST%(8)+1
next
for j=n1+1 to N
    if MATRIX%(j)=1 then TEST%(4)=TEST%(4)+1
    if MATRIX%(j)=2 then TEST%(5)=TEST%(5)+1
    if MATRIX%(j)=3 then TEST%(6)=TEST%(6)+1
    if MATRIX%(j)=4 then TEST%(9)=TEST%(9)+1
    if MATRIX%(j)=5 then TEST%(10)=TEST%(10)+1
next

MAX!=(TEST%(1)+TEST%(4))*(TEST%(5)+TEST%(2))
if ((TEST%(2)=0)or(TEST%(4)=0)) then
    CONN1!(i)=MAX!
else
    CONN1!(i)=TEST%(1)*TEST%(5)/(TEST%(2)*TEST%(4))
end if
if CONN1!(i)>=PAZ!(1) then cont1%=cont1%+1
MAX!=(TEST%(2)+TEST%(5))*(TEST%(6)+TEST%(3))
if ((TEST%(3)=0)or(TEST%(5)=0)) then
    CONN2!(i)=MAX!
else
    CONN2!(i)=TEST%(2)*TEST%(6)/(TEST%(3)*TEST%(5))
end if
if CONN2!(i)>=PAZ!(2) then cont2%=cont2%+1
MAX!=(TEST%(7)+TEST%(9))*(TEST%(8)+TEST%(10))
if ((TEST%(8)=0)or(TEST%(9)=0)) then
    CONN3!(i)=MAX!
else
    CONN3!(i)=TEST%(7)*TEST%(10)/(TEST%(8)*TEST%(9))
end if
if CONN3!(i)>=PAZ!(3) then cont3%=cont3%+1
MAX!=(TEST%(8)+TEST%(10))*(TEST%(7)+TEST%(9))
if ((TEST%(3)=0)or(TEST%(10)=0)) then
    CONN4!(i)=MAX!
else
    CONN4!(i)=TEST%(8)*TEST%(6)/(TEST%(3)*TEST%(10))
end if
if CONN4!(i)>=PAZ!(4) then cont4%=cont4%+1
next

PV1!=(0.5+cont1%)/(niter% + 1):PV2!=(0.5+cont2%)/(niter% + 1)
CALL INVNOR(1-PV1!,torino)
CALL INVNOR(1-PV2!,napoli)
genova=torino+napoli
for i=1 to niter%
    PPV1!(i)=0:PPV2!(i)=0
    for r=1 to niter%
        if CONN1!(r)>=CONN1!(i) then PPV1!(i)=PPV1!(i)+1
        if CONN2!(r)>=CONN2!(i) then PPV2!(i)=PPV2!(i)+1
    next
    PPV1!(i)=(0.5+PPV1!(i))/(niter% + 1):PPV2!(i)=(0.5+PPV2!(i))/(niter% + 1)
    CALL INVNOR(1-PPV1!(i),milano)
    CALL INVNOR(1-PPV2!(i),roma)
    parma=roma+milano
    if parma>=genova then SLOB%=SLOB%+1
next
PVALUE1=1-2*abs((SLOB%/niter%)-0.5)

PV3!=(0.5+cont3%)/(niter% + 1):PV4!=(0.5+cont4%)/(niter% + 1)
CALL INVNOR(1-PV3!,torino)
CALL INVNOR(1-PV4!,napoli)
genova=torino+napoli
for i=1 to niter%
    PPV1!(i)=0:PPV2!(i)=0
    for r=1 to niter%

```

```

    if CONN3!(r)>=CONN3!(i) then PPV1!(i)=PPV1!(i)+1
    if CONN4!(r)>=CONN4!(i) then PPV2!(i)=PPV2!(i)+1
next
PPV1!(i)=(0.5+PPV1!(i))/(niter% + 1):PPV2!(i)=(0.5+PPV2!(i))/(niter% + 1)
CALL INVNOR(1-PPV1!(i),milano)
CALL INVNOR(1-PPV2!(i),roma)
parma=roma+milano
if parma>=genova then LOB%=LOB%+1
next
PVALUE2=1-2*abs((LOB%/niter%)-0.5)

open "results.dat" for append as #1
print 1,tab(2);"*****"
print #1,tab(2);"Allelic Association Test on a single three-allelic locus. "
print #1,tab(2);DATE$
print #1,tab(2);"Nonparametric Permutation Test with monotone restriction on"
print #1,tab(2);"the odds."
print #1,tab(2);"The test is univariate, but there could be 3 alleles, and "
print #1,tab(2);"the system of hypotheses is: "
print #1,tab(2);" H0 ODD-RATIO(AA/Aa) = 1 and ODD-RATIO(Aa/aa) = 1 "
print #1,tab(2);" "
print #1,tab(2);" ODD-RATIO(AA/Aa) >= 1 and ODD-RATIO(Aa/aa) >= 1 "
print #1,tab(2);" H1 OR "
print #1,tab(2);" ODD-RATIO(AA/Aa) <= 1 and ODD-RATIO(Aa/aa) <= 1 "
print #1,tab(2);"The test is performed twice, the first one by comparing the"
print #1,tab(2);"odds obtained with two alleles, one of which is the"
print #1,tab(2);"commonest, the second time by comparing the odds obtained "
print #1,tab(2);"with the commonest allele and the other rare."
print #1,tab(2);"TOTAL CASES:>";tab(56);n1
print #1,tab(2);"TOTAL CONTROLS:>";tab(56);n2
print #1,tab(2);"CASES with genotype (rare1)(rare1):>";tab(56);X1%
print #1,tab(2);"CASES with genotype (rare1)(common):>";tab(56);X2%
print #1,tab(2);"CASES with genotype (common)(common):>";tab(56);X3%
print #1,tab(2);"CASES with genotype (rare2)(rare2):>";tab(56);X4%
print #1,tab(2);"CASES with genotype (rare2)(common):>";tab(56);X5%
print #1,tab(2);"CASES with genotype (rare1)(rare2), not interesting:>"
print #1,tab(56);X6%
print #1,tab(2);"CONTROLS with genotype (rare1)(rare1):>";tab(56);Y1%
print #1,tab(2);"CONTROLS with genotype (rare1)(common):>";tab(56);Y2%
print #1,tab(2);"CONTROLS with genotype (common)(common):>";tab(56);Y3%
print #1,tab(2);"CONTROLS with genotype (rare2)(rare2):>";tab(56);Y4%
print #1,tab(2);"CONTROLS with genotype (rare2)(common):>";tab(56);Y5%
print #1,tab(2);"CONTROLS with genotype (rare1)(rare2), not interesting:>"
print #1,tab(56);Y6%
print #1,tab(2);"Number of CMC-Iterations (the default is 3000):>",niter%
print #1,tab(2);"The p-value for the combined test with the rare allele 1 is"
print #1,tab(56);using "#.###";PVALUE1
print #1,tab(2);"The p-value for the combined test with the rare allele 2 is"
print #1,tab(56);using "#.###";PVALUE2
print #1,tab(2);"*****"
close #1
print tab(2);"The p-value for the combined test with the rare allele 1 is"
print #1,tab(56); using "#.###";PVALUE1
print tab(2);"The p-value for the combined test with the rare allele 2 is"
print #1,tab(56); using "#.###";PVALUE2

fla = fla+1

f1 = (2*Y1% + Y2% + Y6%)/(2*n2)
f2 = (2*Y4% + Y5% + Y6%)/(2*n2)
f3 = 1 - f1 - f2

Y1%=int(f1*f1*n2 + 0.5)
Y2%=int(2*f1*f3*n2 + 0.5)
Y3%=int(f3*f3*n2 + 0.5)

```

```
Y4%=int(f2*f2*n2 + 0.5)
Y5%=int(2*f2*f3*n2 + 0.5)
Y6%=int(2*f1*f2*n2 + 0.5)
```

```
n2hw=Y1%+Y2%+Y3%+Y4%+Y5%+Y6%
```

```
Y3%=n2-(Y1%+Y2%+Y4%+Y5%+Y6%)
```

```
if fla = 1 then
```

```
open "results.dat" for append as #1
```

```
print #1,tab(2);"Controls adjusted to Hardy-Weinberg Equilibrium"
```

```
print #1,tab(2);"TOTAL CONTROLS (approximated):>";tab(56);n2hw
```

```
print #1,tab(2);"TOTAL CONTROLS (real):>";tab(56);n2
```

```
close #1
```

```
print tab(2);"Controls adjusted to Hardy-Weinberg Equilibrium"
```

```
goto 20
```

```
end if
```

```
print tab(2);"The results are saved in the text file 'results.dat'"
```

```
END
```

Routine2.bas

```

10 CLS: ?
    randomize timer
    $stack &H1000
    CLS

print "Nonparametric Permutation Test with monotone restriction on the odds. "
print "The test is bivariate, with 2 loci both bi-allelic, and we test the "
print "effect of each locus conditionally to a specific genotype of the other "
print "locus. In this way, we can obtain an order and a characteristic of all "
print "the types of interaction effects between the loci. "
print "The genotype number 11,12, for example, means that we are considering "
print "the number of individuals (cases or controls, it does not matter) to "
print "which correspond the genotype 11 in first locus and the genotype 12 in "
print "second locus, where 2 is always the rarest allelic form."

15 input "TOTAL CASES:>",n1
    input "TOTALE CONTROLS:>",n2
    input "Number of CASES with genotypes 11,11:>",X1%
    input "Number of CASES with genotypes 11,12:>",X2%
    input "Number of CASES with genotypes 11,22:>",X3%
    input "Number of CASES with genotypes 12,11:>",X4%
    input "Number of CASES with genotypes 12,12:>",X5%
    input "Number of CASES with genotypes 12,22:>",X6%
    input "Number of CASES with genotypes 22,11:>",X7%
    input "Number of CASES with genotypes 22,12:>",X8%
    input "Number of CASES with genotypes 22,22:>",X9%
    input "Number of CONTROLS with genotypes 11,11:>",Y1%
    input "Number of CONTROLS with genotypes 11,12:>",Y2%
    input "Number of CONTROLS with genotypes 11,22:>",Y3%
    input "Number of CONTROLS with genotypes 12,11:>",Y4%
    input "Number of CONTROLS with genotypes 12,12:>",Y5%
    input "Number of CONTROLS with genotypes 12,22:>",Y6%
    input "Number of CONTROLS with genotypes 22,11:>",Y7%
    input "Number of CONTROLS with genotypes 22,12:>",Y8%
    input "Number of CONTROLS with genotypes 22,22:>",Y9%
    input "Number of Conditional Monte Carlo Iterations (3000 is the
default):>",niter%
    if (n1<>X1%+X2%+X3%+X4%+X5%+X6%+X7%+X8%+X9%) or
        (n2<>Y1%+Y2%+Y3%+Y4%+Y5%+Y6%+Y7%+Y8%+Y9%) then
        print "Error in the inserted data. Repeat, please!":goto 15
    end if

    fla = 0

    N=n1+n2

    DIM INDEX%(n1),IX%(n2),INDICE%(N),TEST%(18),MATRIOSS%(N),PVALUE(6)
    DIM MATRIX%(N),MAT1%(n1),MAT2%(n2),CONab1!(niter%),CONab2!(niter%)
    DIM CONaa1!(niter%),CONaa2!(niter%),CONbb1!(niter%),CONbb2!(niter%)
    DIM PPVaa1!(niter%),PPVaa2!(niter%),tor!(12),PV!(12),cont%(12),SLOB%(6)
    DIM PPVab1!(niter%),PPVab2!(niter%),PPVbb1!(niter%),PPVbb2!(niter%)
    DIM PPV11b!(niter%),PPV11a!(niter%),PPV22a!(niter%),PPV22b!(niter%)
    DIM PPV12a!(niter%),PPV12b!(niter%),CON11a!(niter%),CON11b!(niter%)
    DIM CON22a!(niter%),CON22b!(niter%),CON12a!(niter%),CON12b!(niter%)

20 for li%=1 to 6
    SLOB%(li%)=0: cont%(li%)=0: cont%(li%+6)=0
next

    for j=1 to X1%
        MAT1%(j)=1
    next
    for j=X1%+1 to X1%+X2%
        MAT1%(j)=2

```

```

next
for j=X1%+X2%+1 to X1%+X2%+X3%
  MAT1%(j)=3
next
for j=X1%+X2%+X3%+1 to X1%+X2%+X3%+X4%
  MAT1%(j)=4
next
for j=X1%+X2%+X3%+X4%+1 to X1%+X2%+X3%+X4%+X5%
  MAT1%(j)=5
next
for j=X1%+X2%+X3%+X4%+X5%+1 to X1%+X2%+X3%+X4%+X5%+X6%
  MAT1%(j)=6
next
for j=X1%+X2%+X3%+X4%+X5%+X6%+1 to X1%+X2%+X3%+X4%+X5%+X6%+X7%
  MAT1%(j)=7
next
for j=X1%+X2%+X3%+X4%+X5%+X6%+X7%+1 to X1%+X2%+X3%+X4%+X5%+X6%+X7%+X8%
  MAT1%(j)=8
next
for j=X1%+X2%+X3%+X4%+X5%+X6%+X7%+X8%+1 to n1
  MAT1%(j)=9
next

for j=1 to Y1%
  MAT2%(j)=1
next
for j=Y1%+1 to Y1%+Y2%
  MAT2%(j)=2
next
for j=Y1%+Y2%+1 to Y1%+Y2%+Y3%
  MAT2%(j)=3
next
for j=Y1%+Y2%+Y3%+1 to Y1%+Y2%+Y3%+Y4%
  MAT2%(j)=4
next
for j=Y1%+Y2%+Y3%+Y4%+1 to Y1%+Y2%+Y3%+Y4%+Y5%
  MAT2%(j)=5
next
for j=Y1%+Y2%+Y3%+Y4%+Y5%+1 to Y1%+Y2%+Y3%+Y4%+Y5%+Y6%
  MAT2%(j)=6
next
for j=Y1%+Y2%+Y3%+Y4%+Y5%+Y6%+1 to Y1%+Y2%+Y3%+Y4%+Y5%+Y6%+Y7%
  MAT2%(j)=7
next
for j=Y1%+Y2%+Y3%+Y4%+Y5%+Y6%+Y7%+1 to Y1%+Y2%+Y3%+Y4%+Y5%+Y6%+Y7%+Y8%
  MAT2%(j)=8
next
for j=Y1%+Y2%+Y3%+Y4%+Y5%+Y6%+Y7%+Y8%+1 to n2
  MAT2%(j)=9
next

CALL GENET(INDEX%(),n1):CALL GENET(IX%(),n2)
CALL PERMET(INDEX%(),n1,1):CALL PERMET(IX%(),n2,1)
for j=1 to n1
  MATRIOSS%(j)=MAT1%(INDEX%(j))
next
for j=1 to n2
  MATRIOSS%(n1+j)=MAT2%(IX%(j))
next

if ((X2%=0)or(Y1%=0)) then
  PARaa1!=(X1%+Y1%)*(Y2%+X2%)
else
  PARaa1!=X1%*Y2%/(X2%*Y1%)
end if
if ((X3%=0)or(Y2%=0)) then

```

```

    PARaa2!=(X2%+Y2%)*(Y3%+X3%)
else
    PARaa2!=X2%*Y3%/(X3%*Y2%)
end if
if ((X5%=0)or(Y4%=0)) then
    PARab1!=(X4%+Y4%)*(Y5%+X5%)
else
    PARab1!=X4%*Y5%/(X5%*Y4%)
end if
if ((X6%=0)or(Y5%=0)) then
    PARab2!=(X5%+Y5%)*(Y6%+X6%)
else
    PARab2!=X5%*Y6%/(X6%*Y5%)
end if
if ((X8%=0)or(Y7%=0)) then
    PARbb1!=(X7%+Y7%)*(Y8%+X8%)
else
    PARbb1!=X7%*Y8%/(X8%*Y7%)
end if
if ((X9%=0)or(Y8%=0)) then
    PARbb2!=(X8%+Y8%)*(Y9%+X9%)
else
    PARbb2!=X8%*Y9%/(X9%*Y8%)
end if
if ((X4%=0)or(Y1%=0)) then
    PAR11a!=(X1%+Y1%)*(Y4%+X4%)
else
    PAR11a!=X1%*Y4%/(X4%*Y1%)
end if
if ((X7%=0)or(Y4%=0)) then
    PAR11b!=(X4%+Y4%)*(Y7%+X7%)
else
    PAR11b!=X4%*Y7%/(X7%*Y4%)
end if
if ((X5%=0)or(Y2%=0)) then
    PAR12a!=(X2%+Y2%)*(Y5%+X5%)
else
    PAR12a!=X2%*Y5%/(X5%*Y2%)
end if
if ((X8%=0)or(Y5%=0)) then
    PAR12b!=(X5%+Y5%)*(Y8%+X8%)
else
    PAR12b!=X5%*Y8%/(X8%*Y5%)
end if
if ((X6%=0)or(Y3%=0)) then
    PAR22a!=(X3%+Y3%)*(Y6%+X6%)
else
    PAR22a!=X3%*Y6%/(X6%*Y3%)
end if
if ((X9%=0)or(Y6%=0)) then
    PAR22a!=(X6%+Y6%)*(Y9%+X9%)
else
    PAR22b!=X6%*Y9%/(X9%*Y6%)
end if

CALL GENET(INDICE%(),N)
for i=1 to niter%
    CALL PERMET(INDICE%(),N,1)
    for ir%=1 to 18
        TEST%(ir%)=0
    next

    for j=1 to N
        MATRIX%(j)=MATRIOSS%(INDICE%(j))
    next
    for j=1 to n1

```

```

for st% = 1 to 9
  if MATRIX%(j)=st% then TEST%(st%)=TEST%(st%)+1
next
next
for j=n1+1 to N
  for st%=1 to 9
    if MATRIX%(j)=st% then TEST%(st%+9)=TEST%(st%+9)+1
  next
next

if ((TEST%(2)=0)or(TEST%(10)=0)) then
  CONaa1!(i)=(TEST%(1)+TEST%(10))*(TEST%(2)+TEST%(11))
else
  CONaa1!(i)=TEST%(1)*TEST%(11)/(TEST%(2)*TEST%(10))
end if
if CONaa1!(i)>=PARaa1! then cont%(1)=cont%(1)+1

if ((TEST%(3)=0)or(TEST%(11)=0)) then
  CONaa2!(i)=(TEST%(2)+TEST%(11))*(TEST%(3)+TEST%(12))
else
  CONaa2!(i)=TEST%(2)*TEST%(12)/(TEST%(3)*TEST%(11))
end if
if CONaa2!(i)>=PARaa2! then cont%(2)=cont%(2)+1

if ((TEST%(4)=0)or(TEST%(13)=0)) then
  CONab1!(i)=(TEST%(4)+TEST%(13))*(TEST%(5)+TEST%(14))
else
  CONab1!(i)=TEST%(4)*TEST%(14)/(TEST%(5)*TEST%(13))
end if
if CONab1!(i)>=PARab1! then cont%(3)=cont%(3)+1

if ((TEST%(6)=0)or(TEST%(14)=0)) then
  CONab2!(i)=(TEST%(5)+TEST%(14))*(TEST%(6)+TEST%(15))
else
  CONab2!(i)=TEST%(5)*TEST%(15)/(TEST%(6)*TEST%(14))
end if
if CONab2!(i)>=PARab2! then cont%(4)=cont%(4)+1

if ((TEST%(8)=0)or(TEST%(16)=0)) then
  CONbb1!(i)=(TEST%(7)+TEST%(16))*(TEST%(8)+TEST%(17))
else
  CONbb1!(i)=TEST%(7)*TEST%(17)/(TEST%(8)*TEST%(16))
end if
if CONbb1!(i)>=PARbb1! then cont%(5)=cont%(5)+1

if ((TEST%(9)=0)or(TEST%(17)=0)) then
  CONbb2!(i)=(TEST%(8)+TEST%(17))*(TEST%(9)+TEST%(18))
else
  CONbb2!(i)=TEST%(8)*TEST%(18)/(TEST%(9)*TEST%(17))
end if
if CONbb2!(i)>=PARbb2! then cont%(6)=cont%(6)+1

if ((TEST%(4)=0)or(TEST%(10)=0)) then
  CON11a!(i)=(TEST%(1)+TEST%(10))*(TEST%(4)+TEST%(13))
else
  CON11a!(i)=TEST%(1)*TEST%(13)/(TEST%(4)*TEST%(10))
end if
if CON11a!(i)>=PAR11a! then cont%(7)=cont%(7)+1

if ((TEST%(7)=0)or(TEST%(13)=0)) then
  CON11b!(i)=(TEST%(4)+TEST%(13))*(TEST%(7)+TEST%(16))
else
  CON11b!(i)=TEST%(4)*TEST%(16)/(TEST%(7)*TEST%(13))
end if
if CON11b!(i)>=PAR11b! then cont%(8)=cont%(8)+1

```

```

if ((TEST%(1)=0) or (TEST%(11)=0)) then
  CON12a!(i)=(TEST%(2)+TEST%(11))*(TEST%(5)+TEST%(14))
else
  CON12a!(i)=TEST%(2)*TEST%(14)/(TEST%(5)*TEST%(11))
end if
if CON12a!(i)>=PAR12a! then cont%(9)=cont%(9)+1

if ((TEST%(8)=0) or (TEST%(14)=0)) then
  CON12b!(i)=(TEST%(1)+TEST%(10))*(TEST%(2)+TEST%(11))
else
  CON12b!(i)=TEST%(5)*TEST%(17)/(TEST%(8)*TEST%(14))
end if
if CON12b!(i)>=PAR12b! then cont%(10)=cont%(10)+1

if ((TEST%(6)=0) or (TEST%(12)=0)) then
  CON22a!(i)=(TEST%(1)+TEST%(10))*(TEST%(2)+TEST%(11))
else
  CON22a!(i)=TEST%(3)*TEST%(15)/(TEST%(6)*TEST%(12))
end if
if CON22a!(i)>=PAR22a! then cont%(11)=cont%(11)+1

if ((TEST%(9)=0) or (TEST%(15)=0)) then
  CON22b!(i)=(TEST%(1)+TEST%(10))*(TEST%(2)+TEST%(11))
else
  CON22b!(i)=TEST%(6)*TEST%(18)/(TEST%(9)*TEST%(15))
end if
if CON22b!(i)>=PAR22b! then cont%(12)=cont%(12)+1

next

for is%=1 to 12
  PV!(is%)=(0.5+cont%(is%))/(niter% + 1)
  CALL INVNOR(1-PV!(is%),tor!(is%))
next

genaa!=tor!(1)+tor!(2)
genab!=tor!(3)+tor!(4)
genbb!=tor!(5)+tor!(6)
gen11!=tor!(7)+tor!(8)
gen12!=tor!(9)+tor!(10)
gen22!=tor!(11)+tor!(12)

for i=1 to niter%
  PPVaa1!(i)=0:PPVaa2!(i)=0:PPVab1!(i)=0:PPVab2!(i)=0
  PPVbb1!(i)=0:PPVbb2!(i)=0:PPV11a!(i)=0:PPV11b!(i)=0
  PPV12a!(i)=0:PPV12b!(i)=0:PPV22a!(i)=0:PPV22b!(i)=0
  for r=1 to niter%
    if CONaa1!(r)>=CONaa1!(i) then PPVaa1!(i)=PPVaa1!(i)+1
    if CONaa2!(r)>=CONaa2!(i) then PPVaa2!(i)=PPVaa2!(i)+1
    if CONab1!(r)>=CONab1!(i) then PPVab1!(i)=PPVab1!(i)+1
    if CONab2!(r)>=CONab2!(i) then PPVab2!(i)=PPVab2!(i)+1
    if CONbb1!(r)>=CONbb1!(i) then PPVbb1!(i)=PPVbb1!(i)+1
    if CONbb2!(r)>=CONbb2!(i) then PPVbb2!(i)=PPVbb2!(i)+1
    if CON11a!(r)>=CON11a!(i) then PPV11a!(i)=PPV11a!(i)+1
    if CON11b!(r)>=CON11b!(i) then PPV11b!(i)=PPV11b!(i)+1
    if CON12a!(r)>=CON12a!(i) then PPV12a!(i)=PPV12a!(i)+1
    if CON12b!(r)>=CON12b!(i) then PPV12b!(i)=PPV12b!(i)+1
    if CON22a!(r)>=CON22a!(i) then PPV22a!(i)=PPV22a!(i)+1
    if CON22b!(r)>=CON22b!(i) then PPV22b!(i)=PPV22b!(i)+1
  next
  PPVaa1!(i)=(0.5+PPVaa1!(i))/(niter% + 1)
  PPVaa2!(i)=(0.5+PPVaa2!(i))/(niter% + 1)
  PPVab1!(i)=(0.5+PPVab1!(i))/(niter% + 1)
  PPVab2!(i)=(0.5+PPVab2!(i))/(niter% + 1)
  PPVbb1!(i)=(0.5+PPVbb1!(i))/(niter% + 1)

```

```

PPVbb2!(i)=(0.5+PPVbb2!(i))/(niter% + 1)
PPV11a!(i)=(0.5+PPV11a!(i))/(niter% + 1)
PPV11b!(i)=(0.5+PPV11b!(i))/(niter% + 1)
PPV12a!(i)=(0.5+PPV12a!(i))/(niter% + 1)
PPV12b!(i)=(0.5+PPV12b!(i))/(niter% + 1)
PPV22a!(i)=(0.5+PPV22a!(i))/(niter% + 1)
PPV22b!(i)=(0.5+PPV22b!(i))/(niter% + 1)

CALL INVNOR(1-PPVaa1!(i),mil1): CALL INVNOR(1-PPVaa2!(i),mil2)
parm1=mil1+mil2
if genaa!>=parm1 then SLOB%(1)=SLOB%(1)+1
CALL INVNOR(1-PPVab1!(i),mil3): CALL INVNOR(1-PPVab2!(i),mil4)
parm2=mil3+mil4
if genab!>=parm2 then SLOB%(2)=SLOB%(2)+1
CALL INVNOR(1-PPVbb1!(i),mil5): CALL INVNOR(1-PPVbb2!(i),mil6)
parm3=mil5+mil6
if genbb!>=parm3 then SLOB%(3)=SLOB%(3)+1
CALL INVNOR(1-PPV11a!(i),mil7): CALL INVNOR(1-PPV11b!(i),mil8)
parm4=mil7+mil8
if gen11!>=parm4 then SLOB%(4)=SLOB%(4)+1
CALL INVNOR(1-PPV12a!(i),mil9): CALL INVNOR(1-PPV12b!(i),mil10)
parm5=mil9+mil10
if gen12!>=parm5 then SLOB%(5)=SLOB%(5)+1
CALL INVNOR(1-PPV22a!(i),mil11): CALL INVNOR(1-PPV22b!(i),mil12)
parm6=mil11+mil12
if gen22!>=parm6 then SLOB%(6)=SLOB%(6)+1

next

for ki%=1 to 6
  PVALUE(ki%)=1-2*abs((SLOB%(ki%)/niter%)-0.5)
next

open "results.dat" for append as #1
print #1,tab(2);"*****"
print #1,tab(2);"Allelic Association Test on two bi-allelic loci.
Nonparametric Permutation Test with monotone restriction on the odds. The
test is bivariate, with 2 loci both bi-allelic, and we test the effect of
each locus conditionally to a specific genotype of the other locus. In
this way, we can obtain an order and a characteristic of all the types of
interaction effects between the loci. The genotype number 11,12, for
example, means that we are considering the number of individuals (cases
or controls, it does not matter) to which correspond the genotype 11 in
first locus and the genotype 12 in second locus, where 2 is always the
rarest allelic form.
"
print #1,tab(2);DATE$
print #1,tab(2);"TOTAL CASES:>";tab(56);n1
print #1,tab(2);"TOTAL CONTROLS:>";tab(56);n2
print #1,tab(2);"Number of CASES with genotypes 11,11:>";tab(56);X1%
print #1,tab(2);"Number of CASES with genotypes 11,12:>";tab(56);X2%
print #1,tab(2);"Number of CASES with genotypes 11,22:>";tab(56);X3%
print #1,tab(2);"Number of CASES with genotypes 12,11:>";tab(56);X4%
print #1,tab(2);"Number of CASES with genotypes 12,12:>";tab(56);X5%
print #1,tab(2);"Number of CASES with genotypes 12,22:>";tab(56);X6%
print #1,tab(2);"Number of CASES with genotypes 22,11:>";tab(56);X7%
print #1,tab(2);"Number of CASES with genotypes 22,12:>";tab(56);X8%
print #1,tab(2);"Number of CASES with genotypes 22,22:>";tab(56);X9%
print #1,tab(2);"Number of CONTROLS with genotypes 11,11:>";tab(56);Y1%
print #1,tab(2);"Number of CONTROLS with genotypes 11,12:>";tab(56);Y2%
print #1,tab(2);"Number of CONTROLS with genotypes 11,22:>";tab(56);Y3%
print #1,tab(2);"Number of CONTROLS with genotypes 12,11:>";tab(56);Y4%
print #1,tab(2);"Number of CONTROLS with genotypes 12,12:>";tab(56);Y5%
print #1,tab(2);"Number of CONTROLS with genotypes 12,22:>";tab(56);Y6%
print #1,tab(2);"Number of CONTROLS with genotypes 22,11:>";tab(56);Y7%
print #1,tab(2);"Number of CONTROLS with genotypes 22,12:>";tab(56);Y8%

```

```

print #1,tab(1);"Number of CONTROLS with genotypes 22,22:>";tab(56);Y9%
print #1,tab(2);"Number of CMC-Iterations (3000 is the default):>",niter%
print #1,tab(1);"The p-value for the combined test, conditional to genotype
      11 in first locus, is:";tab(56); using "#.###";PVALUE(1)
print #1,tab(2);"The p-value for the combined test, conditional to genotype
      12 in first locus, is:";tab(56); using "#.###";PVALUE(2)
print #1,tab(2);"The p-value for the combined test, conditional to genotype
      22 in first locus, is:";tab(56); using "#.###";PVALUE(3)
print #1,tab(2);"The p-value for the combined test, conditional to genotype
      11 in second locus, is:";tab(56); using "#.###";PVALUE(4)
print #1,tab(2);"The p-value for the combined test, conditional to genotype
      12 in second locus, is:";tab(56); using "#.###";PVALUE(5)
print #1,tab(2);"The p-value for the combined test, conditional to genotype
      22 in second locus, is:";tab(56); using "#.###";PVALUE(6)
print 1 ,tab(2);"*****"
close #1

```

```
fla = fla+1
```

```

f1A = (2*(Y1%+Y2%+Y3%)+Y4%+Y5%+Y6%)/(2*n2)
f2A = 1 - f1A
f1B = (2*(Y1%+Y4%+Y7%)+Y2%+Y5%+Y8%)/(2*n2)
f2B = 1 - f1B

```

```

Y2%=int(2*f1A*f1A*f1B*f2B*n2 + 0.5)
Y3%=int(f1A*f1A*f2B*f2B*n2 + 0.5)
Y4%=int(2*f1A*f2A*f1B*f1B*n2 + 0.5)
Y5%=int(4*f1A*f2A*f1B*f2B*n2 + 0.5)
Y6%=int(2*f1A*f2A*f2B*f2B*n2 + 0.5)
Y7%=int(f2A*f2A*f1B*f1B*n2 + 0.5)
Y8%=int(2*f2A*f2A*f1B*f2B*n2 + 0.5)
Y9%=int(f2A*f2A*f2B*f2B*n2 + 0.5)
Y1%=n2-(Y9%+Y2%+Y3%+Y4%+Y5%+Y6%+Y7%+Y8%)

```

```
if fla = 1 then
```

```
open "results.dat" for append as #1
```

```
print #1,tab(2);"Controls adjusted to Hardy-Weinberg Equilibrium"
```

```
print tab(2);"Controls adjusted to Hardy-Weinberg Equilibrium"
```

```
close #1
```

```
goto 20
```

```
end if
```

```
print tab(2);"The results are saved in the text file 'results.dat'"
```

```
END
```

Routine3.bas

```

10 CLS:?
   randomize timer
   $stack &H1000
   CLS

   print "Nonparametric Permutation Test with monotone restriction on the odds.
         The test is stratified, there is 1 bi-allelic locus with common allele
         'a' and rare allele 'A', and we would like to test if there are some
         significant differences between the effect of the rare allele in
         presence of different levels (1 and 2)of the confounding factor. "

15 input "Insert the number of cases with factor = 1:>",n11
   input "Insert the number of cases with factor = 2:>",n12
   input "Insert the number of controls with factor = 1:>",n21
   input "Insert the number of controls with factor = 2:>",n22
   input "Insert the number of cases with genotype 'aa' and factor = 1:>",X1%
   input "Insert the number of cases with genotype 'Aa' and factor = 1:>",X2%
   input "Insert the number of cases with genotype 'AA' and factor = 1:>",X3%
   input "Insert the number of cases with genotype 'aa' and factor = 2:>",X4%
   input "Insert the number of cases with genotype 'Aa' and factor = 2:>",X5%
   input "Insert the number of cases with genotype 'AA' and factor = 2:>",X6%
   input "Insert the number of controls with genotype 'aa' and factor = 1:>",Y1%
   input "Insert the number of controls with genotype 'Aa' and factor = 1:>",Y2%
   input "Insert the number of controls with genotype 'AA' and factor = 1:>",Y3%
   input "Insert the number of controls with genotype 'aa' and factor = 2:>",Y4%
   input "Insert the number of controls with genotype 'Aa' and factor = 2:>",Y5%
   input "Insert the number of controls with genotype 'AA' and factor = 2:>",Y6%
   input "Number of Conditional Monte Carlo Iterations (3000 is the
default):>",niter%
   if
(n11<>X1%+X2%+X3%)or(n12<>X4%+X5%+X6%)or(n21<>Y1%+Y2%+Y3%)or(n22<>Y4%+Y5%+Y6%)
then
   print "Error in the inserted data. Repeat, please!":goto 15
end if

   fla = 0

   N1=n11+n21: N2=n12+n22: N=N1+N2

   DIM PPV1!(niter%),PPV2!(niter%),MAT12%(n12),MAT22%(n22),IX2%(n22),ODD1!(3)
   DIM INDEX1%(n11),IX1%(n21),INDICE%(N),TEST1%(6),MATRIOS%(N)
   DIM MATRIX%(3,2,2),MAT11%(n11),MAT21%(n21),TEST2%(6),INDEX2%(n12)
   DIM MATRIX1%(N),ODab!(niter%),ODbb!(niter%)

   MATRIX%(1,1,1)=X1%: MATRIX%(1,1,2)=Y1%
   MATRIX%(1,2,1)=X4%: MATRIX%(1,2,2)=Y4%
   MATRIX%(2,1,1)=X2%: MATRIX%(2,1,2)=Y2%
   MATRIX%(2,2,1)=X5%: MATRIX%(2,2,2)=Y5%
   MATRIX%(3,1,1)=X3%: MATRIX%(3,1,2)=Y3%
   MATRIX%(3,2,1)=X6%: MATRIX%(3,2,2)=Y6%

20 SLOB%=0:cont1%=0:cont2%=0

   for j=1 to X1%
     MAT11%(j)=1
   next
   for j=X1%+1 to X1%+X2%
     MAT11%(j)=2
   next
   for j=X1%+X2%+1 to n11
     MAT11%(j)=3
   next
   for j=1 to Y1%

```

```

MAT21%(j)=1
next
for j=Y1%+1 to Y1%+Y2%
  MAT21%(j)=2
next

for j=Y1%+Y2%+1 to n21
  MAT21%(j)=3
next
for j=1 to X4%
  MAT12%(j)=1
next
for j=X4%+1 to X4%+X5%
  MAT12%(j)=2
next
for j=X4%+X5%+1 to n12
  MAT12%(j)=3
next
for j=1 to Y4%
  MAT22%(j)=1
next
for j=Y4%+1 to Y4%+Y5%
  MAT22%(j)=2
next
for j=Y4%+Y5%+1 to n22
  MAT22%(j)=3
next

CALL GENET(INDEX1%(),n11):CALL GENET(IX1%(),n21)
CALL PERMET(INDEX1%(),n11,1):CALL PERMET(IX1%(),n21,1)
CALL GENET(INDEX2%(),n12):CALL GENET(IX2%(),n22)
CALL PERMET(INDEX2%(),n12,1):CALL PERMET(IX2%(),n22,1)

for j=1 to n11
  MATRIOS%(j)=MAT11%(INDEX1%(j))
next
for j=1 to n21
  MATRIOS%(n11+j)=MAT21%(IX1%(j))
next
for j=1 to n12
  MATRIOS%(N1+j)=MAT12%(INDEX2%(j))
next
for j=1 to n22
  MATRIOS%(N1+n12+j)=MAT22%(IX2%(j))
next

for j=1 to 3
  if ((MATRIX%(j,1,1)=0) or (MATRIX%(j,2,2)=0)) then
    ODD1!(j)=9999999999
  else
    ODD1!(j)=MATRIX%(j,1,2)*MATRIX%(j,2,1)/(MATRIX%(j,1,1)*MATRIX%(j,2,2))
  end if
next
if (ODD1!(2)=0) then
  ODD2ab!=9999999999
else
  ODD2ab!=ODD1!(1)/ODD1!(2)
end if
if (ODD1!(3)=0) then
  ODD2bb!=9999999999
else
  ODD2bb!=ODD1!(1)/ODD1!(3)
end if

CALL GENET(INDICE%(),N)

```

```

for i=1 to niter%
  CALL PERMET(INDICE%,N,1)
  TEST1%(1)=0:TEST1%(2)=0:TEST1%(3)=0
  TEST1%(4)=0:TEST1%(5)=0:TEST1%(6)=0
  TEST2%(1)=0:TEST2%(2)=0:TEST2%(3)=0
  TEST2%(4)=0:TEST2%(5)=0:TEST2%(6)=0
  for j=1 to N
    MATRIX1%(j)=MATRIOS%(INDICE%(j))
  next
  for j=1 to n11
    if MATRIX1%(j)=1 then TEST1%(1)=TEST1%(1)+1
    if MATRIX1%(j)=2 then TEST1%(2)=TEST1%(2)+1
    if MATRIX1%(j)=3 then TEST1%(3)=TEST1%(3)+1
  next
  for j=n11+1 to N1
    if MATRIX1%(j)=1 then TEST1%(4)=TEST1%(4)+1
    if MATRIX1%(j)=2 then TEST1%(5)=TEST1%(5)+1
    if MATRIX1%(j)=3 then TEST1%(6)=TEST1%(6)+1
  next
  for j=N1+1 to N1+n12
    if MATRIX1%(j)=1 then TEST2%(1)=TEST2%(1)+1
    if MATRIX1%(j)=2 then TEST2%(2)=TEST2%(2)+1
    if MATRIX1%(j)=3 then TEST2%(3)=TEST2%(3)+1
  next
  for j=N1+n12+1 to N
    if MATRIX1%(j)=1 then TEST2%(4)=TEST2%(4)+1
    if MATRIX1%(j)=2 then TEST2%(5)=TEST2%(5)+1
    if MATRIX1%(j)=3 then TEST2%(6)=TEST2%(6)+1
  next
  if ((TEST1%(1)=0)or(TEST2%(4)=0)) then
    OD1!=9999999999
  else
    OD1!=TEST1%(4)*TEST2%(1)/(TEST1%(1)*TEST2%(4))
  end if
  if ((TEST1%(2)=0)or(TEST2%(5)=0)) then
    OD2!=9999999999
  else
    OD2!=TEST1%(5)*TEST2%(2)/(TEST1%(2)*TEST2%(5))
  end if
  if ((TEST1%(3)=0)or(TEST2%(6)=0)) then
    OD3!=9999999999
  else
    OD3!=TEST1%(6)*TEST2%(3)/(TEST1%(3)*TEST2%(6))
  end if
  if (OD2!=0) then
    ODab!(i)=9999999999
  else
    ODab!(i)=OD1!/OD2!
  end if
  if (OD3!=0) then
    ODbb!(i)=9999999999
  else
    ODbb!(i)=OD1!/OD3!
  end if

  if ODab%(i)>=ODD2ab% then cont1%=cont1%+1
  if ODbb%(i)>=ODD2bb% then cont2%=cont2%+1
next
PV1!=(0.5+cont1%)/(niter% + 1):PV2!=(0.5+cont2%)/(niter% + 1)
CALL INVNOR(1-PV1!,torino): CALL INVNOR(1-PV2!,napoli)
genova=torino+napoli
for i=1 to niter%
  PPV1!(i)=0:PPV2!(i)=0
  for r=1 to niter%
    if ODab%(r)>=ODab%(i) then PPV1!(i)=PPV1!(i)+1
    if ODbb%(r)>=ODbb%(i) then PPV2!(i)=PPV2!(i)+1
  next
next

```

```

next
PPV1!(i)=(0.5+PPV1!(i))/(niter% + 1):PPV2!(i)=(0.5+PPV2!(i))/(niter% + 1)
CALL INVNOR(1-PPV1!(i),milano)
CALL INVNOR(1-PPV2!(i),roma)
parma=roma+milano
if parma>=genova then SLOB%=SLOB%+1
next
PVALUE=1-2*abs((SLOB%/niter%)-0.5)

open "results.dat" for append as #1
print #1,tab(2);"*****"
print #1,tab(2);"Allelic Association Test for the interaction with a
confounder. Nonparametric Permutation Test with monotone
restriction on the odds. The test is stratified, there is 1
bi-allelic locus with common allele 'a' and rare allele 'A',
and we would like to test if there are some significant
differences between the effect of the rare allele in
presence of different levels (1 and 2) of the confounding
factor. "
print #1,tab(2);DATE$
print #1,tab(2);"TOTAL CASES with factor = 1:>";tab(56);n11
print #1,tab(2);"TOTAL CASES with factor = 2:>";tab(56);n12
print #1,tab(2);"TOTAL CONTROLS with factor = 1:>";tab(56);n21
print #1,tab(2);"TOTAL CONTROLS with factor = 2:>";tab(56);n22
print #1,tab(2);"Number of cases with genotype 'aa' and factor = 1:>"
print #1,tab(56);X1%
print #1,tab(2);"Number of cases with genotype 'Aa' and factor = 1:>"
print #1,tab(56);X2%
print #1,tab(2);"Number of cases with genotype 'AA' and factor = 1:>"
print #1,tab(56);X3%
print #1,tab(2);"Number of cases with genotype 'aa' and factor = 2:>"
print #1,tab(56);X4%
print #1,tab(2);"Number of cases with genotype 'Aa' and factor = 2:>"
print #1,tab(56);X5%
print #1,tab(2);"Number of cases with genotype 'AA' and factor = 2:>"
print #1,tab(56);X6%
print #1,tab(2);"Number of controls with genotype 'aa' and factor = 1:>"
print #1,tab(56);Y1%
print #1,tab(2);"Number of controls with genotype 'Aa' and factor = 1:>"
print #1,tab(56);Y2%
print #1,tab(2);"Number of controls with genotype 'AA' and factor = 1:>"
print #1,tab(56);Y3%
print #1,tab(2);"Number of controls with genotype 'aa' and factor = 2:>"
print #1,tab(56);Y4%
print #1,tab(2);"Number of controls with genotype 'Aa' and factor = 2:>"
print #1,tab(56);Y5%
print #1,tab(2);"Number of controls with genotype 'AA' and factor = 2:>"
print #1,tab(56);Y6%
print #1,tab(2);"Number of CMC-Iterations (3000 is the default):>",niter%
print #1,tab(2);"The p-value of the combined test is:";tab(56);
using "#.###";PVALUE
print #1,tab(2);"*****"
close #1
print tab(2);"The p-value of the combined test is:";tab(56);
using "#.###";PVALUE

fla = fla+1

f1 = (2*Y3% + Y2%)/(2*n21)
f2 = 1 - f1

Y1%=int(f2*f2*n21 + 0.5)
Y2%=int(2*f1*f2*n21 + 0.5)
Y3%=int(f1*f1*n21 + 0.5)

Y1%=n21-(Y2%+Y3%)

```

```
f1 = (2*Y6% + Y5%)/(2*n22)
f2 = 1 - f1

Y4%=int(f2*f2*n22 + 0.5)
Y5%=int(2*f1*f2*n22 + 0.5)
Y6%=int(f1*f1*n22 + 0.5)

Y4%=n22-(Y5%+Y6%)

if fla = 1 then
open "results.dat" for append as #1
print #1,tab(2);"Controls adjusted to Hardy-Weinberg Equilibrium"
print tab(2);"Controls adjusted to Hardy-Weinberg Equilibrium"
close #1
goto 20
end if
print tab(2);"The results are saved in the text file 'results.dat'"
```

END

Subroutine utilizzate

```
SUB PERMET(H%(1),nv,nw) 'crea una permutazione casuale delle etichette'  
  local i,l,j,x,z  
  x=RND:z=461.97  
  for i=nv to nw step -1  
    l=H%(i):x=(RND+x)*z:x=x-INT(x):j=INT(x*i)+1:H%(i)=H%(j):H%(j)=l  
  next  
END SUB
```

```
SUB GENET(H%(1),nv) 'crea il vettore base etichette'  
  local i  
  for i=1 to nv:H%(i)=i:next  
END SUB
```

```
SUB INVNOR(x,z)  
  local i,y:i=1  
  IF x<0.5 THEN i=-1:ELSE x=1-x  
  z=SQR(-2*LOG(x))  
  y=2.515517+0.802853*z+0.010328*z^2  
  y=y/(1+1.432788*z+0.189269*z^2+0.001308*z^3)  
  z=i*(z-y)  
END SUB
```