

Università degli Studi di Padova
Facoltà di Scienze Statistiche

Corso di Laurea in Statistica e Tecnologie Informatiche

Tesi di Laurea

**METODI DI SIMULAZIONE PER I
TEST RAPPORTO DI VEROSIMIGLIANZA E
UNIONE INTERSEZIONE**

Relatore:
Ch.mo Prof. Giancarlo Diana

Laureando: Michele Battocchio

Anno Accademico 2005-2006

INDICE

INTRODUZIONE

CAPITOLO 1

Test Rapporto di verosimiglianza (T.R.V.)

- 1.1 Definizione del test T.R.V.
- 1.2 Esempio del test T.R.V.
- 1.3 Correzione di Bartlett analitica
- 1.4 Correzione di Bartlett empirica
- 1.5 Stima della distribuzione del test T.R.V. tramite simulazione
- 1.6 Analisi della bontà dei metodi

CAPITOLO 2

Test Unione-Intersezione (U.I.T.)

- 2.1 Definizione del test U.I.T.
- 2.2 Esempio del test U.I.T.
- 2.3 Stima della distribuzione del test U.I.T. tramite simulazione

CAPITOLO 3

La funzione test

- 3.1 Introduzione al compilatore R
- 3.2 Algoritmo, che stima la distribuzione del test T.R.V.
- 3.3 Algoritmo, che stima la distribuzione del test U.I.T.
- 3.4 Tempi di esecuzione delle simulazioni

CAPITOLO 4

Un'applicazione pratica

- 4.1 Interfaccia della funzione test
- 4.2 Applicazione dei test T.R.V. sul data-set "iris3"
 - 4.2.1 Test sull'ipotesi che $\mu=\mu_0$, con Σ nota "TRV1"
 - 4.2.2 Test sull'ipotesi che $\mu=\mu_0$, con Σ non nota "TRV2"
 - 4.2.3 Test sull'ipotesi che $\Sigma=\Sigma_0$, con μ non nota "TRV3"
 - 4.2.11 Test sull'ipotesi che $\Sigma_1=\dots=\Sigma_g$, test di Box "TRV9"
- 4.3 Applicazione dei test U.I.T. sul data-set "iris3"
 - 4.3.1 Test sull'ipotesi che $\mu=\mu_0$, con Σ nota "UIT1"
 - 4.3.3 Test sull'ipotesi che $\Sigma=\Sigma_0$, con μ non nota "UIT3"

CONCLUSIONI

APPENDICE

Listato del programma

BIBLIOGRAFIA

INTRODUZIONE

Nell'ambito della statistica multivariata, la verifica d'ipotesi sui parametri della distribuzione normale viene abitualmente condotta con l'impiego di test basati sul rapporto di verosimiglianza, stante i risultati almeno asintotici disponibili per le distribuzioni delle statistiche test in ipotesi nulla.

Come è noto, così facendo si va incontro a due ordini di problemi:

- i) le approssimazioni che si ottengono usando la distribuzione limite non sono soddisfacenti se non per campioni di grandi dimensioni, a meno che non si utilizzi la correzione di Bartlett, quando nota;
- ii) il rifiuto dell'ipotesi nulla non consente di verificare, con un adeguato controllo dell'errore di prima specie, se il rifiuto stesso sia da imputare a uno o ad alcuni parametri di specifico interesse.

Al secondo problema si può, in linea teorica, dare soluzione ricorrendo ai test unione intersezione (U.I.T.) introdotti da Roy (1953). Purtroppo però, fatti salvi alcuni semplici casi, non si conoscono le distribuzioni delle statistiche test così ottenute. Presubilmente è per questa ragione che gli U.I.T. hanno avuto scarso sviluppo e impiego in questi anni.

Lo scopo del presente lavoro è mostrare come i metodi di simulazione possano costituire un valido strumento per la soluzione dei problemi descritti. L'attuale velocità di calcolo dei PC consente infatti di effettuare simulazioni ad hoc per ogni specifica situazione in tempi del tutto accettabili.

Nel caso i) si possono stimare direttamente quantità di interesse (percentile o p-value) della distribuzione del test rapporto di verosimiglianza o calcolare, via simulazione, una correzione di Bartlett empirica per migliorare le approssimazioni ottenibili utilizzando la distribuzione asintotica. Si vedrà come la correzione empirica possa essere addirittura migliore di quella analitica.

Nel caso ii) quando, come avviene spesso, non si conosce la distribuzione dell'U.I.T., la stessa può essere oggi facilmente simulata con riferimento alla specifica situazione applicativa d'interesse, consentendo l'impiego del test.

Il presente studio nasce dall'esigenza di disporre di un programma di pubblico dominio, che sia portatile e modificabile, in linea con le esigenze dettate dalla situazione che l'utente si trova ad analizzare.

In considerazione dell'utilizzo che di tale prodotto potrebbe venir fatto in futuro, l'impostazione seguita è rivolta a facilitare un suo possibile utilizzo didattico.

La tesi presenta allora il percorso di studio delineato. Il CAPITOLO 1 ha un carattere introduttivo, in cui si vuole illustrare le possibilità (e le problematiche) del test rapporto verosimiglianza.

Nel CAPITOLO 2 si presenta il test unione-intersezione, con relativa procedura che consente di calcolare i test U.I.T mediante simulazione. Il CAPITOLO 3 ha la funzione di collegamento tra la teoria presentata e la realizzazione al calcolatore. Si presentano cioè gli algoritmi che stanno alla base delle funzioni che simulano i test T.R.V. e U.I.T.

Il CAPITOLO 4 presenta infine qualche esempio specifico su data-set reali.

Infine nelle CONCLUSIONI, si riportano alcune riflessioni finali riguardanti il lavoro svolto.

CAPITOLO 1

TEST RAPPORTO DI VEROSIMIGLIANZA (T.R.V.)

1.1 Definizione del test T.R.V.

I test rapporto di verosimiglianza (T.R.V.), si fondano sulla logica dei test di significatività e portano, pertanto, all'accettazione o al rifiuto di una ipotesi multivariata nulla (${}_M H_0$), a seconda che tali siano le condizioni suggerite dai risultati campionari. Più in particolare, sia $X_{n \times p}$ una matrice di dati campionari da una distribuzione con funzione di probabilità $f(\underline{x}; \underline{\theta})$ dipendente da un parametro vettoriale $\underline{\theta}$. L'interesse è valutare la conformità dei dati ad una certa ipotesi sul vettore di parametri $\underline{\theta}$.

Si procede a saggiare il seguente sistema d'ipotesi:

$${}_M H_0: \underline{\theta} \in \Omega_0$$

$${}_M H_1: \underline{\theta} \in \Omega_1$$

dove:

${}_M H_0$ è l'ipotesi nulla

${}_M H_1$ è l'ipotesi alternativa

Ω_0 e Ω_1 sono spazi parametrici in cui varia $\underline{\theta}$ sotto ${}_M H_0$ e ${}_M H_1$ rispettivamente.

Un possibile modo per condurre la verifica in questione è il seguente.

Per prima cosa si trova il valore assunto dalla verosimiglianza nel punto

$\underline{\theta} = \underline{\theta}_0$ e nel punto $\underline{\theta} = \hat{\underline{\theta}}$, dove $\hat{\underline{\theta}}$ è la stima di massima verosimiglianza.

Indicato con $L(\hat{\theta})$ il valore massimo assunto dalla funzione di verosimiglianza, il rapporto

$$\lambda(X) = L(\theta_0) / L(\hat{\theta}) = L^*_0 / L^*_1$$

dove

$$L^*_i = \max_{\theta \in \Omega_i} L(\theta; X)$$

è una statistica-test per verificare MH_0 contro MH_1 .

Equivalentemente si può usare

$$-2 \log \lambda(X) = 2(\ell^*_1 - \ell^*_0)$$

$$\text{dove: } \ell^*_i = \log L^*_i$$

Sia $\lambda(X)$ che $-2 \log \lambda(X)$ vengono denominate statistica test basata sul rapporto di verosimiglianza.

Il rapporto può assumere valori che variano tra $(0,1]$. Se $\lambda \approx 0$ l'ipotesi nulla appare del tutto insoddisfacente, viceversa se $\lambda=1$, significa che c'è perfetta corrispondenza dei dati all'ipotesi.

Si sceglie un valore critico λ_c , tale per cui se $\lambda(X) < \lambda_c$ rifiuteremo l'ipotesi H_0 .

La regione di rifiuto della statistica rapporto di verosimiglianza è del tipo:

$$R_1 = \{ X \mid \lambda(X) < \lambda_{c1} \}$$

$$\text{ovvero } R_2 = \{ X \mid -2 \log \lambda(X) > \lambda_{c2} \}$$

dove λ_{c1} e λ_{c2} sono costanti tali che

$$\sup_{\theta \in \Omega_0} \Pr(X \in R_i) = \alpha$$

per $i=1,2$ con α prefissato.

Può verificarsi che si decida a favore di MH_1 anche se è vera MH_0 , ciò costituisce un errore di I tipo. Viceversa, può verificarsi che si decida a favore di MH_0 anche se è vera MH_1 , errore di II tipo.

Di conseguenza $\Pr(\lambda(X) < \lambda_c \mid H_0) = \alpha$, significa che si assegna un valore massimo α della probabilità di errore di I tipo e si cerca di rendere minima la probabilità di errore di II tipo.

Comunque non approfondiamo questo argomento, per i nostri scopi è sufficiente l'aver enunciato che nel seguito effettueremo dei test di livello preassegnato, che secondo una consuetudine diffusa, viene posto abitualmente pari a 0.05 o a 0.01.

Il valore critico λ_c , è l'opportuno percentile $1-\alpha$ della distribuzione di λ . Molto spesso però non è possibile la determinazione della distribuzione esatta di λ e quindi si deve ricorrere alla distribuzione asintotica. Il teorema di Wilks, dimostra che se n (numerosità campionaria) è sufficientemente elevata la verifica dell'ipotesi MH_0 tramite il rapporto di verosimiglianza può condursi ricorrendo alla v.c. Chi-Quadrato.

TEOREMA

Se $\Omega_1 \in \mathbb{R}^p$ e $\Omega_0 \in \mathbb{R}^q \subset \mathbb{R}^p$, sotto condizioni di regolarità, $\forall \theta \in \Omega_0$

$$-2\log \lambda(X) \xrightarrow{d} \chi_{p-q}^2 \quad \text{per } n \rightarrow \infty$$

Le approssimazioni che si ottengono usando la distribuzione limite non sono però soddisfacenti se non per campioni di grandi dimensioni.

1.2 Esempio del test T.R.V.

Sia $X_{n \times p}$ una matrice normale di dati da $N(\mu, \Sigma)$ con varianza nota e media ignota nei confronti della quale si formulano le ipotesi

$${}_M H_0: \mu = \mu_0$$

$${}_M H_1: \mu \neq \mu_0$$

che si intende verificare, a livello di significatività α , ricorrendo al test rapporto di verosimiglianza.

Poiché la funzione di verosimiglianza risulta

$$L(\mu) = |2\pi\Sigma|^{-n/2} \exp \left\{ -1/2 (\underline{x} - \mu)' \Sigma^{-1} (\underline{x} - \mu) \right\}$$

allora sotto ${}_M H_0: \mu = \mu_0$

$$\ell_0^* = -n/2 \log |2\pi \Sigma| - n/2 \text{tr} [\Sigma^{-1} S] - n/2 (\bar{\underline{x}} - \mu_0)' \Sigma^{-1} (\bar{\underline{x}} - \mu_0)$$

sotto ${}_M H_1: \mu \neq \mu_0$

$$\ell_1^* = -n/2 \log |2\pi \Sigma| - n/2 \text{tr} [\Sigma^{-1} S]$$

Il rapporto assume la forma

$$-2 \log \lambda = 2 (\ell_1^* - \ell_0^*) = n (\bar{\underline{x}} - \mu_0)' \Sigma^{-1} (\bar{\underline{x}} - \mu_0)$$

La regione di rifiuto di ${}_M H_0$ risulta

$$R_2 = \left\{ -2 \log \lambda \mid -2 \log \lambda = n (\bar{\underline{x}} - \mu_0)' \Sigma^{-1} (\bar{\underline{x}} - \mu_0) > \lambda_{c2} \right\}$$

Per assicurare il prefissato livello di significatività α si può procedere nel modo seguente. Poiché deve essere

$$\alpha = \Pr (-2\log \lambda > \lambda_{c2} \mid {}_M H_0) = \Pr \left\{ n (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) > \lambda_{c2} \mid {}_M H_0 \right\}$$

Ricordando che in questo caso particolare, sotto ${}_M H_0$, $-2\log \lambda$ si distribuisce esattamente come un χ_p^2 allora

$$\Pr (-2\log \lambda > \lambda_{c2} \mid {}_M H_0) = \Pr \left\{ \chi_p^2 > \lambda_{c2} \right\} = \alpha$$

Il valore λ_{c2} viene a coincidere con il valore $\chi_{1-\alpha}^2$, cioè con il percentile di ordine $(1 - \alpha)$ della v.c. χ_p^2 e quindi l'ipotesi H_0 viene rifiutata se

$$-2\log \lambda = n (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \geq \chi_{1-\alpha}^2$$

1.3 Correzione di Bartlett analitica

Quando, come nella maggior parte dei casi, non si conosce la distribuzione esatta del test, si ricorre alla distribuzione asintotica. Viene naturale porsi una domanda: la distribuzione asintotica utilizzata è buona?

Bartlett (1937) propose una correzione al test del rapporto di verosimiglianza (LR), per saggiare l'omogeneità di K varianze nel campionamento da popolazioni normali, noto come test di Bartlett. In termini generali, la correzione di Bartlett, che nel seguito chiameremo correzione di Bartlett analitica, può essere così descritta.

Se LR è un test rapporto di verosimiglianza con distribuzione asintotica χ_p^2 , si cerca quella quantità a tale che $a \rightarrow E(LR) / p$, per cui l'aggiustamento della statistica LR è

$$LR' = LR / a$$

sperando che LR' abbia una distribuzione più vicina al χ_p^2 che non LR.

La quantità a può essere ottenuta considerando lo sviluppo di Taylor di E(LR), da cui risulta che

$$E(LR) = p \{ 1 + b/n + O(n^{-2}) \}$$

e, quindi, prendendo $a = 1 + b/n$, si può considerare la statistica modificata

$$LR' = LR / (1 + b/n)$$

in cui b risulta essere una funzione esprimibile in termini di momenti delle derivate della log-verosimiglianza.

Lawley(1956) ha mostrato che il miglioramento non è solo nell'approssimazione del valore atteso, ma è simultaneo per tutti i momenti non nulli, e quindi per l'intera distribuzione rispetto al χ_p^2 .

Inoltre Barndoff-Nielsen e Hall (1988) hanno mostrato che LR' ha distribuzione Chi-Quadrato con p gradi di libertà con errore di ordine $O(n^{-2})$, anziché di ordine $O(n^{-1})$, come accade per LR.

Al paragrafo 1.6 verrà valutata l'efficacia della correzione di Bartlett analitica.

1.4 Correzione di Bartlett empirica

Accade talvolta, che sia troppo laborioso determinare la correzione di Bartlett analitica. In questi casi si può pensare di risolvere il problema tramite la simulazione.

Infatti, a livello di principio, è possibile ottenere un'approssimazione accurata quanto si vuole di $\{1+b/n+O(n^{-2})\}$.

Questa viene denominata **correzione di Bartlett empirica1**.

A differenza della correzione di Bartlett analitica, l'idea è quella di stimare tramite simulazione il parametro $\{1+b/n+O(n^{-2})\}$, che chiameremo a_1 .

La logica di tale simulazione è abbastanza semplice e può essere così esposta:

Dato che

$$a_1 = E(LR) / p$$

si può stimare a_1 a partire da una stima di $E(LR)$, la cui approssimazione migliora con l'aumentare delle replicazioni.

Vale la pena di sottolineare che a_1 risulta preferibile ad a non solo per la facilità con cui si può ottenere la correzione empirica¹ rispetto alla correzione analitica, ma anche perché la bontà dell'accostamento alla distribuzione limite dipende in questo caso più dal numero N di replicazioni che si vogliono fare che dalla numerosità campionaria con cui si trova ad operare, come avviene per la correzione analitica. In compenso occorre ripetere la simulazione per stimare $E(LR)$ ogni qual volta che cambiano in tutto o in parte n e p .

Proseguendo sulla stessa linea, si è pensato di eguagliare oltre alla media, anche la varianza della statistica test con la media e la varianza della distribuzione χ_p^2 .

Questa viene denominata **correzione di Bartlett empirica²**.

L'idea è quella di determinare, sempre tramite simulazione, due parametri a_2 e c , in modo che LR/a_2+c abbia media e varianza pari a quella della distribuzione χ_p^2 , ossia

$$E(LR/a_2+c) = p$$

$$V(LR/a_2+c) = 2*p$$

da queste espressioni si ricava a_2 e c

$$a_2 = [V(LR) / (2*p)]^{1/2}$$

$$c = p - E(LR) / a_2$$

Come nel caso precedente, andiamo a stimare $V(LR)$ e $E(LR)$ tramite simulazione e quindi a_2 e c .

Grazie a questi parametri a_2 e c , si ha un "controllo" oltre sulla media, anche sulla varianza della statistica test.

L'efficienza di queste due nuove correzioni, verrà verificata nel paragrafo 1.6.

1.5 Stima della distribuzione del test T.R.V. tramite simulazione

Un'alternativa alle correzioni descritte nei paragrafi precedenti è la simulazione.

La simulazione permette di ottenere la stima della distribuzione del test d'interesse mediante delle replicazioni: tramite un generatore di numeri pseudo-casuali, si simulano dei campioni di dati di uguale dimensione al campione originale $X_{n \times p}$, sui quali si calcola il test LR. Da essi verrà poi stimata la distribuzione del test considerato.

L'attenzione rivolta dallo statistico al generatore pseudo-casuale è, perciò, doverosa visto che i risultati della simulazione sono fortemente condizionati dalle sue caratteristiche.

Per i nostri scopi è sufficiente enunciare che nel seguito useremo il generatore *rmvnorm*, contenuto nella libreria *mtvnorm*.

Il problema fondamentale consiste nel determinare il numero di simulazioni per avere una buona stima della distribuzione del test T.R.V. nel caso non fosse nota la vera distribuzione per la verifica dell'ipotesi nulla (MH_0).

La verifica d'ipotesi avviene secondo due possibili approcci:

il primo stimando il percentile $1-\alpha$, il secondo stimando il p-value associato al valore del test calcolato sul campione effettivamente osservato, lasciando così piena libertà all'utente nell'interpretazione dei risultati.

Il primo approccio verrà approfondito nel paragrafo 2.3, mentre di seguito si sviluppa nel dettaglio, il secondo approccio.

L'idea alla base della procedura è abbastanza semplice:

- Si fissa un valore N di simulazioni e in ciascuna di esse si calcola la statistica test LR.

- Si utilizza come valore soglia t_0 , il valore della statistica test calcolata sul campione effettivamente osservato.
- Sia i una variabile conteggio, per ogni simulazione si esegue la seguente verifica:

$$\left. \begin{array}{l} i = 1 \text{ se } LR > t_0 \\ i = 0 \text{ se } LR < t_0 \end{array} \right\}$$

Ne segue che:

$$i \sim \text{Bi}(1, p)$$

dove

$$\hat{p} = (\# \text{volte } i=1) / N$$

- Supponiamo nel seguito di poter ritenere valida l'approssimazione alla normale della distribuzione di \hat{p} .

Ipotizziamo dapprima che sia pari a d l'errore assoluto accettato nella stima di p e che si sia fissata pari a 0.01 la probabilità di commettere un errore maggiore, imponiamo cioè che sia

$$\Pr(|\hat{p} - p| > d) = 0.01$$

Sapendo che

$$Z = \frac{(\hat{p} - p)}{\sqrt{V(\hat{p})}}$$

è prossima ad una normale standardizzata $Z \sim \mathcal{N}(0,1)$

Ne segue che

$$|Z| < d / V(\hat{p})^{1/2} = Z_{1-\alpha/2}$$

$$\frac{d}{\sqrt{V(\hat{p})}} = 2.57 \rightarrow Z_{0.995}$$

dove

$$V(\hat{p}) = p(1-p) / N$$

Basterà dunque ricavare N dall'uguaglianza

$$d = 2.57 * V(\hat{p})^{1/2}$$

ovvero

$$N = (2.57 / d)^2 * p(1-p)$$

- Se ad esempio impostiamo un errore assoluto $d=5e-3$, che è anche il valore imposto di default nelle nostre funzioni e si considera la condizione di massima variabilità, cioè $p=0.5$, si ottiene un N pari a

$$N = 66049$$

In questo modo, compiendo 66049 replicazioni, ci si assicura che

$$\Pr (|\hat{p} - p| > d) = 0.01$$

Questo però, è il caso peggiore, cioè di massima variabilità. Nelle funzioni viene utilizzato un valore di default pari a 0.05.

Viene posto questo valore per due motivi:

- si assegna $\Pr(LR > \lambda_{c2} \mid M H_0) = 0.05$, cioè l'errore di I tipo pari a 0.05
- il tempo di esecuzione dell'algoritmo, è direttamente collegato al numero di cicli che si vuole realizzare nella simulazione.
Con il valore di $p = 0.05$, il numero di replicazioni diviene pari a $N = 12549$, però in questo caso non ci si assicura che

$$\Pr (|\hat{p} - p| > d) = 0.01$$

Per ulteriori informazioni si rimanda al paragrafo 3.4.

Ad ogni modo, l'utente ha la facoltà di modificare il valore di p e dell'errore, lasciando così piena libertà all'utente nell'interpretazione dei risultati.

1.6 Analisi della bontà dei metodi

Lo scopo di questo paragrafo, è valutare l'analisi della bontà dei metodi presentati nei paragrafi precedenti. A tale scopo, sono state create delle funzioni (vedi appendice), con l'obiettivo di stimare in corrispondenza di un α nominale del 5% l' α effettivo, sulla base di $N=10000$ replicazioni, per lo stesso test con correzione di Bartlett analitica, empirica1, empirica2 e simulazione.

Per la stima dei parametri delle correzioni di Bartlett empiriche e per la stima dei percentili della distribuzione di LR è stato usato $d=5e-3$, che equivale a $N=12549$ replicazioni.

I test presi in considerazione sono quelli riportati di seguito.

Test d'ipotesi che $\Sigma=\Sigma_0$, con μ non nota

Sia $X_{n \times p}$ una matrice normale di dati da $N_p(\mu, \Sigma)$, si vuole verificare l'ipotesi:

- $MH_0: \quad \Sigma = \Sigma_0 \quad \rightarrow$ con μ non nota
- $MH_1: \quad \Sigma \neq \Sigma_0$

Per la verifica dell'ipotesi in questione si utilizza il test:

$$\text{TRV3} = \text{LR} = np (a - \log(g) - 1)$$

dove:

g è la media geometrica degli autovalori di $\Sigma_0^{-1} S$

a è la media aritmetica degli autovalori di $\Sigma_0^{-1} S$

sotto MH_0 : $\text{TRV3} \xrightarrow{d} \chi_{p(p+1)/2}^2$ per $n \rightarrow \infty$

La correzione di Bartlett analitica è:

$$a = 1 / [1 - \{ 2 \cdot p + 1 - 2 / (p - 1) \} / (6 \cdot n)]$$

Inoltre, solo per questa verifica d'ipotesi, esiste un'alternativa alla correzione di Bartlett analitica, proposta da Korin(1968).

Nel suo lavoro l'autore mostra come la distribuzione del test LR/b converga alla distribuzione di una F con q_1 e q_2 gradi di libertà, dove:

$$q_1 = v \quad q_2 = (v + 2) / (D_2 - D_1^2)$$

e

$$b = v / (1 - D_1 - v / q_2)$$

con

$$v = \frac{1}{2} p (p + 1)$$

$$D_2 - D_1^2 = \{ 2 p^2 + 2 p - 5 - [4 (p + 2) / (p + 1)^2] \} / (36 n^2)$$

$$D_1 = \{ 2 p + 1 - 2 / (p + 1) \} / (6 n)$$

Utilizzando la funzione Bontà3 dell'appendice, si è ottenuto la seguente tabella:

Bontà dei metodi TRV3

Tab.1.6.1 Stima degli alpha effettivi con alpha nominale del 5%

Variabili P	Numerosità n	G.d.l.	Alpha	Test LR	Bartlett analitico	Korin	Bartlett empirico1	Bartlett empirico2	Simulazione
2	5	3	0.05	0.183	0.125	0.122	0.047	0.048	0.050
2	10	3	0.05	0.096	0.079	0.079	0.052	0.053	0.054
2	20	3	0.05	0.075	0.067	0.067	0.052	0.053	0.052
2	30	3	0.05	0.063	0.058	0.058	0.047	0.048	0.048
2	40	3	0.05	0.059	0.056	0.056	0.050	0.050	0.051
2	50	3	0.05	0.059	0.056	0.056	0.050	0.049	0.049
2	60	3	0.05	0.059	0.056	0.056	0.049	0.049	0.048
2	65	3	0.05	0.058	0.055	0.055	0.055	0.054	0.053
2	70	3	0.05	0.055	0.053	0.053	0.049	0.048	0.048
2	80	3	0.05	0.056	0.054	0.054	0.048	0.047	0.047

2	100	3	0.05	0.058	0.056	0.056	0.055	0.056	0.058
2	130	3	0.05	0.051	0.049	0.049	0.049	0.049	0.050
2	150	3	0.05	0.050	0.050	0.050	0.051	0.053	0.052
3	50	6	0.05	0.063	0.057	0.057	0.051	0.051	0.049
5	70	15	0.05	0.062	0.053	0.053	0.048	0.048	0.047
5	90	15	0.05	0.062	0.053	0.053	0.050	0.051	0.051
8	100	36	0.05	0.073	0.056	0.055	0.050	0.051	0.051
10	50	55	0.05	0.141	0.070	0.068	0.049	0.049	0.050
10	100	55	0.05	0.086	0.056	0.056	0.048	0.046	0.048
15	50	120	0.05	0.434	0.099	0.087	0.055	0.055	0.054
20	150	210	0.05	0.149	0.064	0.063	0.053	0.053	0.053

Come si può vedere, sono stati evidenziati in neretto i casi in cui approssimando alla seconda cifra dopo la virgola, il valore dell'alpha stimato risulta diverso da 0.05.

Dalla tabella si nota, come previsto, che il metodo peggiore risulta essere il test asintotico (LR). Inoltre si può vedere che la correzione di Bartlett analitica, in alcuni casi è "peggiore" del metodo Korin. Infatti la "convergenza" della statistica test verso la distribuzione asintotica $\chi^2_{p(p+1)/2}$, viene raggiunta, a parità dei gradi di libertà, ad una numerosità campionaria più alta.

Consideriamo ora una misura sintetica della loro variabilità.

La deviazione standard dei metodi considerati nella tabella sono:

Deviazione standard	Test LR	Bartlett Analitico	Korin	Bartlett empirico1	Bartlett empirico2	Simulazione
S	69,47e-3	22,52e-3	20,69e-3	2,53e-3	2,83e-3	2,83e-3

Dai risultati ottenuti, si possono ricavare alcune informazioni generali molto interessanti.

Innanzitutto, vengono confermate alcune considerazioni espresse nella parte teorica: la deviazione standard del test LR è la più elevata; quella di Korin, è inferiore a quella di Bartlett analitico.

Si osservi che, le differenze tra la deviazione standard della correzione di Bartlett empirico1, empirico2 e della simulazione risultano minime, e i tre metodi considerati risulterebbero, pertanto, equivalenti.

Test d'ipotesi che $\Sigma_{12}=0$, con μ non nota

Sia $X_{n \times p}$ una matrice normale di dati da $N_p(\mu, \Sigma)$, si vuole verificare l'ipotesi:

- $MH_0: \Sigma_{12} = 0 \rightarrow$ con μ non nota
- $MH_1: \Sigma_{12} \neq 0$

Partizionando la matrice dei dati in due sottomatrici di dimensioni $(n \times p_1)$ e $(n \times p_2)$ rispettivamente, la verifica dell'ipotesi in questione si realizza ricorrendo al test rapporto di verosimiglianza dato da:

$$TRV72 = LR = - (n-1/2(p_1+p_2+3)) \log |I - S^{-1}_{22} S_{21} S^{-1}_{11} S_{12}|$$

sotto MH_0 : $TRV72 \xrightarrow{d} \chi^2_{p_1 p_2}$ per $n \rightarrow \infty$

In questo caso, il rapporto di verosimiglianza ha già presente la correzione di Bartlett analitica, che è pari a:

$$a = - (n-1/2(p_1+p_2+3))$$

Come nel caso precedente, prendendo in considerazione la funzione Bontà72 dell'appendice, otteniamo la seguente tabella:

Bontà dei metodi TRV72

Tab.1.6.2 Stima degli alpha effettivi con alpha nominale del 5%

Variabili p1	Variabili p2	Numerosità N	G.d.l.	Alpha	Bartlett analitico	Bartlett Empirico1	Bartlett empirico2	Simulazione
2	2	5	4	0.05	0.085	0.057	0.052	0.051
2	2	10	4	0.05	0.050	0.048	0.048	0.049
2	3	10	6	0.05	0.056	0.052	0.050	0.051
3	2	15	6	0.05	0.050	0.048	0.049	0.050
4	2	10	8	0.05	0.071	0.057	0.054	0.053
4	2	15	8	0.05	0.053	0.049	0.049	0.051
3	3	20	9	0.05	0.050	0.050	0.050	0.051
4	3	15	12	0.05	0.057	0.051	0.050	0.050
4	4	10	16	0.05	0.128	0.063	0.051	0.049
4	4	20	16	0.05	0.044	0.044	0.044	0.046
4	4	30	16	0.05	0.051	0.049	0.049	0.051
4	4	40	16	0.05	0.050	0.050	0.049	0.051
5	4	40	20	0.05	0.056	0.056	0.056	0.057
5	5	50	25	0.05	0.052	0.054	0.054	0.053
6	5	50	30	0.05	0.051	0.049	0.049	0.047
6	6	50	36	0.05	0.053	0.050	0.050	0.051
7	6	50	42	0.05	0.053	0.049	0.049	0.048
8	7	50	56	0.05	0.055	0.054	0.055	0.055
10	8	50	80	0.05	0.053	0.048	0.047	0.047
10	10	50	100	0.05	0.061	0.052	0.051	0.048
12	10	50	120	0.05	0.064	0.050	0.052	0.053
14	10	50	140	0.05	0.070	0.052	0.052	0.053

Anche in questo caso sono stati evidenziati in neretto i casi in cui approssimando alla seconda cifra dopo la virgola, il valore dell'alpha stimato risulta diverso da 0.05.

Da una prima analisi della tabella si nota, come previsto, tra i metodi considerati quello peggiore risulta essere il test con la correzione di Bartlett analitica.

Passiamo ora ad una analisi più approfondita.

Le deviazioni standard dei metodi considerati sono:

Deviazione Standard	Bartlett Analitico	Bartlett empirico1	Bartlett empirico2	Simulazione
S	20,34e-3	4,36e-3	2,78e-3	2,76e-3

Si può facilmente notare che viene confermata la considerazione precedente. Infatti il valore della deviazione standard della correzione Bartlett analitico è la più grande di tutti.

Per quanto riguarda gli altri metodi, è importante sottolineare che i metodi migliori risultano la simulazione e il Bartlett empirico².

Test d'ipotesi che $\Sigma = I$, con μ non nota

Sia $X_{n \times p}$ una matrice normale di dati da $N_p(\mu, \Sigma)$, si vuole verificare l'ipotesi:

- $MH_0: \quad \Sigma = I \rightarrow$ con μ non nota
- $MH_1: \quad \Sigma \neq I$

La verifica dell'ipotesi in questione si realizza ricorrendo al test:

TRV7 = LR = - n log | R |

dove:

R è la matrice di correlazione campionaria

sotto MH_0 : $TRV7 \xrightarrow{d} \chi^2_{p(p-1)/2}$ per $n \rightarrow \infty$

La statistica corretta è:

$$TRV7' = TRV7 \{1 - (2p+11)/(6n)\}$$

dove

$$a = 1 / \{1 - (2p+11)/(6n)\}$$

Si presentano in una tabella, alcuni esempi di utilizzo della funzione Bontà7.

Bontà dei metodi TRV7

Tab.1.6.3 Stima degli alpha effettivi con alpha nominale del 5%

Variabili P	Numerosità n	G.d.l.	Alpha	Test LR	Bartlett analitico	Bartlett Empirico1	Bartlett empirico2	Simulazione
2	4	1	0.05	0.208	0.039	0.048	0.051	0.051
2	5	1	0.05	0.154	0.040	0.043	0.044	0.042
2	6	1	0.05	0.136	0.046	0.047	0.049	0.048
2	10	1	0.05	0.089	0.047	0.050	0.050	0.051
3	4	3	0.05	0.525	0.057	0.053	0.050	0.051
3	5	3	0.05	0.335	0.051	0.050	0.051	0.053
3	10	3	0.05	0.132	0.049	0.051	0.051	0.048
3	15	3	0.05	0.094	0.044	0.045	0.045	0.050
3	20	3	0.05	0.080	0.051	0.052	0.051	0.051
4	5	6	0.05	0.659	0.101	0.060	0.050	0.050
4	10	6	0.05	0.204	0.053	0.052	0.051	0.050
4	10	6	0.05	0.207	0.053	0.051	0.051	0.052
4	15	6	0.05	0.130	0.052	0.051	0.052	0.052
4	20	6	0.05	0.109	0.057	0.055	0.056	0.056
4	30	6	0.05	0.080	0.049	0.049	0.050	0.050
4	40	6	0.05	0.073	0.050	0.049	0.048	0.047
4	100	6	0.05	0.059	0.050	0.050	0.051	0.052
4	180	6	0.05	0.053	0.050	0.049	0.048	0.047
5	40	10	0.05	0.084	0.052	0.052	0.054	0.054
5	50	10	0.05	0.070	0.048	0.047	0.047	0.049
6	50	15	0.05	0.082	0.050	0.049	0.049	0.049
6	50	15	0.05	0.078	0.048	0.049	0.050	0.049
7	50	21	0.05	0.090	0.048	0.047	0.048	0.047
8	50	28	0.05	0.106	0.048	0.046	0.046	0.046
10	20	45	0.05	0.483	0.066	0.048	0.045	0.045
10	30	45	0.05	0.261	0.059	0.054	0.052	0.052
10	50	45	0.05	0.144	0.053	0.052	0.051	0.051
10	50	45	0.05	0.140	0.050	0.050	0.051	0.052
12	50	66	0.05	0.192	0.048	0.045	0.045	0.045
14	50	91	0.05	0.270	0.054	0.050	0.050	0.050

Come dalle aspettative, i metodi migliori risultano quelli con correzione di Bartlett empirica1, empirica2 e simulazione .

Passiamo ora alla deviazione standard:

Deviazione standard	Test LR	Bartlett Analitico	Bartlett empirico1	Bartlett empirico2	Simulazione
S	19,63e-2	10,81e-3	3,82e-3	2,75e-3	2,96e-3

Prendendo visione di questi valori si possono confermare le considerazioni fatte precedentemente.

Inoltre si può notare la “superiorità” del Bartlett empirico2 e della simulazione rispetto agli altri metodi.

Nel nostro caso, comunque, non possiamo affermare che uno di questi due metodi sia migliore dell’altro. Si noti, infatti, che per la deviazione standard risulta migliore il Bartlett empirico2. Invece considerando l’approssimazione dell’alpha stimato, risulta essere migliore la simulazione.

Test d’ipotesi che $\Sigma_1 = \dots = \Sigma_g$ (test di Box)

Consideriamo g matrici normali di dati da $N_p(\mu, \Sigma)$ con dimensione $X_{n \times p}$, si vuole verificare l’ipotesi:

- $MH_0: \Sigma_1 = \dots = \Sigma_g$
- $MH_1: \Sigma_1 \neq \dots \neq \Sigma_g \rightarrow$ almeno una sia diversa

La verifica dell’ipotesi in questione si realizza ricorrendo al test:

$$TRV9 = LR = \sum_{i=1}^g n_i \log | S_i^{-1} S | = - n \log | S | - \sum_{i=1}^g n_i \log | S_i |$$

sotto MH_0 : $TRV9 \xrightarrow{d} \chi_{p(p+1)(g-1)/2}^2$ per $n \rightarrow \infty$

La correzione di Bartlett analitica è:

$$a = 1 - (2 * p^2 + 3 * p - 1) / (6 * (p+1) * (g-1)) * \left(\sum_{i=1}^g (1 / (n_i - 1)) - (1 / (n - k)) \right)$$

La statistica test corretta si ottiene facendo:

$$TRV9' = a * \sum_{i=1}^g (n_i - 1) * \log | S_{ui}^{-1} S_u |$$

dove

$$S_{ui} = \frac{n_i}{n_i - 1} S_i \quad e \quad S_u = \frac{n}{n - g} S$$

Utilizzando la funzione Bontà9, si ottiene la seguente tabella:

Bontà dei metodi TRV9

Tab.1.6.4 Stima degli alpha effettivi con alpha nominale del 5%

Campioni G	Variabili p	Numerosità Campione	G.d.l.	Alpha	Test LR	Bartlett analitico	Bartlett empirico1	Bartlett empirico2	Simulazione
2	2	5	3	0.05	0.204	0.050	0.050	0.049	0.049
3	3	5	12	0.05	0.586	0.068	0.052	0.050	0.050
3	3	25	12	0.05	0.093	0.053	0.052	0.053	0.054
7	2	5	18	0.05	0.450	0.055	0.052	0.053	0.054
10	2	5	27	0.05	0.541	0.055	0.051	0.051	0.051
10	2	10	27	0.05	0.201	0.051	0.047	0.048	0.048
2	7	10	28	0.05	0.854	0.108	0.057	0.054	0.054
4	4	5	30	0.05	0.978	0.227	0.061	0.049	0.049
4	4	25	30	0.05	0.128	0.050	0.050	0.050	0.050
4	4	220	30	0.05	0.054	0.051	0.052	0.049	0.048
15	2	5	42	0.05	0.688	0.058	0.050	0.051	0.051
10	3	5	54	0.05	0.948	0.105	0.057	0.052	0.051
10	3	30	54	0.05	0.115	0.051	0.053	0.052	0.051
5	5	6	60	0.05	0.999	0.399	0.077	0.053	0.051
6	5	6	75	0.05	0.999	0.450	0.080	0.051	0.048
6	5	25	75	0.05	0.224	0.056	0.054	0.055	0.054
4	7	25	84	0.05	0.341	0.052	0.048	0.046	0.046
10	4	5	90	0.05	0.999	0.430	0.073	0.052	0.050
10	4	20	90	0.05	0.261	0.058	0.055	0.056	0.054
5	7	8	112	0.05	1.000	0.638	0.089	0.051	0.047
5	7	15	112	0.05	0.803	0.074	0.053	0.052	0.051
6	7	20	140	0.05	0.610	0.063	0.052	0.051	0.049
7	7	20	168	0.05	0.654	0.062	0.053	0.051	0.049

Da una prima visione dei valori ottenuti, si possono trarre le stesse conclusioni dei casi precedenti.

Inoltre, possiamo estendere l'ipotesi di Box. Infatti, la correzione di Bartlett analitica fa sì che le cose migliorino decisamente non solo se vengono rispettati i vincoli usualmente posti sui parametri (g e p non superiori a 5 e n almeno 20) ma qualche volta anche quando il numero g di campioni o quello p delle variabili supera la soglia. In questi casi sembra importante che la numerosità campionaria non sia piccola in rapporto ai gradi di libertà.

Passiamo ad una ulteriore analisi, le deviazioni standard:

Deviazioni Standard	Test LR	Bartlett analitico	Bartlett empirico1	Bartlett empirico2	Simulazione
S	62,07e-2	19,20e-2	13,48e-3	2,6e-3	2,4e-3

Dall'ultima analisi condotta, possiamo confermare che i metodi migliori anche in questo caso risultano la simulazione e il Bartlett empirico2.

Dall'analisi effettuate sorge spontaneo affermare che i metodi migliori risultano essere la correzione di Bartlett empirico2 e la simulazione.

Questo però è in parziale contrasto con le aspettative teoriche. Infatti la correzione Bartlett empirico1 introduce un miglioramento per tutti i momenti, e quindi compresa anche la varianza, di conseguenza non dovrebbero esserci delle differenze tra il Bartlett empirico1 e Bartlett empirico2.

Si pone allora il problema di sceglierne uno come “migliore”.

Nel nostro caso, prediligiamo la correzione di Bartlett empirica² quando lavoriamo con più campioni di dimensione uguale. Questo perché si può stimare una sola volta la correzione di Bartlett empirica² e utilizzarla poi per tutti i campioni.

Invece prediligiamo la simulazione quando cambiano in tutto o in parte la numerosità o le variabili del campione d'interesse. Infatti la simulazione deve essere ripetuta ogni volta, perché stima il p-value associato al valore del test calcolato sul campione effettivamente osservato.

Nel capitolo 4 (Un'applicazione pratica), verranno effettuati degli esempi pratici per evidenziare ciò che è stato detto poc'anzi.

CAPITOLO 2

TEST UNIONE-INTERSEZIONE (U.I.T.)

2.1 Definizione del test U.I.T.

L'impiego di test basati sul rapporto di verosimiglianza per la verifica di un'ipotesi multivariata nulla (${}_M H_0$), portano ad un problema. Infatti, il rifiuto dell'ipotesi nulla non consente di verificare, con un adeguato controllo dell'errore di prima specie, se il rifiuto sia da imputare a uno o ad alcuni parametri di specifico interesse.

A questo problema si può, in linea teorica, dare soluzione ricorrendo ai test unione intersezione (U.I.T.) introdotti da Roy (1953).

Procedura di costruzione del test U.I.T.

Data un'ipotesi multivariata del tipo di quelle considerate nel CAPITOLO 1, si decompone l'ipotesi ${}_M H_0$ in un insieme d'ipotesi univariate ${}_{\mu} H_{0a}$, dove \underline{a} è un opportuno vettore di coefficienti in modo da:

- **Accettare** ${}_M H_0$ se ogni ${}_{\mu} H_{0a}$ **viene accettata**

$${}_M H_0 : \bigcap_a {}_{\mu} H_{0a}$$

- **Rifiutare** ${}_M H_0$ se qualche ${}_{\mu} H_{0a}$ **viene rifiutata**

$${}_M H_1 : \bigcup_a {}_{\mu} H_{1a}$$

Regione di rifiuto: $\mathbf{R} = \bigcup_a \mathbf{R}_a$

Regione di accettazione: $\mathbf{A} = \bigcap_a \mathbf{A}_a$

L'obiettivo è quello di costruire una funzione test "ragionevole" per ognuna delle ipotesi componenti ed eliminare la dipendenza dallo specifico μ_a

T_a statistica test per μH_{0a}

Test unilaterali

Nell'ipotesi che al variare di μ_a tutte le ipotesi alternative siano unilaterali, con regione di accettazione del tipo $\{-\infty, T_a\}$, oppure $\{T_a, \infty\}$, si ha come statistica test per $M H_0$:

$$T = \max_a T_a \quad \text{oppure} \quad T = \min_a T_a$$

Si accetta $M H_0$ se:

$$T = \max_a \{T_a\} \leq T_{1-\alpha} \quad \text{oppure} \quad T = \min_a \{T_a\} \geq T_{1-\alpha}$$

con $1-\alpha$ livello di significatività associato a T

Come è noto, se $M H_0$ viene accettata con probabilità $(1 - \alpha)$, alla singola ipotesi componente resta associato un livello di significatività $\geq (1 - \alpha)$.

Infatti se

$$P\{ T \leq T_{1-\alpha} \mid M H_0 \} = 1 - \alpha$$

$\forall a$ si ha

$$T_a \leq T$$

dunque

$$P\{ T_a \leq T_{1-\alpha} \mid M H_0 \} \geq 1 - \alpha$$

Analogamente se

$$P\{ T \geq T_{1-\alpha} \mid M H_0 \} = 1 - \alpha$$

$$\forall a, T_a \geq T$$

e quindi

$$P\{ T_a \geq T_{1-\alpha} \mid M H_0 \} \geq 1 - \alpha$$

Test bilaterali

Per i test bilaterali si costruiscono due regioni critiche in corrispondenza rispettivamente a:

$$T_1 = \max_a \{T_a\} \quad e \quad T_2 = \min_a \{T_a\}$$

Si accetta $M H_0$ se:

$$T_1 \leq T_{1-\alpha/2} \quad e \quad T_2 \geq T_{\alpha/2}$$

Il test Unione-Intersezione (U.I.T.) ha un vantaggio rispetto al test Rapporto di Verosimiglianza (T.R.V.):

Si può stabilire se una qualche $_{\mu}H_{0a}$ d'interesse abbia portato al rifiuto di $_{M}H_0$. Fissato α , si ricava T_{α} e si procede a confrontare ogni determinazione T_a con T_{α} , identificando l'ipotesi componente significativa.

2.2 Esempio del test U.I.T.

Si riprende lo stesso esempio del test T.R.V., in modo tale da notare le differenze dei due metodi. Sia $X_{n \times p}$ una matrice normale di dati da $N_p(\mu, \Sigma)$ con matrice di covarianze nota e vettore media ignoto nei confronti della quale si formula l'ipotesi:

$$_{M}H_0: \mu = \mu_0$$

$$_{M}H_1: \mu \neq \mu_0$$

posto: $y = Xa$, campione casuale di n osservazioni indipendenti, si ottiene:

$$y = a'x \sim (a'\mu, a'\Sigma a)$$

con

$$\bar{y} = \sum_{i=1}^n y_i / n = a' \bar{x}$$

$$E[\bar{y} \mid \mu H_{0a}] = a' \mu_0$$

$$V[\bar{y} \mid \mu H_{0a}] = V[\bar{y}] = a' \Sigma a / n$$

Dunque sotto μH_{0a} :

$$y \sim (\underline{a}' \mu_0, \underline{a}' \Sigma / n \underline{a})$$

Quindi al variare della configurazione di \underline{a} :

$$\underline{a}_1 = \begin{pmatrix} 1 \\ 0 \\ \dots \\ \dots \\ 0 \end{pmatrix}, \dots, \underline{a}_j = \begin{pmatrix} 0 \\ \dots \\ 1 \\ 0 \\ \dots \end{pmatrix}, \dots, \underline{a}_p = \begin{pmatrix} 0 \\ \dots \\ \dots \\ 0 \\ 1 \end{pmatrix} \dots$$

si ottengono infinite **ipotesi univariate** del tipo:

$$\mu H_{0a} : \underline{a}' \mu = \underline{a}' \mu_0$$

La statistica test per μH_{0a} è:

$$z_a = | \underline{a}' \bar{x} - \underline{a}' \mu_0 | / (\underline{a}' \Sigma / n \underline{a})^{1/2}$$

Regione di rifiuto:

$$R_a = \{ z_a : z_a^2 > C_\alpha \} \quad \forall \underline{a}, \text{ ossia}$$

μH_0 viene accettata se e solo se: $z_a^2 \leq C_\alpha, \quad \forall \underline{a},$ ovvero se

$$T = \max_a \{ z_a^2 \} \leq C_\alpha$$

dove $C_\alpha,$ 100(1- α) percentile della distribuzione di T

Eliminazione della dipendenza da \underline{a}

Si cerca un \underline{a}^* tale che la distanza tra l'ipotesi e l'evidenza campionaria sia massima.

Se $T \leq \mathbf{z}_{\underline{a}}^2 \Rightarrow$ accettazione di qualsiasi $\mu_{H_{0a}}$ \Rightarrow accettazione di MH_0

$$T = \mathbf{z}_{\underline{a}^*}^2 = \max_{\underline{a}} \{ n \underline{a}' (\bar{\mathbf{x}} - \underline{\mu}_0)' (\underline{a}' \Sigma \underline{a})^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0) \underline{a} \}$$

si può massimizzare rispetto $\underline{a} \neq 0$

Massimante di $\mathbf{z}_{\underline{a}}^2$

$$\underline{a}^* = \Sigma^{-1}(\bar{\mathbf{x}} - \underline{\mu}_0) / [(\bar{\mathbf{x}} - \underline{\mu}_0)' \Sigma^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0)]^{1/2}$$

\underline{a}^* è l'autovettore della matrice associato all'autovalore massimo della matrice

$$\Sigma^{-1}(\bar{\mathbf{x}} - \underline{\mu}_0)(\bar{\mathbf{x}} - \underline{\mu}_0)'$$

$$T = \mathbf{z}_{\underline{a}^*}^2 = n(\bar{\mathbf{x}} - \underline{\mu}_0)' \Sigma^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0)$$

Ora essendo $(\bar{\mathbf{x}} - \underline{\mu}_0)' \Sigma^{-1} (\bar{\mathbf{x}} - \underline{\mu}_0) \sim \chi_p^2$, risulta che sotto MH_0 :

$$\mathbf{z}_{\underline{a}^*}^2 \sim \chi_p^2 \Rightarrow \Pr\{ \mathbf{z}_{\underline{a}^*}^2 > C_\alpha | H_0 \} = \alpha$$

con C_α , pari al 100(1- α) percentile da χ_p^2

N.B.

Per le singole componenti si accetta $_{\mu}H_{0a}$ se: $z_a^2 \leq C_{\alpha}$

Inoltre, questo è uno dei pochi casi in cui il test U.I.T. **coincide** con il test T.R.V.

In particolare:

- Se si **accetta** $_{M}H_0$, questo implica l'accettazione simultanea di tutte le p ipotesi componenti.
- Se si **rifiuta** $_{M}H_0$, a livello α , allora si rifiuta almeno un'ipotesi $_{\mu}H_{0a}$ e a differenza del test T.R.V., si può verificare se una specifica "ipotesi componente" ha portato al rifiuto.

2.3 Stima della distribuzione del test U.I.T. tramite simulazione

Purtroppo, fatti salvi alcuni semplici casi, non si conoscono le distribuzioni delle statistiche test U.I.T., allora si è pensato, come per il test T.R.V., di ricorrere all'uso della simulazione.

L'idea di base non è più quella di stimare il p-value della statistica test, ma consiste nel determinare il percentile d'interesse della distribuzione del test stimata tramite simulazione.

Come nel caso T.R.V., l'idea è semplice:

- Per prima cosa, si va a simulare N volte il test U.I.T. e si inseriscono in un vettore tutti i valori ottenuti.

- Il numero N , che ci determina il numero di cicli della simulazione, è definito allo stesso modo del caso T.R.V., cioè:

$$N = (2.57/\text{errore})^2 * V(\hat{p}) \frac{1}{2}$$

- Anche in questo caso si va a impostare di default un errore assoluto $d=5e-3$ e $p=5e-2$, che comportano un numero di simulazioni pari a 12549. Il valore p , è di fondamentale importanza, perché oltre a determinare il numero di simulazioni, determina anche il valore del percentile d'interesse.
- Dopo aver inserito tutti i valori dei test in un vettore, li si ordina in ordine crescente e si prende come valore d'interesse il percentile $1-p$.
- Infine, si utilizza il percentile $1-p$, come valore soglia. Infatti, se il test calcolato nel campione effettivamente osservato è più piccolo del percentile d'interesse, allora si accetta l'ipotesi H_0 , altrimenti la si rifiuta.

Anche in questo caso, l'utente ha la facoltà di modificare il valore di p e dell'errore.

Vale la pena di sottolineare che questa procedura è valida per il test scalare, nel caso non lo fosse, occorrerebbe eseguire delle semplici modifiche alla procedura.

CAPITOLO 3

LA FUNZIONE TEST

3.1 Introduzione al compilatore R

La vasta gamma di programmi disponibili per l'elaborazione dei dati, pone il problema di sceglierne uno. Per la selezione dei candidati ci si è basati sulle seguenti considerazioni: Il programma deve:

- prevedere l'accettazione dei dati presenti in un file codificato secondo lo standard ASCII;
- prevedere, se non tutte, almeno gran parte delle tecniche statistiche presenti nella guida automatizzata.

Si è deciso di usare il pacchetto R per la sua facilità d'utilizzo.

R è un ambiente integrato che permette di elaborare dati, eseguire calcoli ed effettuare rappresentazioni grafiche. Si parla di ambiente e non di software perché è un linguaggio object-oriented, che può essere facilmente esteso dall'utente finale.

Nasce come estensione del linguaggio di programmazione S sviluppato nei Laboratori Bell americani dell'AT&T.

A differenza di altri ambienti statistici, R è di pubblico dominio (open-source) che, nel rispetto dei diritti d'autore, può essere copiato, distribuito e modificato da chiunque. Grazie a questa sua flessibilità, molti utenti esperti sono stimolati a svilupparlo e aggiornarlo con frequenza.

3.2 Algoritmo, che stima la distribuzione del test T.R.V.

Dopo aver presentato, nel capitolo 1, la teoria per ottenere la stima della distribuzione dei test T.R.V., si passa ora alla realizzazione tramite PC.

Le funzioni, che via simulazione ci portano a costruire i test T.R.V., si fondano tutte sulla stessa idea di base, cioè quella di calcolare il p-value, associato al test calcolato sui dati del campione effettivamente osservato, al fine di lasciare maggior libertà all'utente nel prendere una decisione.

L'algoritmo che ci permette questo è il seguente:

```
#calcola la statistica test sui dati "originali"
t<-trv(dati,...)

#determina il numero di cicli della simulazione
N<-round((2.57/errore)^2*alpha*(1-alpha))

#Ciclo che mi identifica il numero di volte che ripeto la simulazione
for(i in 1:N)
{
dati<-NULL
#simula n dati con distribuzione normale multivariata  $N_p(\text{media},\text{sigma})$ 
dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)

#calcola la statistica test sui dati "simulati"
x<-trv(dati0,...)

#Controlla se x > valore di soglia ( $t_0$ )
if(x>t)
{
```

```

        #Conta le volte che x > valore di soglia (t0)
        rif<-rif+1
    }
}
#calcola il p-value
pvalue<-rif/N

```

Questo breve e semplice algoritmo, è la base che sta sotto ogni funzione che simula la distribuzione dei test T.R.V.

3.3 Algoritmo, che stima la distribuzione del test U.I.T.

In questo caso a differenza dell'algoritmo precedente, basato sulla stima del p-value, ci si basa nel calcolo del percentile d'interesse della distribuzione del test U.I.T.

Si tratta di una scelta obbligata, maturata dalla struttura dell'U.I.T..

L'algoritmo è il seguente:

```

#determina il numero di cicli della simulazione
N<-round( ( 2.57/errore ) ^2*alpha*( 1-alpha ) )

#Ciclo che mi identifica il numero di volte che ripeto la simulazione
for(i in 1:N)
{
dati0<-NULL
#simula n dati con distribuzione normale multivariata Np(media,sigma)
dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)

```

```
#calcola la statistica test sui dati "simulati"
k<-uit(dati0,...)

#crea un vettore con tutti i valori del test U.I.T.
k1<-rbind(k1,k)
}
#ordina i valori in modo crescente
x1<-sort(k1)

#determina il percentile d'interesse della statistica U.I.T.
qmax<-quantile(x1,1-alpha)
```

Le istruzioni appena descritte sono la base che sta sotto ogni funzione che simula la distribuzione dei test U.I.T.

3.4 Tempi di esecuzione delle simulazioni

Lo scopo di questo lavoro consisteva nella progettazione e realizzazione di un programma che, data una matrice normale di dati $X_{n \times p}$, mi calcolasse il test per la verifica d'ipotesi d'interesse e che quando non fosse nota la vera distribuzione del test, lasciasse piena libertà all'utente nell'interpretazione dei risultati ottenuti.

Poiché le scelte fatte sono tutte opinabili, si è preferito realizzare un programma facile da comprendere, modificare e in grado di raggiungere dei risultati sufficientemente precisi in tempo reale.

Per noi, dunque, l'efficienza del programma è correlata al tempo di esecuzione, che dipende dalla macchina sulla quale il programma viene eseguito, dal particolare compilatore R usato e dal grado di approssimazione dei risultati.

Si sottolinea che le approssimazioni dei risultati dipendono dal numero N di replicazioni usate nella simulazione, che a livello di principio possono essere aumentate fino a quanto si vuole.

Per ridurre il tempo d'attesa, si è deciso di perdere nell'accuratezza dei risultati, utilizzando un numero ragionevole di replicazioni .

Per comprendere meglio faremo un esempio.

Consideriamo una matrice normale di dati $X_{n \times p}$, di dimensione:

```
dim(dati)
[1] 100  8
```

Si vuole verificare l'ipotesi:

- $MH_0: \Sigma = \Sigma_0 \rightarrow$ con μ non nota
- $MH_1: \Sigma \neq \Sigma_0$

Si suppone di usare la funzione **s.TRV3**.

Per vedere il tempo di esecuzione della funzione, utilizziamo l'istruzione `system.time`.

```
system.time(S.TRV3(dati,sigma0))
[1] "Risultati della simulazione"
[1] "*****"
[1] 23.06 0.00 23.19 NA NA
```

Il tempo di esecuzione è di 23.06 secondi, eseguendo la simulazione con un pentium4 (2.5GHz e 512 MB di Ram).

Nell'esempio fatto, si è utilizzato un errore assoluto $d=5e-3$ e $p=5e-2$, che equivale a $N=12549$ replicazioni.

Se volessimo usare $p=0.5$, cioè di massima variabilità, il numero di replicazioni aumenterebbe a 66049, che comporterebbero un tempo di esecuzione troppo oneroso, nell'ordine di qualche minuto. Vale la pena di sottolineare che i risultati così ottenuti sarebbero però migliori rispetto a quelli precedenti.

Resta da dire che una volta che l'utente ha deciso quali valori usare per d e p , lasciamo piena libertà nell'interpretazione dei risultati.

L'incremento della velocità, dovuto sia a migliori programmi che a macchine più avanzate, accrescerà negli anni futuri la nostra capacità di trattare questi problemi. Oggi possiamo solamente immaginare la frontiera di ciò che è possibile fare con i calcolatori.

CAPITOLO 4

UN'APPLICAZIONE PRATICA

4.1 Interfaccia della funzione test

Prendiamo in esame il data-set “iris3” presente in R. Questo data-set contiene le misurazioni in centimetri delle variabili peso e grandezza dei sepali e dei petali, di 150 fiori provenienti da 3 specie diverse. Le specie sono setosa, versicolor e virginica.

Dopo averlo opportunamente caricato, suddividiamo le popolazioni oggetto di studio:

```
data(iris3)
setosa<-iris3[, ,1]
versic<-iris3[, ,2]
virginica<-iris3[, ,3]
```

Per effettuare l'analisi dei dati, dobbiamo caricare il “pacchetto” **TRV** contenente le routine che ci servono. Procediamo, quindi, con la presentazione del “pacchetto” TRV che abbiamo sviluppato. Il “pacchetto” oggetto di studio, permette all'utente, di richiamare tutti i possibili test rapporto verosimiglianza, digitando la seguente istruzione.

```
test('trv')
[1] "                TEST T.R.V.                "
[1] "Ipotesi: mu=m0, sigma nota                ==> TRV1(dati,m0,sigma)"
[1] "Ipotesi: mu=m0, sigma non nota            ==> TRV2(dati,m0)"
[1] "Ipotesi: sigma=sigma0, mu non nota        ##> TRV3()?"
[1] "Ipotesi: R*mu=r, mu non nota              ==> TRV4(dati,sigma,R,r)"
[1] "Ipotesi: R*mu=r, sigma non nota           ==> TRV5(dati,R,r)"
[1] "Ipotesi: sigma=k*sigma0, mu e k non note  ##> TRV6()?"
[1] "Ipotesi: sigma12=0, mu non nota           ##> TRV72()?"
```

```

[1] "Ipotesi: sigma=I, mu non nota          ##> TRV7()?"
[1] "Ipotesi: mul=mu2, con signal=sigma2    ==> TRV82(dati,num)"
[1] "Ipotesi: mul=.=mug, con signal=.=sigmag ##> TRV8()?"
[1] "Ipotesi: signal=.=sigmag, test di Box  ##> TRV9()?"
[1] "Ipotesi: mul=.=mug, e signal=.=sigmag  ##> TRV10()?"
[1] "-----"
[1] "##> indica distribuzione asintotica    ?indica scelta dell'utente"

```

Come si può vedere a video, oltre a comparire tutti i test rapporto verosimiglianza eseguibili con **TRV**, compaiono anche le varie funzioni da digitare, con i relativi parametri.

Inoltre possiamo notare alla fine della schermata una legenda, la quale mi indica se la distribuzione del test è asintotica (`##> indica distribuzione asintotica`) e i test per i quali c'è possibilità di scelta da parte dell'utente (`?indica scelta dell'utente`).

La stessa cosa è stata pensata per i test unione-intersezione. Infatti digitando l'istruzione `"trv('uit')"`, otteniamo a video:

```

test('uit')
[1] "                TEST U.I.T. GLOBALE                "
[1] "Ipotesi: mu=m0, sigma nota          ==> UIT1(dati,m0,sigma,alpha)"
[1] "Ipotesi: mu=m0, sigma non nota      ==> UIT2(dati,m0,alpha)"
[1] "Ipotesi: sigma=sigma0, mu non nota  NN> UIT3(dati,sigma0,alpha,errore)"
[1] "Ipotesi: R*mu=r, mu non nota        ==> UIT4(dati,sigma,R,r,alpha)"
[1] "Ipotesi: R*mu=r, sigma non nota     ==> UIT5(dati,R,r,alpha)"
[1] "Ipotesi: signal2=0, mu non nota     NN> UIT72(dati,p1,p2,alpha,errore)"
[1] "Ipotesi: mul=mu2, con signal=sigma2 ==> UIT82(dati,num,alpha)"
[1] "Ipotesi: mul=.=mug, con signal=.=sigmag NN> UIT8(dati,num,alpha,errore)"
[1] "-----"
[1] "##> non nota la distribuzione                "

```

Nei prossimi due paragrafi, faremo degli esempi di alcuni test sul data-set "iri3".

4.2 Applicazione dei test T.R.V. sul data-set “iris3”

4.2.1 Ipotesi che $\mu = \mu_0$, con Σ nota

“TRV1”

Consideriamo solo la popolazione setosa, con tutte le sue variabili (peso e grandezza dei petali e dei sepali).

Supponiamo che i dati si distribuiscano come $N(\mu, \Sigma)$ con Σ nota.

Dato un vettore medie $\mu_0 = (5, 3, 1.4, 0.2)$, vogliamo verificare l'ipotesi:

- $MH_0: \mu = \mu_0 \quad \rightarrow$ con Σ nota
- $MH_1: \mu \neq \mu_0$

Il test è:

$$TRV1 = LR = n (\bar{\mu} - \mu_0)' \Sigma^{-1} (\bar{\mu} - \mu_0)$$

sotto $MH_0: TRV1 \sim \chi_p^2$

Per effettuare tale test ci siamo costruiti una funzione apposita, “TRV1”.

La funzione dopo aver calcolato il test, stamperà a video:

- il vettore delle medie campionarie
- il vettore μ_0
- il valore del test con i relativi gradi di libertà (gl)
- il p-value del test

Funzione: **TRV1**

Parametri:

dati: matrice di dati

m0: vettore μ_0

sigma: matrice di varianza e covarianza nota

Applicazione:

```
dati<-iris3[, ,1] → dati=setosa  
m0<-c(5,3.4,1.4,0.24) → vettore ipotesi  $\mu_{H0}$ 
```

```
sigma → matrice var. cov. nota
```

	Sepal L.	Sepal W.	Petal L.	Petal W.
Sepal L.	0.12	0.10	0.02	0.01
Sepal W.	0.10	0.15	0.01	0.01
Petal L.	0.02	0.01	0.03	0.01
Petal W.	0.01	0.01	0.01	0.03

Adesso tramite la funzione TRV1, eseguo il test :

```
TRV1(dati,m0,sigma)  
[1] "Risultati del test"  
[1] "*****"  
$"medie campionarie"  
Sepal L. Sepal W. Petal L. Petal W.  
5.006 3.428 1.462 0.246  
  
$"m0"  
[1] 5.00 3.40 1.40 0.24  
  
$test  
  
[ ,1]  
[1,] 8.589234  
  
$gl  
[1] 4
```

\$"p-value"

```
      [,1]  
[1,] 0.0722281
```

Si può accettare l'ipotesi nulla che $\mu = \mu_0$.

4.2.2 Ipotesi che $\mu = \mu_0$, con Σ non nota "TRV2"

Con gli stessi dati, verifichiamo l'ipotesi del caso precedente, però con Σ non nota.

In questo caso il test T.R.V. è:

$$\text{TRV2} = \text{LR} = (n-p)/p (\bar{\mu} - \mu_0)' S^{-1} (\bar{\mu} - \mu_0)$$

sotto MH_0 : $\text{TRV2} \sim F_{p,p-n}$

Funzione: TRV2

Parametri:

dati: matrice di dati

m0: vettore μ_0

Applicazione:

TRV2(dati,m0)

```
[1] "Risultati del test"
```

```
[1] "-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*"
```

```
 $"medie campionarie"
```

```
Sepal L. Sepal W. Petal L. Petal W.
      5.006      3.428      1.462      0.246
```

```
 $"m0"
 [1] 5.00 3.40 1.40 0.24
```

```
 $test
```

```
           [,1]
 [1,] 1.733327
```

```
 $g11
 [1] 4
```

```
 $g12
 [1] 46
```

```
 $"p-value"
```

```
           [,1]
 [1,] 0.1588863
```

Accettiamo che il vettore delle medie assegnato è uguale alle medie dei dati.

4.2.3 Ipotesi che $\Sigma = \Sigma_0$, con μ non nota “TRV3”

Sempre con gli stessi dati, proviamo a verificare l'ipotesi che $\Sigma = \Sigma_0$, con Σ_0 pari a:

Σ_0

	Sepal L.	Sepal W.	Petal L.	Petal W.
Sepal L.	0.12	0.10	0.02	0.01
Sepal W.	0.10	0.15	0.01	0.01
Petal L.	0.02	0.01	0.03	0.01
Petal W.	0.01	0.01	0.01	0.03

L'ipotesi risulta essere:

- $MH_0: \Sigma = \Sigma_0 \rightarrow$ con μ non nota
- $MH_1: \Sigma \neq \Sigma_0$

Il test per la verifica d'ipotesi in questione è:

$$\mathbf{TRV3} = -2\log \lambda = np (\mathbf{a} - \log(\mathbf{g}) - 1)$$

dove:

\mathbf{g} è la media geometrica degli autovalori di $\Sigma_0^{-1} \mathbf{S}$

\mathbf{a} è la media aritmetica degli autovalori di $\Sigma_0^{-1} \mathbf{S}$

sotto MH_0 : $\mathbf{TRV3} \xrightarrow{d} \chi_{p(p+1)/2}^2$ per $n \rightarrow \infty$

Questo, è uno di quei casi in cui non si conosce la distribuzione esatta e quindi possiamo ricorrere ad uno dei cinque metodi già descritti nel capitolo 1. La funzione “**TRV3**”, stamperà a video le 5 opzioni possibili, lasciando all'utente la facoltà di sceglierne una.

TRV3()

```
[1] "Scegliere una delle 5 opzioni"
[1] "-----"
[1] "Bartlett analitico          ##> B.a.TRV3(dati,sigma0,alpha,errore) "
[1] "Bartlett empirico1(media)  ##> B.e1.TRV3(dati,sigma0,a1,alpha,errore) "
[1] "Bartlett empirico2(media e var.) ##> B.e2.TRV3(dati,sigma0,a2,c,alpha,errore) "
[1] "Korin(approssimazione con la F) ##> K.TRV3(dati,sigma0,alpha,errore) "
[1] "Simulazione                ##> S.TRV3(dati,sigma0,alpha,errore) "
```

Noi ci limitiamo a utilizzare il metodo della simulazione.

La funzione usata è **s.TRV3**, dopo aver calcolato il test, stamperà a video il valore del test con i relativi gradi di libertà, il p-value stimato con il suo intervallo di confidenza e infine il numero di replicazioni eseguite per la simulazione.

Funzione: **s.TRV3**

Parametri:

dati: matrice di dati

sigma0: matrice di varianza e covarianza dell'ipotesi MH_0

alpha: parametro che determina il livello di significatività del test (impostato di default pari a 0.05)

errore: parametro che determina l'ampiezza dell'intervallo del p-value stimato (impostato di default pari a $5e-3$)

Applicazione:

```
s.TRV3(dati,sigma0)
[1] "Risultati della simulazione"
[1] "-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*"
$test
[1] 22.76146

$gl
[1] 10

$"p-value"
[1] 0.01800940

$"intervallo di confidenza del p-value"
[1] 0.01495848 0.02106033

$Nreplicazioni
[1] 12549
```

Sulla base del p-value stimato, lasciamo piena libertà all'utente nell'interpretazione del risultato.

4.2.4 Ipotesi che $\Sigma_1 = \dots = \Sigma_g$ (test di Box) “TRV9”

Consideriamo le specie setosa e versicolor, si vuole verificare l'ipotesi:

- $MH_0: \Sigma_{\text{setosa}} = \Sigma_{\text{versicolor}}$
- $MH_1: \Sigma_{\text{setosa}} \neq \Sigma_{\text{versicolor}}$

La verifica dell'ipotesi in questione si realizza ricorrendo al test:

$$\text{TRV9} = \text{LR} = \sum_{i=1}^g n_i \log | \mathbf{S}_i^{-1} \mathbf{S} | = -n \log | \mathbf{S} | - \sum_{i=1}^g n_i \log | \mathbf{S}_i |$$

sotto MH_0 : $\text{TRV9} \xrightarrow{d} \chi_{p(p+1)(g-1)/2}^2$ per $n \rightarrow \infty$

Poiché non si conosce la distribuzione esatta, tramite la funzione “TRV9” lasciamo all'utente la facoltà di scegliere tra le 4 opzioni possibili.

TRV9()

```
[1] "Scegliere una delle 4 opzioni"
[1] "-----"
[1] "Bartlett analitico           ==> B.a.TRV9(dati,num)"
[1] "Bartlett empirico1(media)   ==> B.e1.TRV9(dati,num,a1,alpha,errore)"
[1] "Bartlett empirico2(media e varianza) ==> B.e2.TRV9(dati,num,a2,c,alpha,errore)"
[1] "Simulazione                 ==> S.TRV9(dati,num,alpha,errore)"
```

Utilizziamo la funzione **B.e2.TRV9**, che dopo aver calcolato il test, stamperà a video il valore del test con i relativi gradi di libertà, il p-value stimato e infine i parametri di correzione a_2 e c .

Funzione: **B.e2.TRV9**

Parametri:

dati: matrice di dati

num: vettore che indica la numerosità dei campioni

a2: parametro di correzione (impostato di default pari a 0)

c: parametro di correzione (impostato di default pari a 0)

alpha: parametro che determina il livello di significatività del test (impostato di default pari a 0.05)

errore: parametro che determina l'ampiezza dell'intervallo del p-value stimato (impostato di default pari a 5e-3)

Applicazione:

```
dati<-rbind(setosa,versic)
```

```
B.e2.TRV9(dati,num)
```

```
[1] "Risultati del test"
```

```
[1] "*****"
```

```
$test
```

```
[1] 54.70822
```

```
$gl
```

```
[1] 10
```

```
 $"p-value"
```

```
[1] 3.580148e-08
```

```
 $"parametri di correzione a2 e c"
```

```
[1] 1.263623 -1.718838
```

Come si può vedere, viene rifiutata l'ipotesi MH_0 .

Adesso, supponiamo di voler verificare l'ipotesi:

- $MH_0: \Sigma_{setosa} = \Sigma_{virginica}$
- $MH_1: \Sigma_{setosa} \neq \Sigma_{virginica}$

La verifica dell'ipotesi in questione, come previsto, si realizza ricorrendo allo stesso test precedente.

In questo caso, per la verifica dell'ipotesi, conviene scegliere la stessa funzione di prima, **B.e2.TRV9**, poiché non cambiando la numerosità e le variabili dei campioni d'interesse, possiamo utilizzare gli stessi parametri di correzione, evitando così l'attesa per il calcolo della loro stima.

Applicazione:

```
dati<-rbind(setosa,virginica)
B.e2.TRV9(dati,num,1.263,-1.718)
[1] "Risultati del test"
[1] "*****"
$test
[1] 90.6408

$gl
[1] 10

$"p-value"
[1] 3.996803e-15

$"parametri di correzione a2 e c"
[1] 1.263 -1.718
```

Anche in questo caso rifiutiamo l'ipotesi MH_0 .

Come verifica, proviamo ad eseguire la funzione senza inserire i parametri a_2 e c .

```
B.e2.TRV9(dati,num)
```

```
[1] "Risultati del test"
```

```
[1] "-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*"
```

```
$test
```

```
[1] 90.71059
```

```
$gl
```

```
[1] 10
```

```
 $"p-value"
```

```
[1] 3.885781e-15
```

```
 $"parametri di correzione a2 e c"
```

```
[1] 1.261340 -1.749727
```

Come si può vedere dai risultati ottenuti, vengono confermate le considerazioni fatte prima.

4.3 Applicazione dei test U.I.T. sul data-set “iris3”

4.3.1 Ipotesi che $\mu=\mu_0$, con Σ nota

“UIT1”

A titolo di esempio, dato un vettore medie $\mu_0=(6.5, 3, 5.5, 2.2)$, vogliamo stabilire se sono uguali a quelle delle variabili della popolazione virginica.

L'ipotesi risulta essere:

- ${}_M H_0: \quad \mu = \mu_0 \quad \rightarrow \text{con } \Sigma \text{ nota}$
- ${}_M H_1: \quad \mu \neq \mu_0$

Il test globale coincide in questo caso con il TRV:

$$\text{UIT1} = \text{LR} = n (\underline{\mu} - \underline{\mu}_0)' \Sigma^{-1} (\underline{\mu} - \underline{\mu}_0)$$

sotto ${}_M H_0: \text{UIT1} \sim \chi_p^2$

Per effettuare tale test ci siamo costruiti una funzione apposita, “UIT1”.

La funzione dopo aver calcolato il test e averlo confrontato con un adeguata distribuzione chi-quadrato, stamperà a video:

- il grafico della densità del test UIT1 , due rette, una rossa che mi indica il percentile $1-\alpha$ stimato della distribuzione del test globale e una blu che indica il valore del test globale calcolato sui dati originali.
- il valore del test U.I.T. globale.
- il p-value del test .
- il valore critico della distribuzione del test, relativo all' α usato.
- il valore α usato.

- una frase che indica se l'ipotesi globale viene accettata o rifiutata.

Funzione: `UIT1`

Parametri:

dati: matrice di dati

m0: vettore μ_0

sigma: matrice di varianza e covarianza nota

alpha: parametro che determina il livello di significatività del test (impostato di default pari a 0.05)

Applicazione:

```
dati<-virginica
```

```
m0<-c(6.5,3,5.5,2.2) ➔ vettore ipotesi  $\mu_{H0}$ 
```

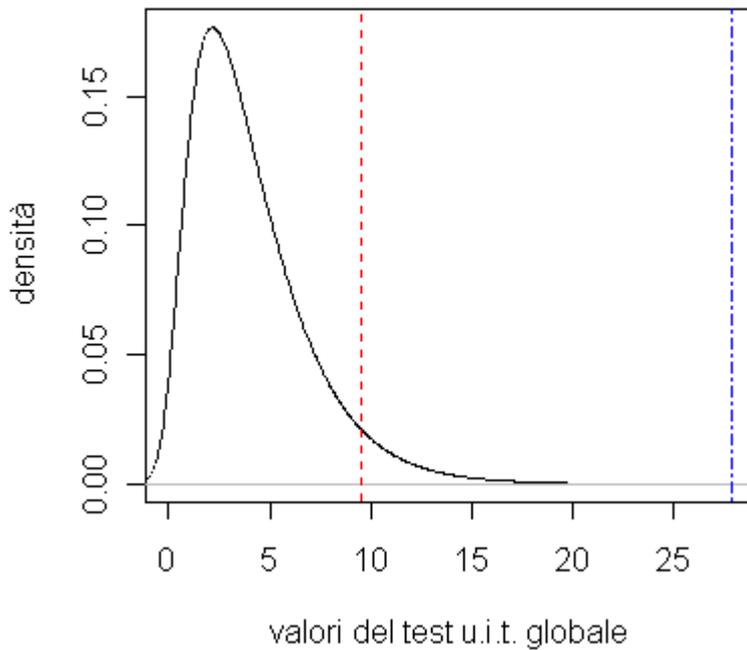
```
sigma<-var(dati)
```

```
> sigma
```

	Sepal L.	Sepal W.	Petal L.	Petal W.
Sepal L.	0.40434286	0.09376327	0.30328980	0.04909388
Sepal W.	0.09376327	0.10400408	0.07137959	0.04762857
Petal L.	0.30328980	0.07137959	0.30458776	0.04882449
Petal W.	0.04909388	0.04762857	0.04882449	0.07543265

UIT1(dati,m0,sigma)

Distribuzione del test U.I.T. globale



Pause. Press <Enter> to continue...

[1] "Risultati del test"

[1] "*****"

\$"Valore del test U.I.T. globale"

[,1]

[1,] 27.79926

\$"p-value"

[,1]

[1,] 1.369754e-05

\$"Valore critico della distribuzione del Test"

[1] 9.487729

\$alpha

[1] 0.05

\$Ipotesi

[1] "Rifiutiamo l'ipotesi globale"

[2] "Si può verificare se una qualche <ipotesi componente> abbia"

[3] "portato al rifiuto con UIT1.a(a) dove a rappresenta il"

[4] "vettore p-dimensionale di un'ipotesi componente"

Se, come in questo caso l'ipotesi MH_0 è rifiutata, allora, come viene riportato nell'output, si utilizza `UIT1.a(a)`.

Questa funzione calcola il test per la generica componente nella forma:

$$\text{UIT1.a} = \text{LR} = n \mathbf{a}' (\boldsymbol{\mu} - \boldsymbol{\mu}_0)' (\boldsymbol{\mu} - \boldsymbol{\mu}_0) \mathbf{a} / \mathbf{a}' \boldsymbol{\Sigma}^{-1} \mathbf{a}$$

sotto MH_0 : $\text{UIT1.a} \sim \chi_p^2$

Funzione: `UIT1.a`

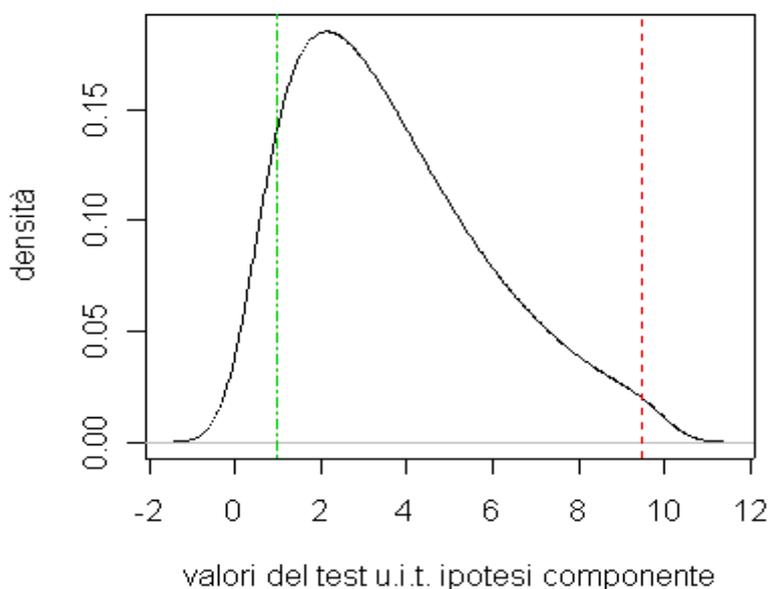
Parametri:

a: vettore p-dimensionale di un'ipotesi componente

Applicazione:

`a=c(1,0,0,0)` → consideriamo solo la prima variabile
`UIT1.a(a)`

Distribuzione del test U.I.T.



Pause. Press <Enter> to continue...

[1] "Risultati del test"

```

[1] "*****"
$"Vettore a:"
[1] 1 0 0 0

$"Valore del test U.I.T. ipotesi componente
      [,1]
[1,] 0.9576032

$"p-value"
      [,1]
[1,] 0.9161551

$"Valore critico della distribuzione del Test"
[1] 9.487729

$alpha
[1] 0.05

$Ipotesi
[1] "Accettiamo l'ipotesi componente"

```

Accettiamo l'ipotesi componente, cioè l'ipotesi che la media della variabile Sepal L. è uguale a 6.5 (primo elemento del vettore m_0). Questo lo si vede anche nel grafico, nel quale la retta verde che indica il valore del test dell'ipotesi componente sui dati originali (virginica), è a sinistra, quindi inferiore al 95 percentile della distribuzione del test.

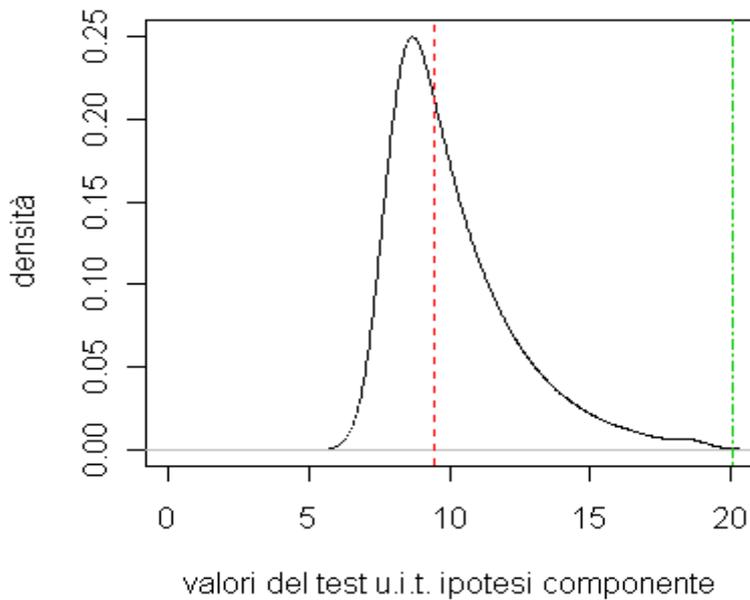
La stessa cosa viene eseguita anche per le altre componenti e si vedrà che un'ipotesi componente che ha portato al rifiuto il test globale è la media della variabile Petal W che è diversa da 2.2.

```

a<-c(0,0,0,1)
UIT1.a(a)

```

Distribuzione del test U.I.T. componente



Pause. Press <Enter> to continue...

```
[1] "Risultati del test"
```

```
[1] "-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*"
```

```
 $"Vettore a:"
```

```
[1] 0 0 0 1
```

```
 $"Valore del test U.I.T. ipotesi componente"
```

```
      [,1]
```

```
[1,] 20.06823
```

```
 $"p-value"
```

```
      [,1]
```

```
[1,] 0.0004841459
```

```
 $"Valore critico della distribuzione del Test"
```

```
[1] 9.487729
```

```
 $alpha
```

```
[1] 0.05
```

```
 $Ipotesi
```

```
[1] "Rifiutiamo l'ipotesi componente"
```

4.3.2 Ipotesi che $\Sigma = \Sigma_0$, con μ non nota

“UIT3”

Consideriamo la popolazione setosa, vogliamo verificare l'ipotesi:

- ${}_M H_0: \Sigma = \Sigma_0 \rightarrow$ con μ non nota
- ${}_M H_1: \Sigma \neq \Sigma_0$

Il test globale (UIT3) in questo caso è la coppia di autovalori:

$$\lambda_1(\Sigma^{-1}_0 \mathbf{S}) \quad \lambda_p(\Sigma^{-1}_0 \mathbf{S})$$

dove rifiutiamo ${}_M H_0$ se $\lambda_1(\Sigma^{-1}_0 \mathbf{S}) > c_{1a}$ o $\lambda_p(\Sigma^{-1}_0 \mathbf{S}) < c_{2a}$

$$c_{1a} = \max_a \frac{a' S a}{a' \Sigma_0 a} = \lambda_1(\Sigma^{-1}_0 \mathbf{S}) \quad c_{2a} = \min_a \frac{a' S a}{a' \Sigma_0 a} = \lambda_p(\Sigma^{-1}_0 \mathbf{S})$$

In questo caso, non conoscendo la distribuzione del test UIT3, si ricorre alla funzione “UIT3”, che ci permette di calcolare i percentili d'interesse della distribuzione del test tramite la simulazione.

La funzione dopo aver calcolato il test stamperà a video:

- il grafico della densità stimata tramite simulazione del test UIT3, quattro rette, due rosse che mi indicano i percentili $\alpha/2$ e $1-\alpha/2$ stimati dalla distribuzione del test globale e due blu che indicano i valori del test globale calcolato sui dati originali.
- il valori del test U.I.T. globale.
- il valori critici della distribuzione del test, relativo all' α usato.
- il valore α usato.
- una frase che indica se l'ipotesi globale viene accettata o rifiutata.

Funzione: **UIT3**

Parametri:

dati: matrice di dati

sigma0: matrice di varianza e covarianza dell'ipotesi MH_0

alpha: parametro che determina il livello di significatività del test (impostato di default pari a 0.05)

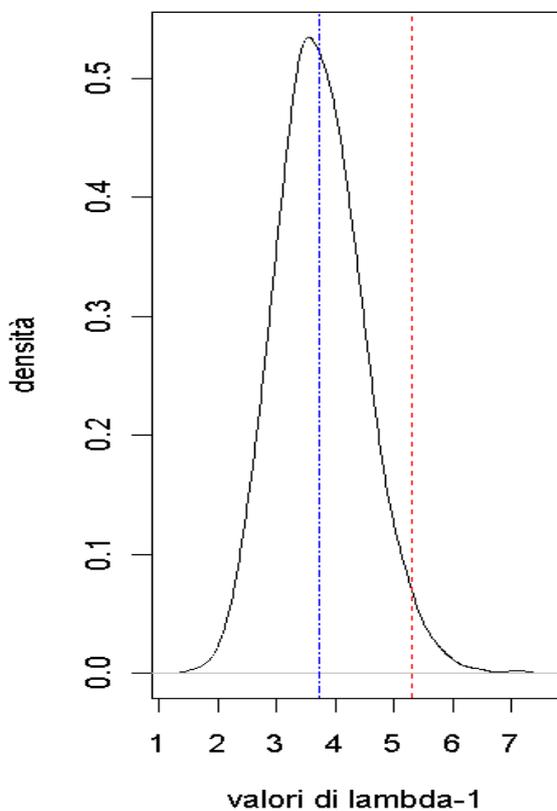
errore: parametro che determina l'ampiezza dell'intervallo del p-value stimato (impostato di default pari a $5e-3$)

Applicazione:

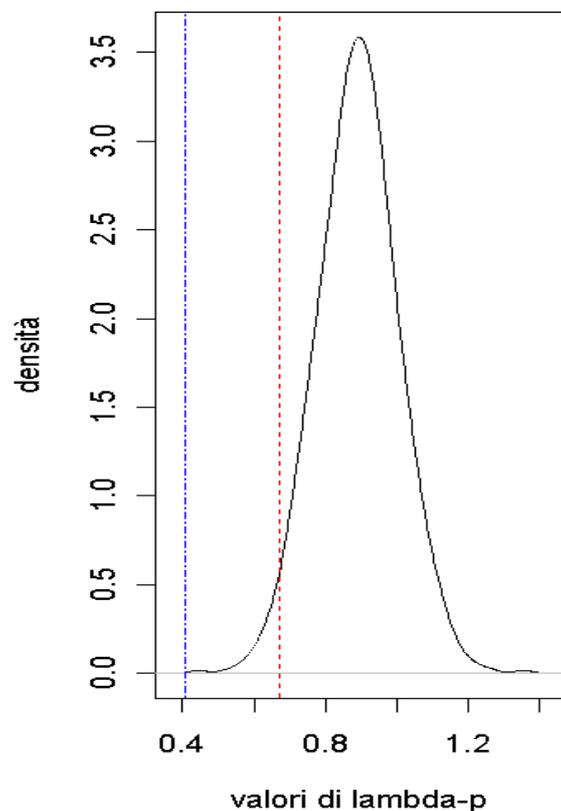
```
dati<-setosa
```

```
UIT3(dati,sigma0)
```

Distribuzione lambda-1



Distribuzione lambda-p



```

Pause. Press <Enter> to continue...
[1] "Risultati della simulazione"
[1] "*****"
$"Valori del test U.I.T. globale"
[1] 0.4063127 3.7182478

$"Valori critici della distribuzione del Test"
      2.5%      97.5%
0.6698843 5.3118419

$alpha.effettivo
[1] 0.05031875

$Ipotesi
[1] "Rifiutiamo l'ipotesi globale"
[2] "Si può verificare se una qualche <ipotesi componente> abbia"
[3] "portato al rifiuto con UIT3.a(a) dove a rappresenta il"
[4] "vettore p-dimensionale di un'ipotesi componente"

$Nreplicazioni
[1] 12549

```

Il test in questione è più difficile da gestire e diversamente da quanto avviene per gli altri esempi, calcoliamo l'alpha.effettivo facendo la somma dell'alpha nominale più il numero delle volte che i due autovalori calcolati per ogni simulazione, escono contemporaneamente dalle zone di accettazione visibili dai grafici e quindi ci portano al rifiuto del test globale. Si tratta di una scelta obbligata, maturata dalla struttura del test UIT3.

Se, come in questo caso l'ipotesi MH_0 è rifiutata, allora, come viene riportato nell'output, si utilizza **UIT3.a(a)**.

Questa funzione calcola il test per la generica componente nella forma:

$$\mathbf{UIT3.a} = \mathbf{LR} = \mathbf{a' S a} / \mathbf{a' \Sigma_0 a}$$

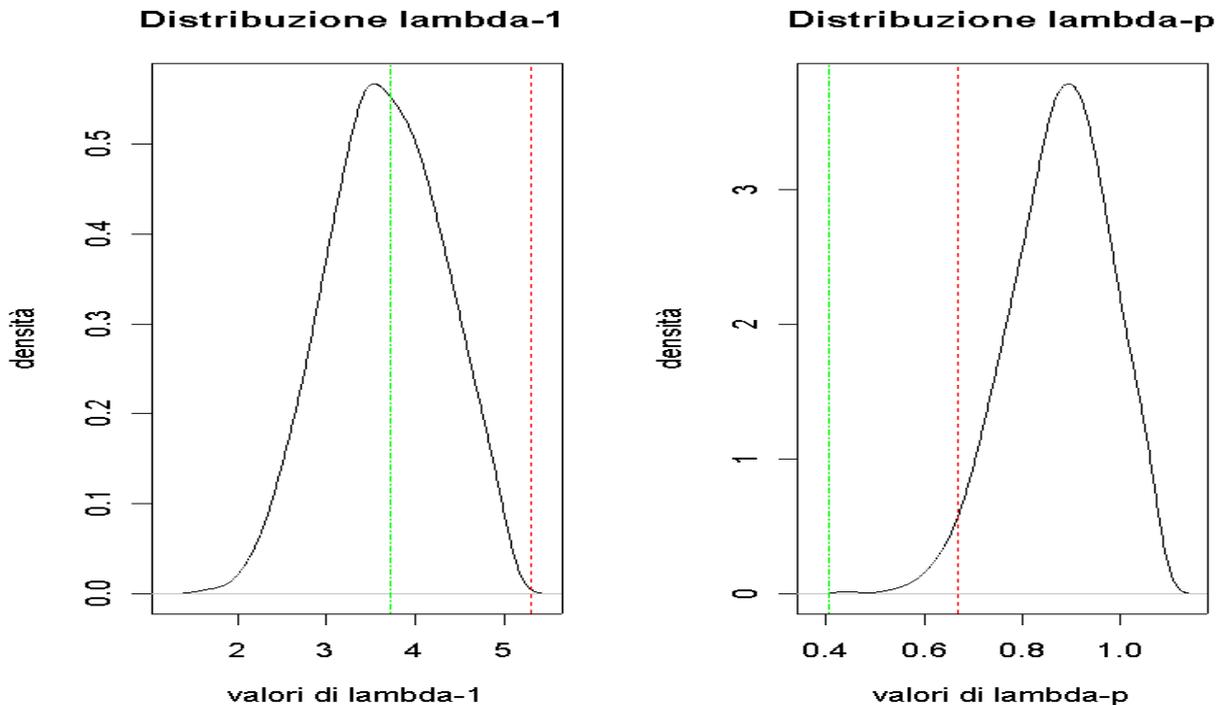
Funzione: **UIT3.a**

Parametri:

a: vettore p-dimensionale di un'ipotesi componente

Applicazione:

`a<-c(1,0,0,0)` → consideriamo solo la prima variabile
`UIT3.a(a)`



```
Pause. Press <Enter> to continue...
[1] "Risultati della simulazione"
[1] "*****"
$"Vettore a:"
[1] 1 0 0 0

$"Valore del test U.I.T. componente"
[1] 1.0147

$"Valori critici della distribuzione del Test"
      2.5%      97.5%
0.6670678 5.3009577

$alpha
[1] 0.05

$Ipotesi
[1] "Accettiamo l'ipotesi componente"
```

Accettiamo l'ipotesi componente.

La stessa cosa viene eseguita anche per le altre componenti.

Conclusioni

Il lavoro svolto in questa tesi consisteva nella progettazione e realizzazione di un programma in grado di raggiungere dei risultati sufficientemente precisi in tempo reale; questo lo si è fatto con la realizzazione di alcune procedure che risolvono ognuna uno specifico problema di verifica d'ipotesi multidimensionale. Di tali procedure si è già parlato nel corso della tesi.

Credo che la cosa principale messa in luce da questa tesi è che la simulazione può facilitare molto l'applicazione dell'inferenza asintotica, tuttavia non può eliminare ogni difficoltà.

Tutt'altro: va sottolineato che particolarizzare i metodi asintotici mediante dei programmi come quelli presentati nella tesi, rimane un compito non semplice, mediamente laborioso, attuabile solo da chi abbia una certa pratica sia di metodi asintotici che di pacchetti di calcolo statistico, oltre che di una buona dose di pazienza.

L'invito è allora a non considerare la simulazione come un mezzo per eliminare totalmente l'applicazione delle distribuzioni asintotiche, ma occorre tener conto che la sua applicazione in problemi effettivi non va considerata automatica, ma può presentare difficoltà di varia natura (a livello di sviluppo del programma, o dell'utilizzo stesso da parte di un'utente).

I metodi di simulazione presentati in questa tesi sembrano poter costituire uno strumento efficiente in alcuni problemi inferenziali, e gli esempi presentati nel CAPITOLO 4, nei loro limiti, sembrano complessivamente confermare tale impressione.

Per concludere, giacché il pacchetto è stato costruito in modo modulare, esso può essere ampliato nelle sue funzionalità con l'inserimento di nuove parti che realizzino ulteriori algoritmi, qualora se ne presenti la necessità.

Appendice: listato funzioni in R

Questo appendice contiene gli algoritmi utilizzati per la realizzazione degli esempi del CAPITOLO 4. Parte di questi sono già stati presentati nella tesi.

All'inizio di ognuno sono state poste alcune righe di commento nelle quali viene spiegata la funzione del particolare algoritmo considerato.

Qui sotto viene data la lista degli algoritmi presenti in appendice.

NOME FUNZIONE	UTILIZZO
Test Rapporto di verosimiglianza (T.R.V.)	
TRV1	Test sull'ipotesi che $\mu=\mu_0$, con Σ nota
TRV2	Test sull'ipotesi che $\mu=\mu_0$, con Σ non nota
TRV3	Test sull'ipotesi che $\Sigma=\Sigma_0$, con μ non nota
TRV4	Test sull'ipotesi che $R^*\mu=r$, con μ non nota
TRV5	Test sull'ipotesi che $R^*\mu=r$, con Σ non nota
TRV6	Test sull'ipotesi che $\Sigma=k*\Sigma_0$, con μ e k non note
TRV72	Test sull'ipotesi che $\Sigma_{12}=0$, con μ non nota
TRV7	Test sull'ipotesi che $\Sigma=I$, con μ non nota
TRV82	Test sull'ipotesi che $\mu_1=\mu_2$, con $\Sigma_1=\Sigma_2$
TRV8	Test sull'ipotesi che $\mu_1=\dots=\mu_g$, con $\Sigma_1=\dots=\Sigma_g$
TRV9	Test sull'ipotesi che $\Sigma_1=\dots=\Sigma_g$, test di Box
TRV10	Test sull'ipotesi che $\mu_1=\dots=\mu_g$ e $\Sigma_1=\dots=\Sigma_g$
Test Unione-Intersezione (U.I.T.)	
UIT1	Test sull'ipotesi che $\mu=\mu_0$, con Σ nota
UIT2	Test sull'ipotesi che $\mu=\mu_0$, con Σ non nota
UIT3	Test sull'ipotesi che $\Sigma=\Sigma_0$, con μ non nota
UIT4	Test sull'ipotesi che $R^*\mu=r$, con μ non nota
UIT5	Test sull'ipotesi che $R^*\mu=r$, con Σ non nota

UIT72	Test sull'ipotesi che $R^*\mu=r$, con Σ non nota
UIT82	Test sull'ipotesi che $\mu_1=\mu_2$, con $\Sigma_1=\Sigma_2$
UIT8	Test sull'ipotesi che $\mu_1=\dots=\mu_g$, con $\Sigma_1=\dots=\Sigma_g$
Analisi della bontà dei metodi	
Bontà3	Analisi della bontà dei metodi sul test TRV3
Bontà72	Analisi della bontà dei metodi sul test TRV72
Bontà7	Analisi della bontà dei metodi sul test TRV7
Bontà9	Analisi della bontà dei metodi sul test TRV9

```
#####
#####TEST#####
#####
```

```
test<-function(q=0)
{
```

```
uit<-as.character(c("uit"))
trv<-as.character(c("trv"))
```

```
if( q==trv )
{
print("                TEST T.R.V.                ")
print("Ipotesi: mu=m0, sigma nota                ==> TRV1(dati,m0,sigma)")
print("Ipotesi: mu=m0, sigma non nota            ==> TRV2(dati,m0)")
print("Ipotesi: sigma=sigma0, mu non nota        ##> TRV3()?")
print("Ipotesi: R*mu=r, mu non nota              ==> TRV4(dati,sigma,R,r)")
print("Ipotesi: R*mu=r, sigma non nota           ==> TRV5(dati,R,r)")
print("Ipotesi: sigma=k*sigma0, mu e k non note   ##> TRV6()?")
print("Ipotesi: sigma2=0, mu non nota             ##> TRV72()?")
print("Ipotesi: sigma=I, mu non nota              ##> TRV7()?")
print("Ipotesi: mul=mu2, con sigmal=sigma2        ==> TRV82(dati,num)")
print("Ipotesi: mul=.=mug, con sigmal=.=sigmag    ##> TRV8()?")
print("Ipotesi: sigmal=.=sigmag, test di Box     ##> TRV9()?")
print("Ipotesi: mul=.=mug, e sigmal=.=sigmag     ##> TRV10()?")
print("-----")
print("##> indica distribuzione asintotica      ? indica scelta dell'utente ")
}
}
```

```
if( q==uit )
{
print("                TEST U.I.T. GLOBALE                ")
print("Ipotesi: mu=m0, sigma nota                ==> UIT1(dati,m0,sigma,alpha)")
print("Ipotesi: mu=m0, sigma non nota            ==> UIT2(dati,m0,alpha)")
print("Ipotesi: sigma=sigma0, mu non nota        NN> UIT3(dati,sigma0,alpha,errore)")
print("Ipotesi: R*mu=r, mu non nota              ==> UIT4(dati,sigma,R,r,alpha)")
print("Ipotesi: R*mu=r, sigma non nota           ==> UIT5(dati,R,r,alpha)")
print("Ipotesi: sigma2=0, mu non nota            NN> UIT72(dati,p1,p2,alpha,errore)")
print("Ipotesi: mul=mu2, con sigmal=sigma2        ==> UIT82(dati,num,alpha)")
print("Ipotesi: mul=.=mug, con sigmal=.=sigmag    NN> UIT8(dati,num,alpha,errore)")
print("-----")
print("##> non nota la distribuzione                ")
}
}
```

```
if(q!=uit & q!=trv) {stop("Inserire 'trv' oppure 'uit'")}
}
```

```
#####
#####TEST RAPPORTO DI VEROSIMIGLIANZA#####
#####
```

```
TRV1<-function(dati,m0,sigma)
{
# mu=m0, sigma nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(length(m0)!=p) {stop("Test valido solo se length(m0)=length(medie)")}
m<-apply(dati,2,FUN=mean)
gl<-p

test<-n*t(m-m0)%*%solve(sigma)%*(m-m0)
pvalue<-1-pchisq(test,gl)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('medie campionarie'=m, 'vettore m0'=m0, test=test,
      gl=gl, 'p-value'=pvalue )
}
```

```
#####
```

```
TRV2<-function(dati,m0)
{
# mu=m0, sigma non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(length(m0)!=p) {stop("Test valido solo se length(m0)=length(medie)")}
m<-apply(dati,2,FUN=mean)
s<-var(dati)*(n-1)/n
gll<-p
gl2<-n-p

test<-(n-p)/p*t(m-m0)%*%solve(s)%*(m-m0)
pvalue<-1-pf(test,gll,gl2)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('medie campionarie'=m, 'vettore m0'=m0, test=test,
      gll=gll, gl2=gl2, 'p-value'=pvalue)
}
```

```
#####
```

```
TRV3<-function(scelta=0)
{
print("Scegliere una delle 5 opzioni")
print("-----")
print("Bartlett analitico          ##> B.a.TRV3(dati,sigma0,alpha,errore)")
print("Bartlett empirico1(media)   ##> B.e1.TRV3(dati,sigma0,a1,alpha,errore)")
print("Bartlett empirico2(media e var.) ##> B.e2.TRV3(dati,sigma0,a2,c,alpha,errore)")
print("Korin(approssimazione con la F) ##> K.TRV3(dati,sigma0,alpha,errore)")
print("Simulazione                   ##> S.TRV3(dati,sigma0,alpha,errore)")
}
```

```

#-----

B.a.TRV3<-function(dati,sigma0)
{
# Sigma=sigma0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}
gl<-p*(p+1)/2
s<-var(dati)*(n-1)/n
ma<-mean(diag(solve(sigma0)%*%s))
mg<-det(solve(sigma0)%*%s)^(1/p)
a<-(1-(2*p+1-2/(p+1))/(6*n))
test<-n*p*(ma-log(mg)-1)*a

pvalue<-1-pchisq(test,gl)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue)
}

#-----

B.e1.TRV3<-function(dati,sigma0,a1=0,alpha=0.05,errore=5e-3)
{
# Sigma=sigma0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}
gl<-p*(p+1)/2
inv.sigma0<-solve(sigma0)

if(a1==0)
{
h<-empirico.trv3(dati,alpha,errore)
a1<-h[1]
t<-trv3(dati,inv.sigma0)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

else
{
t<-trv3(dati,inv.sigma0)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametro di correzione a1'=a1)
}

```

```

#-----

B.e2.TRV3<-function(dati,sigma0,a2=0,c=0,alpha=0.05,errore=5e-3)
{
# Sigma=sigma0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}
gl<-p*(p+1)/2
inv.sigma0<-solve(sigma0)

if(a2==0 & c==0)
{
h<-empirico.trv3(dati,alpha,errore)
a2<-h[2]
c<-h[3]
q<-c(a2,c)
t<-trv3(dati,inv.sigma0)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

else
{
q<-c(a2,c)
t<-trv3(dati,inv.sigma0)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametri di correzione a2 e c'=q)
}

#-----

K.TRV3<-function(dati,sigma0)
{
# Sigma=sigma0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}
s<-var(dati)*(n-1)/n
ma<-mean(diag(solve(sigma0)%*%s))
mg<-det(solve(sigma0)%*%s)^(1/p)
v<-0.5*p*(p+1)
D1<-(2*p+1-2/(p+1))/(6*n)
D2<-(p-1)*(p+2)/(6*n^2)
w<-D2-D1^2
gl1<-round(v)
gl2<-round((v+2)/(w))
b<-v/(1-D1-v/gl2)

test<-n*p*(ma-log(mg)-1)/(b)
pvalue<-1-pf(test,gl1,gl2)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
}

```

```

list(test=test, gl1=gl1, gl2=gl2, 'p-value'=pvalue)
}

#-----

S.TRV3<-function(dati,sigma0,alpha=0.05,errore=5e-3)
{
# Sigma=sigma0, mu non nota

pvalue<-0
rif<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}
gl<-p*(p+1)/2
inv.sigma0<-solve(sigma0)
s<-var(dati)*(n-1)/n
media<-c(rep(0,p))

t<-trv3(dati,inv.sigma0)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for(i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,sigma0,tol=1e-6,empirical=FALSE)
x<-trv3(dati0,inv.sigma0)

#Controllo se il valore appartiene o meno alla zona di rifiuto
if(x>t)
{
#Conto le osservazioni che cadono nella zona di rifiuto
rif<-rif+1
}
}

pvalue<-rif/N
errstimato<-((pvalue*(1-pvalue))/N)^(1/2)
er99<-2.57*errstimato
intervallo<-c(pvalue-er99,pvalue+er99)

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=t, gl=gl, 'p-value'=pvalue, 'intervallo di confidenza del p-value'=intervallo,
Nreplicazioni=N)
}

#-----

trv3<-function(dati,inv.sigma0)
{

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n
ma<-mean(diag(inv.sigma0%%s))
mg<-det(inv.sigma0%%s)^(1/p)
test<-n*p*(ma-log(mg)-1)
}

```



```

list(test=test, gl=gl, 'p-value'=pvalue, 'parametri di correzione a2 e c'=q)
}
#-----

S.TRV6<-function(dati,sigma0,alpha=0.05,errore=5e-3)
{
# Sigma=k*sigma, mu e k non note

pvalue<-0
rif<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}
gl<-(p-1)*(p+2)/2
inv.sigma0<-solve(sigma0)
s<-var(dati)*(n-1)/n
media<-c(rep(0,p))

t<-trv6(dati,inv.sigma0)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for(i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,sigma0,tol=1e-6,empirical=FALSE)
x<-trv6(dati0,inv.sigma0)

#Controllo se il valore appartiene o meno alla zona di rifiuto
if(x>t)
{
#Conto le osservazioni che cadono nella zona di rifiuto
rif<-rif+1
}
}

pvalue<-rif/N
errstimato<-((pvalue*(1-pvalue))/N)^(1/2)
er99<-2.57*errstimato
intervallo<-c(pvalue-er99,pvalue+er99)

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=t, gl=gl, 'p-value'=pvalue, 'intervallo di confidenza del p-value'=intervallo,
Nreplicazioni=N)
}

#-----

trv6<-function(dati,inv.sigma0)
{

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n
g<-det(inv.sigma0%%s)^(1/p)
test<-n*p*log(g)
}

#-----

empirico.trv6<-function(dati,alpha=0.05,errore=5e-3)
{

```

```

y1<-NULL
Sy<-0
v<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
gl<-(p-1)*(p+2)/2
media<-c(rep(0,p))
sigma<-diag(100,p,p)
inv.sigma<-solve(sigma)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
  {
    dati0<-NULL
    dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
    y<-trv6(dati0,inv.sigma)
    y1<-c(y1,y)
    Sy<-Sy+y
  }

Ey<-Sy/N
v<-var(y1)

a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

h<-c(a2,c)
}

#####

TRV72<-function(scelta=0)
{
print("Scegliere una delle 5 opzioni")
print("-----")
print("Test valido per p1=1,2 (0 p2=1,2) ==> V.TRV72(dati,p1,p2)")
print("Bartlett analitico ##> B.a.TRV72(dati,p1,p2)")
print("Bartlett empirico1(media) ##> B.e1.TRV72(dati,p1,p2,a1,alpha,errore)")
print("Bartlett empirico2(media e var.) ##> B.e2.TRV72(dati,p1,p2,a2,c,alpha,errore)")
print("Simulazione ##> S.TRV72(dati,p1,p2,alpha,errore)")
}

#-----

V.TRV72<-function(dati,p1,p2)
{
# sigma12=0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(p1+p2>p) {stop("Errore p1+p2>variabili(dati)")}
if(p1==0 | p2==0) {stop("Test valido solo per p1 e p2 diverso da zero")}
if(!((p1==1 | p1==2)|(p2==1 | p2==2))) {stop("Test valido solo per p1=1,2 o p2=1,2")}

I<-diag(1,p2,p2)

S11<-var(dati[,1:p1])*(n-1)/n
S22<-var(dati[, (p1+1):(p1+p2)])*(n-1)/n

```

```

S12<-cov(dati[,1:p1],dati[, (p1+1):(p1+p2)])*(n-1)/n
S21<-cov(dati[, (p1+1):(p1+p2)],dati[,1:p1])*(n-1)/n

l<-solve(S22)%*%S21%%solve(S11)%*%S12

k<-det(I-l)

if(p1==1)
{
test<-(n-p1-p2)/p2*(1-k)/(k)

g11<-p2
g12<-n-p1-p2
pvalue<-1-pf(test,g11,g12)
}

if(p1==2)
{
test<-(n-p1-p2)/p2*(1-sqrt(k))/sqrt(k)

g11<-2*p2
g12<-2*(n-p1-p2)
pvalue<-1-pf(test,g11,g12)
}

if(p2==1)
{
test<-(n-1-p1)/p1*(1-k)/(k)

g11<-p1
g12<-n-1-p1
pvalue<-1-pf(test,g11,g12)
}

if(p2==2)
{
test<-(n-2-p1)/p1*(1-sqrt(k))/(sqrt(k))

g11<-2*p1
g12<-2*(n-2-p1)
pvalue<-1-pf(test,g11,g12)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, g11=g11, g12=g12, 'p-value'=pvalue)
}

#-----

B.a.TRV72<-function(dati,p1,p2)
{
# sigma12=0, mu non nota

p<-dim(dati)[2]
if(p1+p2>p) {stop("Errore p1+p2>variabili(dati)")}

gl<-p1*p2
test<-trv72(dati,p1,p2)

pvalue<-1-pchisq(test,gl)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")

```

```

list(test=test, gl=gl, 'p-value'=pvalue)
}
#-----

B.e1.TRV72<-function(dati,p1,p2,a1=0,alpha=0.05,errore=5e-3)
{
# sigma12=0, mu non nota

p<-dim(dati)[2]
if(p1+p2>p) {stop("Errore p1+p2>variabili(dati)")}

gl<-p1*p2

if(a1==0)
{
h<-empirico.trv72(dati,p1,p2,alpha,errore)
a1<-h[1]
t<-trv72(dati,p1,p2)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

else
{
t<-trv72(dati,p1,p2)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametro di correzione a1'=a1)
}

#-----

B.e2.TRV72<-function(dati,p1,p2,a2=0,c=0,alpha=0.05,errore=5e-3)
{
# Sigma=sigma0, mu non nota

p<-dim(dati)[2]
if(p1+p2>p) {stop("Errore p1+p2>variabili(dati)")}

gl<-p1*p2

if(a2==0 & c==0)
{
h<-empirico.trv72(dati,p1,p2,alpha,errore)
a2<-h[2]
c<-h[3]
q<-c(a2,c)
t<-trv72(dati,p1,p2)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

else
{
q<-c(a2,c)
t<-trv72(dati,p1,p2)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}
}

```

```

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametri di correzione a2 e c'=q)
}

#-----

S.TRV72<-function(dati,p1,p2,alpha=0.05,errore=5e-3)
{
# sigma12=0, mu non nota

pvalue<-0
rif<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
if(p1+p2>p) {stop("Errore p1+p2>variabili(dati)")}
gl<-p1*p2
s<-var(dati[,1:(p1+p2)])*(n-1)/n
media<-c(rep(0,p))
I<-diag(10,p,p)

t<-trv72(dati,p1,p2)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for(i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,I,tol=1e-6,empirical=FALSE)
x<-trv72(dati0,p1,p2)

#Controllo se il valore appartiene o meno alla zona di rifiuto
if(x>t)
{
#Conto le osservazioni che cadono nella zona di rifiuto
rif<-rif+1
}
}
pvalue<-rif/N
errstimato<-((pvalue*(1-pvalue))/N)^(1/2)
er99<-2.57*errstimato
intervallo<-c(pvalue-er99,pvalue+er99)

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=t, gl=gl, 'p-value'=pvalue, 'intervallo di confidenza del p-value'=intervallo,
Nreplicazioni=N)
}

#-----

trv72<-function(dati,p1,p2)
{
n<-dim(dati)[1]
I<-diag(1,p2,p2)

S11<-var(dati[,1:p1])*(n-1)/n
S22<-var(dati[, (p1+1):(p1+p2)])*(n-1)/n
S12<-cov(dati[,1:p1],dati[, (p1+1):(p1+p2)])*(n-1)/n
S21<-cov(dati[, (p1+1):(p1+p2)],dati[,1:p1])*(n-1)/n

```

```

l<-solve(S22)%*%S21%*%solve(S11)%*%S12

test<--(n-0.5*(p1+p2+3))*log(det(I-l))
}

#-----

empirico.trv72<-function(dati,p1,p2,alpha=0.05,errore=5e-3)
{

y1<-NULL
Sy<-0
v<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
gl<-p1*p2
media<-c(rep(0,p))
I<-diag(1,p,p)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,I,tol=1e-6,empirical=FALSE)
y<-trv72(dati0,p1,p2)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)

a1<-Ey/gl
a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

h<-c(a1,a2,c)
}

#####

TRV7<-function(scelta=0)
{

print("Scegliere una delle 4 opzioni")
print("-----")
print("Bartlett analitico          ##> B.a.TRV7(dati)")
print("Bartlett empirico1(media)   ##> B.e1.TRV7(dati,a1,alpha,errore)")
print("Bartlett empirico2(media e varianza) ##> B.e2.TRV7(dati,a2,c,alpha,errore)")
print("Simulazione                   ##> S.TRV7(dati,alpha,errore)")

}

#-----

B.a.TRV7<-function(dati)
{
# sigma=I, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
gl<-p*(p-1)/2

```

```

test<-(n-1/6*(2*p+1))*(trv7(dati))/n
pvalue<-1-pchisq(test,gl)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue)
}

#-----

B.e1.TRV7<-function(dati,a1=0,alpha=0.05,errore=5e-3)
{
# sigma=I, mu non nota

p<-dim(dati)[2]
gl<-p*(p-1)/2

if(a1==0)
{
h<-empirico.trv7(dati,alpha,errore)
a1<-h[1]
t<-trv7(dati)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

else
{
t<-trv7(dati)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametro di correzione a1'=a1)
}

#-----

B.e2.TRV7<-function(dati,a2=0,c=0,alpha=0.05,errore=5e-3)
{
# Sigma=I, mu non nota

p<-dim(dati)[2]
gl<-p*(p-1)/2

if(a2==0 & c==0)
{
h<-empirico.trv7(dati,alpha,errore)
a2<-h[2]
c<-h[3]
q<-c(a2,c)
t<-trv7(dati)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

else
{
q<-c(a2,c)

```

```

t<-trv7(dati)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametri di correzione a2 e c'=q)
}

#-----

S.TRV7<-function(dati,alpha=0.05,errore=5e-3)
{
# Sigma diagonale, mu non nota

pvalue<-0
rif<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
gl<-p*(p-1)/2
s<-var(dati)*(n-1)/n
media<-c(rep(0,p))
I<-diag(1,p,p)

t<-trv7(dati)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for(i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,I,tol=1e-6,empirical=FALSE)
x<-trv7(dati0)

#Controllo se il valore appartiene o meno alla zona di rifiuto
if(x>t)
{
#Conto le osservazioni che cadono nella zona di rifiuto
rif<-rif+1
}
}
pvalue<-rif/N
errstimato<-((pvalue*(1-pvalue))/N)^(1/2)
er99<-2.57*errstimato
intervallo<-c(pvalue-er99,pvalue+er99)

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=t, gl=gl, 'p-value'=pvalue, 'intervallo di confidenza del p-value'=intervallo,
Nreplicazioni=N)
}

#-----

trv7<-function(dati)
{
R<-cor(dati)
test<-n*log(det(R))
}

```

```

#-----
empirico.trv7<-function(dati,alpha=0.05,errore=5e-3)
{
y1<-NULL
Sy<-0
v<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
gl<-p*(p-1)/2
media<-c(rep(0,p))
I<-diag(1,p,p)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,I,tol=1e-6,empirical=FALSE)
y<-trv7(dati0)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)

a1<-Ey/gl
a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

h<-c(a1,a2,c)
}

#####

TRV82<-function(dati,num)
{
# mu1=mu2, con sigma1=sigma2
# T^2 a due campioni

g<-length(num)
if(g!=2) {stop("Test valido solo per due campioni")}
dati<-as.matrix(dati)
n<-dim(dati)[1]
p<-dim(dati)[2]
n1<-num[1]
n2<-num[2]
W<-matrix(rep(0,p^2),p,p)
ba<-1
m1<-apply(dati[1:n1,],2,FUN=mean)
m2<-apply(dati[(n1+1):n,],2,FUN=mean)
for(i in 1:g){
ni<-num[i]
sn<-var(dati[ba:(ni+ba-1),])*ni
W<-W+sn
ba<-ba+ni
}
g11<-p
g12<-n-p-1
test<-t(m1-m2)%*%solve(W)%*%(m1-m2)*n1*n2*gl2/(n*g11)

pvalue<-1-pf(test,g11,g12)
}

```



```

empirico.trv8<-function(dati,num,alpha=0.05,errore=5e-3)
{
y1<-NULL
Sy<-0
v<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
gl<-p*(g-1)
media<-c(rep(0,p))
sigma<-diag(100,p,p)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
y<-trv8(dati0,num)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)

a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

h<-c(a2,c)
}

#####

TRV9<-function(scelta=0)
{
print("Scegliere una delle 4 opzioni")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
print("Bartlett analitico          ==> B.a.TRV9(dati,num)")
print("Bartlett empirico1(media)    ==> B.e1.TRV9(dati,num,a1,alpha,errore)")
print("Bartlett empirico2(media e varianza) ==> B.e2.TRV9(dati,num,a2,c,alpha,errore)")
print("Simulazione                   ==> S.TRV9(dati,num,alpha,errore)")
}

#-----

B.a.TRV9<-function(dati,num)
{
# sigma1=...=sigmag

r<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
gl<-p*(p+1)*(g-1)/2
test<-0
W<-matrix(rep(0,p^2),p,p)
ba<-1
for(i in 1:g)

```

```

        {
        ni<-num[i]
        sn<-var(dati[ba:(ni+ba-1),])
        ba<-ba+ni
        test<-test+(ni-1)*log(det(sn))
        W<-W+sn*(ni-1)
        r<-r+1/(ni-1)
        }
r<-r-1/(n-g)
su<-W/(n-g)
testB<-(n-g)*log(det(su))-test

f<-1-((2*p^2+3*p-1)/(6*(p+1)*(g-1)))*r

test<-f*testB

pvalue<-1-pchisq(test,gl)

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue)
}

#-----

B.e1.TRV9<-function(dati,num,a1=0,alpha=0.05,errore=5e-3)
{
# sigma1=...=sigmag

n<-dim(dati)[1]

p<-dim(dati)[2]
g<-length(num)
gl<-p*(p+1)*(g-1)/2

if(a1==0)
{
h<-empirico.trv9(dati,num,alpha,errore)
a1<-h[1]
t<-trv9(dati,num)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

else
{
t<-trv9(dati,num)
test<-t/a1
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametro di correzione a1'=a1)
}

#-----

B.e2.TRV9<-function(dati,num,a2=0,c=0,alpha=0.05,errore=5e-3)
{
# sigma1=...=sigmag

n<-dim(dati)[1]

```



```

intervallo<-c(pvalue-er99,pvalue+er99)

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=t, gl=gl, 'p-value'=pvalue, 'intervallo di confidenza del p-value'=intervallo,
      Nreplicazioni=N)
}

#-----

trv9<-function(dati,num)
{

n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
gl<-p*(p+1)*(g-1)/2
test<-0
W<-matrix(rep(0,p^2),p,p)
ba<-1
for(i in 1:g)
  {
    ni<-num[i]
    sn<-var(dati[ba:(ni+ba-1),])*ni
    ba<-ba+ni
    test<-test+ni*log(det(sn/ni))
    W<-W+sn
  }

test<-n*log(det(W/n))-test
}

#-----

empirico.trv9<-function(dati,num,alpha=0.05,errore=5e-3)
{

y1<-NULL
Sy<-0
v<-0
g<-length(num)
n<-dim(dati)[1]
p<-dim(dati)[2]
gl<-p*(p+1)*(g-1)/2
media<-c(rep(0,p))
sigma<-diag(100,p,p)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
  {
    dati0<-NULL
    dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
    y<-trv10(dati0,num)
    y1<-c(y1,y)
    Sy<-Sy+y
  }

Ey<-Sy/N
v<-var(y1)

a1<-Ey/gl
a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

```

```

h<-c(a1,a2,c)
}

#####

TRV10<-function(scelta=0)
{

print("Scegliere una delle 2 opzioni")
print("-----")
print("Bartlett empirico2(media e var.)    ##>  B.e2.TRV10(dati,num,a2,c,alpha,errore)")
print("Simulazione                        ##>  S.TRV10(dati,num,alpha,errore)")

}

#-----

B.e2.TRV10<-function(dati,num,a2=0,c=0,alpha=0.05,errore=5e-3)
{
# mul=...=mug e sigmal=...=sigmag

n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
gl<-p*(p+3)*(g-1)/2

if(a2==0 & c==0)
{
h<-empirico.trv10(dati,num,alpha,errore)
a2<-h[1]
c<-h[2]
q<-c(a2,c)
t<-trv10(dati,num)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

else
{
q<-c(a2,c)
t<-trv10(dati,num)
test<-t/a2+c
pvalue<-1-pchisq(test,gl)
}

print("Risultati del test")
print("-----")
list(test=test, gl=gl, 'p-value'=pvalue, 'parametri di correzione a2 e c'=q)
}

#-----

S.TRV10<-function(dati,num,alpha=0.05,errore=5e-3)
{
# mul=...=mug e sigmal=...=sigmag

pvalue<-0
rif<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
gl<-p*(p+3)*(g-1)/2

```

```

media<-c(rep(0,p))
sigma<-diag(100,p,p)

t<-trv10(dati,num)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for(i in 1:N)
  {
  dati0<-NULL
  dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
  x<-trv10(dati0,num)

  #Controllo se il valore appartiene o meno alla zona di rifiuto
  if(x>t)
    {
    #Conto le osservazioni che cadono nella zona di rifiuto
    rif<-rif+1
    }
  }

pvalue<-rif/N
errstimato<-((pvalue*(1-pvalue))/N)^(1/2)
er99<-2.57*errstimato
intervallo<-c(pvalue-er99,pvalue+er99)

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list(test=t, gl=gl, 'p-value'=pvalue, 'intervallo di confidenza del p-value'=intervallo,
      Nreplicazioni=N)
}

#-----

trv10<-function(dati,num)
{

dati<-as.matrix(dati)
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
test<-0
ba<-1
for(i in 1:g){
  ni<-num[i]
  sn<-var(dati[ba:(ni+ba-1),])*(ni-1)
  test<-test-ni*log(det(sn/ni))
  ba<-ba+ni
}
test<-test+n*log(det(var(dati)*(n-1)/n))
}

#-----

empirico.trv10<-function(dati,num,alpha=0.05,errore=5e-3)
{

y1<-NULL
Sy<-0
v<-0

n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)

```

```

gl<-p*(p+3)*(g-1)/2
media<-c(rep(0,p))
sigma<-diag(100,p,p)

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
y<-trv10(dati0,num)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)

a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

h<-c(a2,c)
}

#####
#####TEST UNIONE-INTERSEZIONE#####
#####

UIT1<-function(dati,m0,sigma,alpha=0.05)
{
# mu=m0, sigma nota

n<-dim(dati)[1]
p<-dim(dati)[2]
m<-apply(dati,2,FUN=mean)
if(length(m0)!=p) {stop("Test valido solo se length(m0)=length(medie)")}
gl<-p
unlink("salva.Rdata")

test<-n*t(m-m0)%*%solve(sigma)%*(m-m0)

pvalue<-1-pchisq(test,gl)
qmax<-qchisq(1-alpha,gl)

if(test>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,xlim=c(0,test),main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con UIT1.a(a) dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente ")
e<-c(f,g,h,k)
}

if(test<qmax)
{

```

```

pvec<-seq(0,(1-alpha),0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
e<-c("Accettiamo l'ipotesi globale")
}

pause()

save(dati,m0,sigma,alpha, file = "salva.Rdata")

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=test, 'p-value'=pvalue,
      'Valore critico della distribuzione del Test'=qmax, alpha=alpha,
      Ipotesi=e)
}

#-----

UIT1.a<-function(a)
{
# mu=m0, sigma nota

load("salva.Rdata")
n<-dim(dati)[1]
p<-dim(dati)[2]
if(length(m0)!=p) {stop("Test valido solo se length(m0)=length(medie)")}
if(length(a)!=p) {stop("Test valido solo se length(a)=varabili dei dati")}
m<-apply(dati,2,FUN=mean)
gl<-p

test.a<-n*t(a)**(m-m0)**t(m-m0)**a/(t(a)**sigma**a)

pvalue<-1-pchisq(test.a,gl)
qmax<-qchisq(1-alpha,gl)

if(test.a>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,xlim=c(0,test.a),main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Rifiutiamo l'ipotesi componente")
}

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T.",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
}

```

```

abline(v=test.a, lty=6, col="green3")
e<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a: '=a, 'Valore del test U.I.T. ipotesi componente '=test.a,
      'p-value '=pvalue, 'Valore critico della distribuzione del Test '=qmax,
      alpha=alpha, Ipotesi=e)
}

#####

UIT2<-function(dati,m0,alpha=0.05)
{
# mu=m0, sigma non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
if(length(m0)!=p) {stop("Test valido solo se length(m0)=length(medie)")}
m<-apply(dati,2,FUN=mean)
s<-var(dati)*(n-1)/n
g11<-p
g12<-n-p
unlink("salva.Rdata")

test<-(n-p)/p*t(m-m0)%*%solve(s)%*%(m-m0)

pvalue<-1-pf(test,g11,g12)
qmax<-qf(1-alpha,g11,g12)

if(test>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,xlim=c(0,test),main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con          UIT2.a(a)          dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente ")
e<-c(f,g,h,k)
}

if(test<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
e<-c("Accettiamo l'ipotesi globale")
}

pause()

```

```

save(dati,m0,alpha, file= "salva.Rdata")

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=test, 'p-value'=pvalue,
     'Valore critico della distribuzione del Test'=qmax, alpha=alpha,
     Ipotesi=e)
}

#-----

UIT2.a<-function(a)
{
# mu=m0, sigma non nota

load("salva.Rdata")
n<-dim(dati)[1]
p<-dim(dati)[2]
if(length(m0)!=p) {stop("Test valido solo se length(m0)=length(medie)")}
if(length(a)!=p) {stop("Test valido solo se length(a)=varabili dei dati")}
m<-apply(dati,2,FUN=mean)
s<-var(dati)*(n-1)/n
g11<-p
g12<-n-p

test.a<-(n-p)/p*t(a)**(m-m0)**t(m-m0)**a/(t(a)**s**a)

pvalue<-1-pf(test.a,g11,g12)
qmax<-qf(1-alpha,g11,g12)

if(test.a>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,xlim=c(0,test.a),main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Rifiutiamo l'ipotesi componente")
}

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T.",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a: '=a, 'Valore del test U.I.T. ipotesi componente'=test.a,
     'p-value'=pvalue, 'Valore critico della distribuzione del Test'=qmax,

```

```

        alpha=alpha, Ipotesi=e)
}

#####

UIT3<-function(dati,sigma0,alpha=0.05,errore=5e-3)
{
# sigma=sigma0, con mu non nota

p<-dim(dati)[2]
if(dim(sigma0)[1]!=p) {stop("Test valido solo se dim(sigma0)=dim(sigma)")}

simula3(dati,sigma0,alpha,errore)
load("salva.Rdata")

q<-c(qmin,qmax)
par(mfrow=c(1,2))

if(t[2]>qmax | t[1]<qmin)
{
pvec<-seq(0,1,0.001)

m<-quantile(d[,1],pvec)
m<-density(m)
plot(m,main="Distribuzione lambda-1 ",
xlab="valori di lambda-1",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=t[2], lty=6, col="blue")

n<-quantile(d[,2],pvec)
n<-density(n)
plot(n,main="Distribuzione lambda-p",
xlab="valori di lambda-p",ylab="densità")
abline(v=qmin,lty=2, col="red")
abline(v=t[1], lty=6, col="blue")
f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con      UIT3.a(a)      dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente")
e<-c(f,g,h,k)
}

if(t[2]<qmax & t[1]>qmin)
{
pvec<-seq(0,(1-alpha),0.001)

m<-quantile(d[,1],pvec)
m<-density(m)
plot(m,main="Distribuzione lambda-1 ",
xlab="valori di lambda-1",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=t[2], lty=6, col="blue")

n<-quantile(d[,2],pvec)
n<-density(n)
plot(n,main="Distribuzione lambda-p",
xlab="valori di lambda-p",ylab="densità")
abline(v=qmin,lty=2, col="red")
}

```

```

abline(v=t[1], lty=6, col="blue")
e<-c("Accettiamo l'ipotesi globale")
}

par(mfrow=c(1,1))
pause()

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=t,
      'Valori critici della distribuzione del Test'=q,
      alpha.effettivo=alpha.effettivo, Ipotesi=e, Nreplicazioni=N)
}

#-----

simula3<-function(dati,sigma0,alpha,errore)
{
n<-dim(dati)[1]
p<-dim(dati)[2]
media<-c(rep(0,p))
inv.sigma0<-solve(sigma0)
x1<-NULL
k1<-NULL
j<-0
unlink("salva.Rdata")

N<-round((2.57/errore)^2*alpha*(1-alpha))

  for(i in 1:N)
  {
    dati0<-NULL
    dati0<-mvrnorm(n=n,media,sigma0,tol=1e-6,empirical=FALSE)
    k<-uit3(dati0,inv.sigma0)
    k1<-rbind(k1,k)
  }

x2<-k1[,1]
y2<-k1[,p]
x1<-sort(x2)
y1<-sort(y2)
d<-cbind(x1,y1)

qmax<-quantile(x1,1-alpha/2)
qmin<-quantile(y1,0+alpha/2)

  for(i in 1:N)
  {
    if(x2[i]>qmax & y2[i]<qmin) {j<-j+1}
  }

alpha.effettivo<-alpha+j/N
test<-uit3(dati,inv.sigma0)
t1<-test[1]
tp<-test[p]
t<-c(tp,t1)

save(d,qmax,qmin,t,dati,sigma0,alpha.effettivo,alpha,N, file = "salva.Rdata")
}

```

```

#-----

uit3<-function(dati,inv.sigma0)
{
# Sigma=sigma0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n

test<-eigen(inv.sigma0*s)$values
}

#-----

UIT3.a<-function(a)
{
# Sigma=sigma0, mu non nota

load("salva.Rdata")
q<-c(qmin,qmax)
par(mfrow=c(1,2))

p<-dim(dati)[2]
if(length(a)!=p) {stop("Test valido solo se length(a)=variabili dei dati")}

test.a<-uit3.a(dati,sigma0,a)

if(test.a>qmax | test.a<qmin)
{
pvec<-seq(0,1,0.001)

m<-quantile(d[,1],pvec)
m<-density(m)
plot(m,main="Distribuzione lambda-1 ",
xlab="valori di lambda-1",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=t[2], lty=6, col="blue")

n<-quantile(d[,2],pvec)
n<-density(n)
plot(n,main="Distribuzione lambda-p",
xlab="valori di lambda-p",ylab="densità")
abline(v=qmin,lty=2, col="red")
abline(v=t[1], lty=6, col="blue")
p<-c("Rifiutiamo l'ipotesi componente")
}

if(test.a<qmax & test.a>qmin)
{
pvec<-seq(0,(1-alpha),0.001)

m<-quantile(d[,1],pvec)
m<-density(m)
plot(m,main="Distribuzione lambda-1 ",
xlab="valori di lambda-1",ylab="densità")
abline(v=qmax,lty=2, col="red")
}

```

```

abline(v=t[2], lty=6, col="green")

n<-quantile(d[,2],pvec)
n<-density(n)
plot(n,main="Distribuzione lambda-p",
xlab="valori di lambda-p",ylab="densità")
abline(v=qmin,lty=2, col="red")
abline(v=t[1], lty=6, col="green")
p<-c("Accettiamo l'ipotesi componente")
}

par(mfrow=c(1,1))
pause()

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a: '=a, 'Valore del test U.I.T. componente '=test.a,
      'Valori critici della distribuzione del Test '=q,
      alpha=alpha, Ipotesi=p)
}

#-----

uit3.a<-function(dati,sigma0,a)
{
# Sigma=sigma0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n

test<-eigen(t(a)%*%s%*a/t(a)%*%sigma0%*a)$values
}

#####

UIT4<-function(dati,sigma,R,r,alpha=0.05)
{
# Rmu=r, sigma nota

n<-dim(dati)[1]
p<-dim(dati)[2]
m<-apply(dati,2,FUN=mean)
q<-dim(R)[1]
gl<-q
unlink("salva.Rdata")

test<-n*t(R%*%m-r)%*%solve(R%*%sigma%*%t(R))%*%(R%*%m-r)

pvalue<-1-pchisq(test,gl)
qmax<-qchisq(1-alpha,gl)

if(test>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,xlim=c(0,test),main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
}
}

```

```

abline(v=test, lty=6, col="blue")

f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con      UIT4.a(a)      dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente")
e<-c(f,g,h,k)
}

if(test<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
e<-c("Accettiamo l'ipotesi globale")
}

pause()

save(dati,sigma,R,r,alpha, file= "salva.Rdata")

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=test,pvalue=pvalue,
      'Valore critico della distribuzione del Test'=qmax, alpha=alpha,
      Ipotesi=e)
}

#-----

UIT4.a<-function(a)
{
# Rmu=r, sigma nota

load("salva.Rdata")
n<-dim(dati)[1]
p<-dim(dati)[2]
m<-apply(dati,2,FUN=mean)
q<-dim(R)[1]
gl<-q

test.a<-n*t(a)**(R**m-r)**t(R**m-r)**a/t(a)**(R**sigma**t(R))**a

pvalue<-1-pchisq(test.a,gl)
qmax<-qchisq(1-alpha,gl)

if(test.a>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,xlim=c(0,test.a),main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Rifiutiamo l'ipotesi componente")
}
}

```

```

}

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qchisq(pvec,gl)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T.",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a:'=a, 'Valore del test U.I.T. ipotesi componente'=test.a,
      pvalue=pvalue, 'Valore critico della distribuzione del Test'=qmax,
      alpha=alpha, Ipotesi=e)
}

#####

UIT5<-function(dati,R,r,alpha=0.05)
{
# Rmu=r, sigma non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
m<-apply(dati,2,FUN=mean)
s<-var(dati)*(n-1)/n
q<-dim(R)[1]
gl1<-q
gl2<-n-q
unlink("salva.Rdata")

test<-((n-q)/q)*t(R%*%m-r)%*%solve(R%*%s%*%t(R))%*%(R%*%m-r)

pvalue<-1-pf(test,gl1,gl2)
qmax<-qf(1-alpha,gl1,gl2)

if(test>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qf(pvec,gl1,gl2)
f<-density(f)
plot(f,xlim=c(0,test),main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con          UIT5.a(a)          dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente")
e<-c(f,g,h,k)
}

if(test<qmax)
{

```

```

pvec<-seq(0,(1-alpha),0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
e<-c("Accettiamo l'ipotesi globale")
}

pause()

save(dati,R,r,alpha, file= "salva.Rdata")

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=test,pvalue=pvalue,
      'Valore critico della distribuzione del Test'=qmax, alpha=alpha,
      Ipotesi=e)
}

#-----

UIT5.a<-function(a)
{
# Rmu=r, sigma non nota

load("salva.Rdata")
n<-dim(dati)[1]
p<-dim(dati)[2]
m<-apply(dati,2,FUN=mean)
s<-var(dati)*(n-1)/n
q<-dim(R)[1]
g11<-q
g12<-n-q

test.a<-((n-q)/q)*t(a)**(R**m-r)**t(R**m-r)**a/t(a)**(R**S**t(R))**a

pvalue<-1-pf(test.a,g11,g12)
qmax<-qf(1-alpha,g11,g12)

if(test.a>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,xlim=c(0,test.a),main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Rifiutiamo l'ipotesi componente")
}

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T.",
xlab="valori del test u.i.t. ipotesi componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
}

```

```

abline(v=test.a, lty=6, col="green3")
e<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a:'=a, 'Valore del test U.I.T. ipotesi componente'=test.a,
      pvalue=pvalue, 'Valore critico della distribuzione del Test'=qmax,
      alpha=alpha, Ipotesi=e)
}

#####

UIT72<-function(dati,p1,p2,alpha=0.05,errore=5e-3)
{
# sigma12=0, mu non nota

p<-dim(dati)[2]
if(p1+p2>p)      {stop("Errore p1+p2>variabili(dati)")}

if((p1==1 | p1==2)|(p2==1 | p2==2))
{

I<-diag(1,p2,p2)

S11<-var(dati[,1:p1])*(n-1)/n
S22<-var(dati[, (p1+1):(p1+p2)])*(n-1)/n
S12<-cov(dati[,1:p1],dati[, (p1+1):(p1+p2)])*(n-1)/n
S21<-cov(dati[, (p1+1):(p1+p2)],dati[,1:p1])*(n-1)/n

l<-solve(S22)%*%S21%*%solve(S11)%*%S12

k<-det(I-l)

if(p1==1)
{
test<-(n-p1-p2)/p2*(1-k)/(k)

g11<-p2
g12<-n-p1-p2
}

if(p1==2)
{
test<-(n-p1-p2)/p2*(1-sqrt(k))/sqrt(k)

g11<-2*p2
g12<-2*(n-p1-p2)
}

if(p2==1)
{
test<-(n-1-p1)/p1*(1-k)/(k)

g11<-p1
g12<-n-1-p1
}
}

```



```

simula72<-function(dati,p1,p2,alpha,errore)
{
n<-dim(dati)[1]
p<-dim(dati)[2]
media<-c(rep(0,p))
sigma<-diag(100,p,p)
x1<-NULL
k1<-NULL
unlink("salva.Rdata")

test<-uit72(dati,p1,p2)

N<-round((2.57/errore)^2*alpha*(1-alpha))

  for(i in 1:N)
  {
    dati0<-NULL
    dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
    k<-uit72(dati0,p1,p2)
    k1<-rbind(k1,k)
  }

x1<-sort(k1)
qmax<-quantile(x1,1-alpha)

save(x1,qmax,dati,p1,p2,test,alpha,N, file = "salva.Rdata")
}

#-----

uit72<-function(dati,p1,p2)
{
# sigma12=0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
g1<-p1*p2
I<-diag(1,p2,p2)

S11<-var(dati[,1:p1])*(n-1)/n
S22<-var(dati[, (p1+1):(p1+p2)])*(n-1)/n
S12<-cov(dati[,1:p1],dati[, (p1+1):(p1+p2)])*(n-1)/n
S21<-cov(dati[, (p1+1):(p1+p2)],dati[,1:p1])*(n-1)/n

l<-eigen(solve(S22)%*%S21)%*%solve(S11)%*%S12)$values
t<-l[1]
}

#-----

UIT72.a<-function(a,d)
{
# sigma12=0, mu non nota

load("salva.Rdata")

p<-dim(dati)[2]
if(length(a)!=p1) {stop("Test valido solo se length(a)=p1 varabili")}
if(length(d)!=p2) {stop("Test valido solo se length(d)=p2 varabili")}

test.a<-uit72.a(dati,p1,p2,a,d)

```

```

if(test.a>qmax)
{
pvec<-seq(0.9,1,0.001)

f<-quantile(x1,pvec)
f<-density(f)
plot(f,xlim=c(0,test.a),main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
p<-c("Rifiutiamo l'ipotesi componente")
}

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-quantile(x1,pvec)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
p<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a: '=a, 'Vettore d: '=d, 'Valore del test U.I.T. componente '=test.a,
'Valore critico della distribuzione del Test '=qmax,
alpha=alpha, Ipotesi=p)
}

#-----

uit72.a<-function(dati,p1,p2,a,d)
{
# sigma12=0, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati[,1:(p1+p2)])*(n-1)/n
S11<-var(dati[,1:p1])*(n-1)/n
S22<-var(dati[, (p1+1):(p1+p2)])*(n-1)/n
S12<-cov(dati[,1:p1],dati[, (p1+1):(p1+p2)])*(n-1)/n
S21<-cov(dati[, (p1+1):(p1+p2)],dati[,1:p1])*(n-1)/n

test<-(t(a)%*%S12%*%d)^2/t(a)%*%S11%*%a%*%t(d)%*%S22%*%d
}

#####

UIT82<-function(dati,num,alpha=0.05)
{
# mu1=mu2, con sigma1=sigma2
# T^2 a due campioni

g<-length(num)
if(g!=2) {stop("Test valido solo per due campioni")}
dati<-as.matrix(dati)
n<-dim(dati)[1]
p<-dim(dati)[2]
n1<-num[1]

```

```

n2<-num[2]
W<-matrix(rep(0,p^2),p,p)
ba<-1
m1<-apply(dati[1:n1],2,FUN=mean)
m2<-apply(dati[(n1+1):n],2,FUN=mean)
for(i in 1:g){
  ni<-num[i]
  sn<-var(dati[ba:(ni+ba-1),])*(ni-1)
  W<-W+sn
  ba<-ba+ni
}
g11<-p
g12<-n-p-1
unlink("salva.Rdata")

test<-t(m1-m2)%*%solve(W)%*%(m1-m2)*n1*n2*g12/(n*g11)

pvalue<-1-pf(test,g11,g12)
qmax<-qf(1-alpha,g11,g12)

if(test>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,xlim=c(0,test),main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con UIT82.a(a) dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente")
e<-c(f,g,h,k)
}

if(test<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
e<-c("Accettiamo ipotesi globale")
}

pause()

save(dati,num,alpha, file="salva.Rdata")

print("Risultati del test")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=test, 'p-value'=pvalue,
'Valore critico della distribuzione del Test'=qmax, alpha=alpha,
Ipotesi=e)
}

#-----

```

```

UIT82.a<-function(a)
{
# mul=...=mug, con signal=...=sigmag

load("salva.Rdata")
dati<-as.matrix(dati)
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
media<-mean(dati)
w<-0
q<-0
ba<-1
for(i in 1:g){
  ni<-num[i]
  sn<-(ni-1)%*%t(a)%*%var(dati[ba:(ni+ba-1),])%*%a
  w<-w+sn
  m<-apply(dati[ba:(ni+ba-1),],2,FUN=mean)
  q1<-ni%*%(t(a)%*%(m-media))^2
  q<-q+q1
  ba<-ba+ni
}

test.a<-q/w

g11<-p
g12<-n-p-1
pvalue<-1-pf(test.a,g11,g12)
qmax<-qf(1-alpha,g11,g12)

if(test.a>qmax)
{
pvec<-seq(0,0.999,0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. componente ",
xlab="valori del test u.i.t. componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Rifiutiamo l'ipotesi componente")
}

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-qf(pvec,g11,g12)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
e<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati del test")
print("*****")
list('Vettore a':a, 'Valore del test U.I.T. componente'=test.a,
'p-value'=pvalue, 'Valore critico della distribuzione del Test'=qmax,
alpha=alpha, Ipotesi=e)

```

```

}

#####

UIT8<-function(dati,num,alpha=0.05,errore=5e-3)
{
# mul=...=mug, con sigma1=...=sigmag

simula8(dati,num,alpha,errore)
load("salva.Rdata")

if(test>qmax)
{
pvec<-seq(0,1,0.001)

f<-quantile(x1,pvec)
f<-density(f)
plot(f,xlim=c(0,test),main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
f<-c("Rifiutiamo l'ipotesi globale")
g<-c("Si può verificare se una qualche <ipotesi componente> abbia portato")
h<-c("al rifiuto con UIT8.a(a) dove a rappresenta il vettore")
k<-c("p-dimensionale di un'ipotesi componente")
e<-c(f,g,h,k)
}

if(test<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-quantile(x1,pvec)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. globale",
xlab="valori del test u.i.t. globale",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test, lty=6, col="blue")
e<-c("Accettiamo l'ipotesi globale")
}

pause()

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Valore del test U.I.T. globale'=test,
'Valore critico della distribuzione del Test'=qmax, alpha=alpha,
Ipotesi=e, Nreplicazioni=N)
}

#-----

simula8<-function(dati,num,alpha,errore)
{
n<-dim(dati)[1]
p<-dim(dati)[2]
media<-c(rep(0,p))
sigma<-diag(100,p,p)
x1<-NULL
k1<-NULL
unlink("salva.Rdata")

test<-uit8(dati,num)

```

```

N<-round((2.57/errore)^2*alpha*(1-alpha))

  for(i in 1:N)
  {
    dati0<-NULL
    dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
    k<-uit8(dati0,num)
    k1<-rbind(k1,k)
  }

x1<-sort(k1)
qmax<-quantile(x1,1-alpha)

save(x1,qmax,dati,num,test,alpha,N, file = "salva.Rdata")
}

#-----

uit8<-function(dati,num)
{

dati<-as.matrix(dati)
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
T<-var(dati)*(n-1)
w<-matrix(rep(0,p^2),p,p)
ba<-1
for(i in 1:g){
  ni<-num[i]
  sn<-var(dati[ba:(ni+ba-1),])*ni
  ba<-ba+ni
  w<-w+sn
}
B<-T-w
t<-eigen(B/w)$values
test<-t[1]
}

#-----

UIT8.a<-function(a)
{
# mul=...=mug, con sigmal=...=sigmag

load("salva.Rdata")

test.a<-uit8.a(dati,num,a,qmax)

if(test.a>qmax)
{
pvec<-seq(0,1,0.001)

f<-quantile(x1,pvec)
f<-density(f)
plot(f,xlim=c(0,test.a),main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
p<-c("Rifiutiamo l'ipotesi componente")
}
}

```

```

if(test.a<qmax)
{
pvec<-seq(0,(1-alpha),0.001)

f<-quantile(x1,pvec)
f<-density(f)
plot(f,main="Distribuzione del test U.I.T. componente",
xlab="valori del test u.i.t. componente",ylab="densità")
abline(v=qmax,lty=2, col="red")
abline(v=test.a, lty=6, col="green3")
p<-c("Accettiamo l'ipotesi componente")
}

pause()

print("Risultati della simulazione")
print("-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*-*")
list('Vettore a:'=a, 'Valore del test U.I.T. componente'=test.a,
'Valore critico della distribuzione del Test'=qmax,
alpha=alpha, Ipotesi=p)

}

#-----

uit8.a<-function(dati,num,a,qmax)
{
# mul=...=mug, con signal=...=sigmag

dati<-as.matrix(dati)
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
media<-mean(dati)
w<-0
q<-0
ba<-1
for(i in 1:g){
ni<-num[i]
sn<-(ni-1)*t(a)**var(dati[ba:(ni+ba-1),])**a
w<-w+sn
m<-apply(dati[ba:(ni+ba-1),],2,FUN=mean)
q1<-ni**t(a)**(m-media)^2
q<-q+q1
ba<-ba+ni
}

test<-q/w
}

#####

pause<-function ()
{
cat("Pause. Press <Enter> to continue...")
readline()
invisible()
}

library(MASS)

```

```
#####
#####Bontà#####
#####
```

```
bontà3<-function(p,n,alpha=0.05,errore=5e-3)
{
nr<-10000
pvasintotico<-0
pvanalitica<-0
pvempirical<-0
pvempirica2<-0
pvkorin<-0
pvsimulazione<-0

media<-c(rep(0,p))
sigma<-diag(100,p,p)
inv.sigma<-solve(sigma)

dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)

h<-p.e.trv3(n,p,alpha,errore)
a1<-h[1]
a2<-h[2]
c<-h[3]
k<-trv3.korin(dati,inv.sigma)

gl<-p*(p+1)/2
q1<-k[2]
q2<-k[3]

soglia<-qchisq(1-alpha,gl)
sogliasimulazione<-h[4]
sogliakorin<-qf(1-alpha,q1,q2)

for(i in 1:nr)
{
dati<-NULL
dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
t<-trv3.n(dati,inv.sigma)
x<-trv3.bart(dati,inv.sigma)
e1<-t/a1
e2<-t/a2+c
b<-trv3.korin(dati,inv.sigma)[1]

if(t>soglia)          pvasintotico<-pvasintotico+1
if(x>soglia)          pvanalitica<-pvanalitica+1
if(e1>soglia)         pvempirical<-pvempirical+1
if(e2>soglia)         pvempirica2<-pvempirica2+1
if(b>sogliakorin)     pvkorin<-pvkorin+1
if(t>sogliasimulazione) pvsimulazione<-pvsimulazione+1
}

pvasintotico<-pvasintotico/nr
pvanalitica<-pvanalitica/nr
pvempirical<-pvempirical/nr
pvempirica2<-pvempirica2/nr
pvkorin<-pvkorin/nr
pvsimulazione<-pvsimulazione/nr

h<-c(gl,pvasintotico,pvanalitica,pvempirical,pvempirica2,pvkorin,pvsimulazione)
print(h)
}
}
```

```
#-----
```

```

trv3.n<-function(dati,inv.sigma)
{

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n
ma<-mean(diag(inv.sigma%%s))
mg<-det(inv.sigma%%s)^(1/p)
test<-n*p*(ma-log(mg)-1)
}

#-----

trv3.bart<-function(dati,inv.sigma)
{
# Sigma=sigma, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n
ma<-mean(diag(inv.sigma%%s))
mg<-det(inv.sigma%%s)^(1/p)
a<-(1-(2*p+1-2/(p+1)))/(6*n))
test<-n*p*(ma-log(mg)-1)*a

test
}

#-----

trv3.korin<-function(dati,inv.sigma)
{
# Sigma=sigma, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]
s<-var(dati)*(n-1)/n
ma<-mean(diag(inv.sigma%%s))
mg<-det(inv.sigma%%s)^(1/p)
v<-0.5*p*(p+1)
D1<-(2*p+1-2/(p+1))/(6*n)
D2<-(p-1)*(p+2)/(6*n^2)
w<-D2-D1^2
q1<-v
q2<-(v+2)/(w)
b<-v/(1-D1-v/q2)

test<-n*p*(ma-log(mg)-1)/(b)

h<-c(test,q1,q2)
}
#-----

p.e.trv3<-function(n,p,alpha,errore)
{

y1<-NULL
Sy<-0
v<-0
media<-c(rep(0,p))
sigma<-diag(100,p,p)
g1<-p*(p+1)/2
inv.sigma<-solve(sigma)

N<-round((2.57/errore)^2*alpha*(1-alpha))

```

```

for( i in 1:N)
  {
    dati<-NULL
    dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
    y<-trv3.n(dati,inv.sigma)
    y1<-c(y1,y)
    Sy<-Sy+y
  }

Ey<-Sy/N
v<-var(y1)
a1<-Ey/gl
a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

y1<-sort(y1)
percentile<-quantile(y1,1-alpha)

h<-c(a1,a2,c,percentile)
}

library(MASS)

#####

bontà72<-function(p1,p2,n,alpha=0.05,errore=5e-3)
{

nr<-10000
pvanalitica<-0
pvempirical<-0
pvempirica2<-0
pvsimulazione<-0

p<-p1+p2
media<-c(rep(0,p))
sigma<-diag(100,p,p)

dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)

h<-p.e.trv72(n,p1,p2,alpha,errore)
a1<-h[1]
a2<-h[2]
c<-h[3]

gl<-p1*p2

soglia<-qchisq(1-alpha,gl)
sogliasimulazione<-h[4]

for(i in 1:nr)
  {
    dati<-NULL
    dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
    t<-trv72.n(dati,p1,p2)
    e1<-t/a1
    e2<-t/a2+c

    if(x>soglia)          pvanalitica<-pvanalitica+1
    if(e1>soglia)        pvempirical<-pvempirical+1
    if(e2>soglia)        pvempirica2<-pvempirica2+1
    if(t>sogliasimulazione)  pvsimulazione<-pvsimulazione+1
  }
}

```

```

    }

pvanalitica<-pvanalitica/nr
pvempirical<-pvempirical/nr
pvempirica2<-pvempirica2/nr
pvsimulazione<-pvsimulazione/nr

h<-c(g1,pvanalitica,pvempirical,pvempirica2,pvsimulazione)
print(h)
}

#-----

trv72.n<-function(dati,p1,p2)
{
n<-dim(dati)[1]
I<-diag(1,p2,p2)

S11<-var(dati[,1:p1])*(n-1)/n
S22<-var(dati[, (p1+1):(p1+p2)])*(n-1)/n
S12<-cov(dati[,1:p1],dati[, (p1+1):(p1+p2)])*(n-1)/n
S21<-cov(dati[, (p1+1):(p1+p2)],dati[,1:p1])*(n-1)/n

l<-solve(S22)%*%S21%*%solve(S11)%*%S12

test<--(n-0.5*(p1+p2+3))*log(det(I-l))
}

#-----

p.e.trv72<-function(n,p1,p2,alpha,errore)
{

y1<-NULL
Sy<-0
v<-0
p<-p1+p2
media<-c(rep(0,p))
sigma<-diag(100,p,p)
g1<-p1*p2

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
y<-trv72.n(dati0,p1,p2)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)
a1<-Ey/g1
a2<-sqrt(v/(2*g1))
b<-g1-(Ey/a2)

y1<-sort(y1)
percentile<-quantile(y1,1-alpha)

h<-c(a1,a2,c,percentile)
}

```

```

library(MASS)

#####

bontà7<-function(p,n,alpha=0.05,errore=5e-3)
{

nr<-10000
pvasintotico<-0
pvanalitica<-0
pvempirical<-0
pvempirica2<-0
pvsimulazione<-0

media<-c(rep(0,p))
sigma<-diag(100,p,p)

dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)

h<-p.e.trv7(n,p,alpha,errore)
a1<-h[1]
a2<-h[2]
c<-h[3]
gl<-p*(p-1)/2

soglia<-qchisq(1-alpha,gl)
sogliasimulazione<-h[4]

for(i in 1:nr)
{
dati<-NULL
dati<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
t<-trv7.n(dati)
x<-trv7.Bart(dati)
e1<-t/a1
e2<-t/a2+c

if(t>soglia)          pvasintotico<-pvasintotico+1
if(x>soglia)          pvanalitica<-pvanalitica+1
if(e1>soglia)         pvempirical<-pvempirical+1
if(e2>soglia)         pvempirica2<-pvempirica2+1
if(t>sogliasimulazione) pvsimulazione<-pvsimulazione+1
}

pvasintotico<-pvasintotico/nr
pvanalitica<-pvanalitica/nr
pvempirical<-pvempirical/nr
pvempirica2<-pvempirica2/nr
pvsimulazione<-pvsimulazione/nr

h<-c(gl,pvasintotico,pvanalitica,pvempirical,pvempirica2,pvsimulazione)
print(h)
}

#-----

trv7.n<-function(dati)
{
n<-dim(dati)[1]
R<-cor(dati)
test<--n*log(det(R))
}

```

```

#-----

trv7.Bart<-function(dati)
{
# sigma=I, mu non nota

n<-dim(dati)[1]
p<-dim(dati)[2]

test<-(n-1/6*(2*p+11))/n*(trv7.n(dati))
}

#-----

p.e.trv7<-function(n,p,alpha,errore)
{

y1<-NULL
Sy<-0
v<-0
media<-c(rep(0,p))
sigma<-diag(100,p,p)
gl<-p*(p-1)/2

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=n,media,sigma,tol=1e-6,empirical=FALSE)
y<-trv7.n(dati0)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)
a1<-Ey/gl
a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

y1<-sort(y1)
percentile<-quantile(y1,1-alpha)

h<-c(a1,a2,c,percentile)
}

library(MASS)

#####

bontà9<-function(g,p,n,alpha=0.05,errore=5e-3)
{

nr<-10000
pvasintotico<-0
pvanalitica<-0
pvempirical<-0
pvempirica2<-0
pvsimulazione<-0

nc<-n*g

```

```

num<-c(rep(nc/g,g))
media<-c(rep(0,p))
sigma<-diag(100,p,p)

dati<-mvrnorm(n=nc,media,sigma,tol=1e-6,empirical=FALSE)

h<-p.e.trv9(nc,p,num,alpha,errore)
a1<-h[1]
a2<-h[2]
c<-h[3]
gl<-p*(p+1)*(g-1)/2

soglia<-qchisq(1-alpha,gl)
sogliasimulazione<-h[4]

for(i in 1:nr)
{
  dati<-NULL
  dati<-mvrnorm(n=nc,media,sigma,tol=1e-6,empirical=FALSE)
  t<-trv9.n(dati,num)
  x<-trv9.Bart(dati,num)
  e1<-t/a1
  e2<-t/a2+c

  if(t>soglia)          pvasintotico<-pvasintotico+1
  if(x>soglia)          pvanalitica<-pvanalitica+1
  if(e1>soglia)         pvempirical<-pvempirical+1
  if(e2>soglia)         pvempirical2<-pvempirical2+1
  if(t>sogliasimulazione) pvsimulazione<-pvsimulazione+1
}

pvasintotico<-pvasintotico/nr
pvanalitica<-pvanalitica/nr
pvempirical<-pvempirical/nr
pvempirical2<-pvempirical2/nr
pvsimulazione<-pvsimulazione/nr

h<-c(gl,pvasintotico,pvanalitica,pvempirical,pvempirical2,pvsimulazione)
print(h)
}

#-----

trv9.n<-function(dati,num)
{

n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
test<-0
W<-matrix(rep(0,p^2),p,p)
ba<-1
for(i in 1:g)
{
  ni<-num[i]
  sn<-var(dati[ba:(ni+ba-1),])*ni
  ba<-ba+ni
  test<-test+ni*log(det(sn/ni))
  W<-W+sn
}

test<-n*log(det(W/n))-test
}

#-----

trv9.Bart<-function(dati,num)

```

```

{
# sigma1=...=sigmag

r<-0
n<-dim(dati)[1]
p<-dim(dati)[2]
g<-length(num)
test<-0
W<-matrix(rep(0,p^2),p,p)
ba<-1
for(i in 1:g)
{
ni<-num[i]
sn<-var(dati[ba:(ni+ba-1),])
ba<-ba+ni
test<-test+(ni-1)*log(det(sn))
W<-W+sn*(ni-1)
r<-r+1/(ni-1)
}
r<-r-1/(n-g)
su<-W/(n-g)
testB<-(n-g)*log(det(su))-test

f<-1-((2*p^2+3*p-1)/(6*(p+1)*(g-1)))*r

test<-f*testB
}

#-----

p.e.trv9<-function(nc,p,num,alpha,errore)
{

y1<-NULL
Sy<-0
v<-0
media<-c(rep(0,p))
sigma<-diag(100,p,p)
g<-length(num)
gl<-p*(p+1)*(g-1)/2

N<-round((2.57/errore)^2*alpha*(1-alpha))

for( i in 1:N)
{
dati0<-NULL
dati0<-mvrnorm(n=nc,media,sigma,tol=1e-6,empirical=FALSE)
y<-trv9.n(dati0,num)
y1<-c(y1,y)
Sy<-Sy+y
}

Ey<-Sy/N
v<-var(y1)
a1<-Ey/gl
a2<-sqrt(v/(2*gl))
c<-gl-Ey/a2

y1<-sort(y1)
percentile<-quantile(y1,1-alpha)

h<-c(a1,a2,c,percentile)
}

library(MASS)

```

BIBLIOGRAFIA

- Azzalini, A. (2000). *Inferenza Statistica. Un'introduzione basata sul concetto di verosimiglianza*. Springer-Verlag, Berlin-Heidelberg.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceeding Royal Society, Series A*, **160**, 268-282.
- Bortot P., Salvan A. e Ventura L. (2000). *Inferenza Statistica: Applicazioni con S-Plus e R*. Cedam, Padova.
- Cribari-Neto, F. and Cordeiro, G. M. (1996). On Bartlett-type corrections. *Econometric Reviews*, **15**, 339-367.
- Iacus S. M. e Masarotto G. (2003). *Laboratorio di statistica con R*. McGraw-Hill, Milano.
- Jesen, J. L. (1993). A historical sketch and some new results on the improved Likelihood statistic. *Scandinavian Journal of Statistic*, **20**, 1-15.
- Korin, B. P. (1968). On the distribution of a statistic used for testing a covariance matrix. *Biometrika*, **55**, 171-178
- Landenna G., Marasini D. e Ferrari P. (1998). *La verifica di ipotesi statistiche*. Il Mulino, Bologna.
- Pace L. e Salvan A. (2001). *Introduzione alla statistica*. Cedam, Padova.

Piccolo, D. (1998). *Statistica*. Il Mulino, Bologna.

Mardia, K. V., Kent J. T. e Bibby J. M. (1979). *Multivariate Analysis*.
Academic Press.

Olkin, I. and Tomsy, T. L. (1975). A new class of multivariate tests based on
the union-intersection principle. *Bull. Inst. Statist. Inst.*, **46**, 202-204