Politecnico di Milano
Facoltà di Ingegneria dei Sistemi

ESAME DI STATISTICA APPLICATA

Milano, 8 Febbraio 2006

Indice

1 qw

Esercizio 1

Siano $X_1 \sim N_1(1,2)$ e $X_2 \sim N_1(1,2)$.
Cosa possiamo asserire riguardo alla distribuzione congiunta di $\begin{pmatrix} X_1 + X_2 \\ X_2 \end{pmatrix}$ rispettivamente nel caso in cui:

1. $E[X_1 X_2] = 1$
2. $E[X_1 X_2] = 0$
3. $E[X_1 X_2] = -1$

Esercizio 2

Nella seguente tabella sono riportate altezza, peso ed età dei 5 dottorandi in Ingegneria Matematica del XX ciclo:

Nome	Altezza (cm)	Peso (kg)	Età (anni)
A	160	60	25
B	165	60	28
C	180	65	26
D	185	70	26
E	185	85	26

Si vuole effettuare una clusterizzazione dei dottorandi utilizzando un algoritmo gerarchico agglomerativo di tipo Single-Linkage, utilizzando esclusivamente le variabili Altezza e Peso ed utilizzando la distanza Manhattan per descrivere la dissimilarità tra i dottorandi:

1. Disegnate il dendrogramma corrispondente mettendo bene in evidenza ascisse ad ordinate.
2. Calcolate il coefficiente cofenetico e commentatene il valore.
3. Partendo dalla dichiarazione della matrice dei dati, riportate i comandi R necessari al fine di ottenere il grafico del dendrogramma.

Esercizio 3

Un campione casuale di ampiezza uguale a 30 è estratto da una popolazione normale bivariata di media μ e matrice di covarianza Σ ignote.
La matrice di covarianza campionaria osservata è:

$$S = \begin{pmatrix} 3 & \sqrt{5/4} \\ \sqrt{5/4} & 1 \end{pmatrix}$$

Dato S come sopra, rappresentate come funzione della media campionaria la regione critica dei seguenti test:

1. Test (livello 5%) per la media:

$$H_0 : \mu = \underline{0} \; vs \; H_1 : \mu \neq \underline{0}$$

2. Test simultanei (livello globale 5%) per le componenti della media basato sugli intervalli-T^2:

$$H_{0i} : \mu_i = 0 \; vs \; H_{1i} : \mu_i \neq 0 \text{ con } i = 1, 2$$

3. Test simultanei (livello globale 5%) per le componenti della media basato sugli intervalli di Bonferroni:

$$H_{0i} : \mu_i = 0 \; vs \; H_{1i} : \mu_i \neq 0 \text{ con } i = 1, 2$$

Sebbene in tutti e tre i test il livello nominale sia pari a 5%, sappiamo che il livello reale potrebbe anche non coincidere con quello dichiarato.

4. Ordinate in senso crescente i livelli reali dei test 1, 2 e 3.

5. Cosa possiamo affermare invece riguardo alle potenze dei test 1, 2 e 3?

Giustificate brevemente le risposte alle domande 4 e 5.

Esercizio 4

In un piccolo ospedale viene organizzato un esperimento clinico per testare l'efficacia di un farmaco contro l'ipertensione: il DePression.
Vengono scelti casualmente per l'esperimento 9 pazienti soggetti ad ipertensione. A 6 di questi viene somministrato per un mese il DePression, mentre ai rimanenti un placebo. A fine mese viene misurata ai 9 pazienti la pressione:

Paziente	Pressione	Trattamento
Pz1	100	Farmaco
Pz2	102	Farmaco
⋮	⋮	⋮
Pz6	110	Farmaco
Pz7	130	Placebo
Pz8	127	Placebo
Pz9	120	Placebo

Si decide di effettuare un'analisi della varianza tramite un modello lineare del tipo:

$$Pressione = \beta_0 X_0 + \beta_1 X_1 + \epsilon$$

ove X_0 e X_1 sono variabili dummy opportune e $\epsilon \sim N(0, \sigma^2)$.

Sono proposte tre matrici disegno di rango pieno:

Matrice A	X_0	X_1
Pz1	1	1
Pz2	1	1
$\vdots$	$\vdots$	$\vdots$
Pz6	1	1
Pz7	1	-2
Pz8	1	-2
Pz9	1	-2

Matrice B	X_0	X_1
Pz1	1	1
Pz2	1	1
$\vdots$	$\vdots$	$\vdots$
Pz6	1	1
Pz7	1	0
Pz8	1	0
Pz9	1	0

Matrice C	X_0	X_1
Pz1	1	-1/2
Pz2	1	-1/2
$\vdots$	$\vdots$	$\vdots$
Pz6	1	-1/2
Pz7	1	1
Pz8	1	1
Pz9	1	1

Indichiamo con μ_F il valor medio della pressione per la popolazione che ha ricevuto il farmaco e con μ_P il valor medio della pressione per la popolazione che ha ricevuto il placebo ed esprimiamo μ_F e μ_P in funzione di tre nuovi parametri:

$$\mu_F = \mu + \tau_F$$
$$\mu_P = \mu + \tau_P$$

Per ciascuna delle tre matrici disegno:

1. Si descriva in dettaglio il significato dei coefficienti β_0 e β_1.
2. Si esprimano la relazioni che nei tre casi legano i coefficienti β_0 e β_1 ai parametri μ, μ_F e μ_P mettendo in evidenza anche il vincolo imposto sui parametri τ_F e τ_P.

Presso l'ospedale vengono implementati in R i seguenti comandi (i valori di pressione sono registrati nel file Pressione.txt):

```
> y <- read.table('Pressione.txt')
> y <- as.vector(as.matrix(y))
> x <- c(rep(1,6),rep(0,3))
> fit <- lm(y~x)
> summary(fit)


Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-8.18566  0.02980  0.37587  3.23867  5.76133
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  122.047      2.858  42.706 1.01e-09 ***
x            -15.732      3.500  -4.495  0.00282 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.95 on 7 degrees of freedom
Multiple R-Squared: 0.7427,     Adjusted R-squared: 0.7059
F-statistic:  20.2 on 1 and 7 DF,  p-value: 0.002817

> influence(fit)$hat
        1         2         3         4         5         6
0.1666667 0.1666667 0.1666667 0.1666667 0.1666667 0.1666667
        7         8         9
0.3333333 0.3333333 0.3333333
```

3. Individuate la matrice disegno utilizzata.
4. Fornite le stime di μ, μ_F, μ_P, τ_F e τ_P.
5. Fornite il *p-value* ed interpretate il significato del test:

   $H_0 : \beta_1 = 0 \;\; vs \;\; H_1 : \beta_1 \neq 0$
6. Fornite il *p-value* ed interpretate il significato del test:

   $H_0 : \beta_1 = 0 \;\; vs \;\; H_1 : \beta_1 < 0$
7. Costruite la tabella ANOVA, ovvero fornite i valori di SS_{cor}, SS_{tr} e SS_{res}.
8. Fornite delle stime non distorte delle varianze dei residui $\hat{\epsilon}_1$, $\hat{\epsilon}_2$, $\hat{\epsilon}_3$, $\hat{\epsilon}_4$, $\hat{\epsilon}_5$, $\hat{\epsilon}_6$, $\hat{\epsilon}_7$, $\hat{\epsilon}_8$ e $\hat{\epsilon}_9$. Commentare tali stime.

Esercizio 5

Sia X una variabile aleatoria a valori in $\Re^6$ di legge $N_6(\underline{0}, \Sigma)$ con:

$$\Sigma = \begin{pmatrix} 1 & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\ 1/\sqrt{2} & 1 & 1/\sqrt{2} & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1 & 1/\sqrt{2} \\ 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1 \end{pmatrix}$$

Si individuino le componenti principali e le relative varianze. Si commenti il risultato dell'analisi.