

Politecnico di Milano
Facoltà di Ingegneria dei Sistemi

II APPELLO DI STATISTICA APPLICATA

Milano, 8 Marzo 2006

©I diritti d'autore sono riservati. Ogni sfruttamento commerciale non autorizzato sarà perseguito.

Esercizio 1

Un navigatore satellitare è uno strumento in grado di stimare posizione e velocità di un veicolo sulla base delle informazioni che periodicamente riceve dai satelliti geostazionari che orbitano intorno alla Terra.

L'XNavi06 è attualmente montato su di un veicolo ed è in comunicazione con 4 satelliti identici da ciascuno dei quali riceve ogni minuto una stima delle due coordinate geografiche del veicolo al momento dell'invio del segnale.

L'Xnavi06 conserva in memoria le informazioni relative agli ultimi due invii (esprese in *km* da un determinato punto di riferimento). In particolare alle h 12:31 sono presenti in memoria le informazioni arrivate alle 12:30 ed alle 12:31 (i dati completi sono presenti nel file `sat.txt`):

sat	Lat _{12:30}	Lon _{12:30}	Lat _{12:31}	Lon _{12:31}
sat1	1.45	1.90	2.75	2.10
⋮	⋮	⋮	⋮	⋮
sat4	2.40	2.00	3.00	1.50

In prima approssimazione la variabilità delle misure è attribuibile esclusivamente al rumore elettronico causato dai circuiti dei satelliti. Possiamo quindi assumere che il vettore delle quattro coordinate, ricevute nei due diversi istanti temporali da ciascuno dei quattro satelliti, abbia distribuzione normale e che i vettori di coordinate relativi a satelliti diversi siano indipendenti e identicamente distribuiti:

$$\begin{pmatrix} Lat_{12:30} \\ Lon_{12:30} \\ Lat_{12:31} \\ Lon_{12:31} \end{pmatrix} \sim N_4\left(\begin{pmatrix} \mu_{12:30} \\ \mu_{12:31} \end{pmatrix}, \begin{pmatrix} \Sigma & \tilde{\Sigma} \\ \tilde{\Sigma}' & \Sigma \end{pmatrix}\right)$$

L'XNavi06 alle 12:31 (come ogni minuto) effettua numerose operazioni due delle quali sono riportate in seguito:

- Stima le due componenti della velocità media $\begin{pmatrix} v_{Lat} \\ v_{Lon} \end{pmatrix}$ del veicolo tra le 12:30 e le 12:31 e le riporta sul monitor espresse in *km/h*.
- Effettua un test di livello 10%:

$$H_0: \text{velocità media} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ vs } H_1: \text{velocità media} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

e nel caso di non rifiuto dell'ipotesi H_0 riporta sul monitor la scritta "Veicolo Fermo".

Cosa dovrebbe comparire sul monitor dell'XNavi06 alla 12:31, rispettivamente:

1. se si assumono i rilevamenti relativi alle 12:31 stocasticamente indipendenti dai rilevamenti relativi alle 12:30.
2. se, rispetto al caso precedente, rimuoviamo l'ipotesi di indipendenza tra i rilevamenti provenienti in due minuti successivi dallo stesso satellite.

In seguito sono riportati alcuni comandi R utili per la risoluzione dell'esercizio.

```
> Sat <- read.delim('sat.txt')
> Sat <- data.frame(Sat)
> Sat
```

	Lat_12.30	Lon_12.30	Lat_12.31	Lon_12.31
sat1	1.45	1.90	2.75	2.10
sat2
sat3
sat4	2.40	2.00	3.00	1.50

```
> mean(Sat)
```

Lat_12.30	Lon_12.30	Lat_12.31	Lon_12.31
1.00	1.00	3.70	1.00

```
> cov(Sat)
```

	Lat_12.30	Lon_12.30	Lat_12.31	Lon_12.31
Lat_12.30	1.00	1.00	1.00	0.00
Lon_12.30	1.00	2.00	0.00	1.00
Lat_12.31	1.00	0.00	3.00	1.00
Lon_12.31	0.00	1.00	1.00	4.00

Esercizio 2

Sia $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\mu, \Sigma)$ con $\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$.

Si individuino le distribuzioni condizionate di:

1. $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Big| X_1 + X_2 = c$
2. $X_1 + X_2 \mid X_1 = c$
3. $X_1 + X_2 \mid X_2 = c$

Esercizio 3

Da nove anni una centralina per la misurazione della qualità dell'aria situata in Piazza Piola registra quotidianamente il livello di PM10 nell'aria. Un termometro situato sempre in Piazza Piola registra invece la temperatura minima raggiunta nella notte precedente. Nella tabella seguente sono riportate le misure relative al giorno 8 marzo:

Anno	1997	1998	1999	2000	2001	2002	2003	2004	2005
Temperatura	-1	0	1	1	2	-1	3	0	1
PM10	Alto	Medio	Alto	Basso	Medio	Alto	Basso	Alto	Alto

L'Istituto di Meteorologia dell'Aeroporto di Linate afferma inoltre che le seguenti ipotesi possono essere ritenute vere:

- Temperatura | PM10 = Alto $\sim N(\mu_{Alto}, \sigma^2)$
- Temperatura | PM10 = Medio $\sim N(\mu_{Medio}, \sigma^2)$
- Temperatura | PM10 = Basso $\sim N(\mu_{Basso}, \sigma^2)$
- $\pi(\text{PM10} = \text{Alto}) = \pi(\text{PM10} = \text{Medio}) = \pi(\text{PM10} = \text{Basso})$

Si vuole costruire un criterio discriminante che, sulla base della temperatura minima registrata nella notte, permetta di prevedere, minimizzando la probabilità di errore, il livello di PM10 nell'aria nel giorno successivo:

1. Si calcoli l'espressione analitica dei *discriminant scores* e se ne plotti l'andamento su di un grafico.
2. Si individuino le regioni R_{Alto} , R_{Medio} ed R_{Basso} .
3. Si calcoli l'*Apparent Error Rate* (APER). Attenzione alle *prior*!
4. La temperatura minima raggiunta stanotte in Piazza Piola è stata 1.25 °C.
Che livello di PM10 prevediamo per la giornata odierna?
Qual è invece il livello di PM10 meno probabile?

Esercizio 4

Nel dataframe **skate** sono contenuti alcuni dei tempi (espressi in secondi) registrati alle ultime Olimpiadi Invernali di Torino 2006 nella gare di pattinaggio su ghiaccio - 500m. In seguito sono riportati alcuni output di R con alcune quantità omesse.

1. Si riportino sul foglio protocollo le 4 tabelle ANOVA opportunamente completate delle parti mancanti.
2. Si effettui il test al 10%:

$$\begin{aligned} H_0: & \text{ non c'è interazione tra i fattori Nazione e Sesso} \\ & \text{vs} \\ H_1: & \text{ c'è interazione tra i fattori Nazione e Sesso} \end{aligned}$$

3. Si calcolino gli R^2 relativi ai quattro modelli.
4. Sulla base di quanto calcolato nei punti precedenti si individui il modello che ritenete più adatto a spiegare la variabile Tempo, giustificando la scelta fatta.

Output di R:

```
> skate
```

	Atleta (Età)	Nazione	Tempo	Sesso
1	Joey Cheek (26)	USA	69.76	M
2	Kang-Seok Lee (20)	KOR	70.43	M
3	Yuya Oikawa (25)	JPN	70.56	M
4	Fengtong Yu (21)	CHN	70.68	M
5	Joji Kato (21)	JPN	70.78	M
6	Mike Ireland (32)	CAN	70.88	M
7	Jae-Bong Choi (25)	KOR	71.04	M
8	Jeremy Wotherspoon (29)	CAN	71.05	M
...
...
33	Seung-Yong Choi (26)	KOR	79.02	F
34	Yu-Rim Kim (16)	KOR	79.25	F
35	Kerry Simpson (24)	CAN	79.34	F
36	Krisy Myers (27)	CAN	79.43	F
37	Elli Ochowicz (22)	USA	79.48	F
38	Bo-Ra Lee (19)	KOR	79.73	F
39	Kim Weger (25)	CAN	79.99	F
40	Chris Witty (30)	USA	80.69	F

```

> mean(skate$Tempo)
[1] 75.5055

> var(skate$Tempo)
[1] 23.27928

> table(skate$Nazione, skate$Sesso)

      F M
CAN   4 4
CHN   4 4
JPN   4 4
KOR   4 4
USA   4 4

> fit1 <- aov(Tempo ~ Nazione * Sesso, data=skate)
> fit2 <- aov(Tempo ~ Nazione + Sesso, data=skate)
> fit3 <- aov(Tempo ~ Nazione, data=skate)
> fit4 <- aov(Tempo ~ Sesso, data=skate)

```

```
> fit1
Call:
aov(formula = Tempo ~ Nazione * Sesso, data = skate)

Terms:
              Nazione      Sesso Nazione:Sesso Residuals
Sum of Squares      ...      ...      ...      ...
Deg. of Freedom       4       1       4       30
```

Residual standard error: 3.868797

```
> fit2
Call:
aov(formula = Tempo ~ Nazione + Sesso, data = skate)
```

```
Terms:
              Nazione      Sesso Residuals
Sum of Squares      ...      ...      ...
Deg. of Freedom       4       1       34
```

Residual standard error: ...

```
> fit3
Call:
aov(formula = Tempo ~ Nazione, data = skate)
```

```
Terms:
              Nazione Residuals
Sum of Squares      ...      ...
Deg. of Freedom       4       35
```

Residual standard error: 4.915629

```
> fit4
Call:
aov(formula = Tempo ~ Sesso, data = skate)
```

```
Terms:
              Sesso Residuals
Sum of Squares      ...      ...
Deg. of Freedom       1       38
```

Residual standard error: 3.839764