Modelli Statistici 1	
A.A. 2002-2003	**Homework 4**

Esercizio 4.1. Si riconsiderino i dati contenuti nel file `cherry.dat` (si veda il laboratorio 6), riferiti a misurazioni rilevate su alberi di ciliegio. Le variabili sono rispettivamente: diametro (a 4.5 piedi dal suolo, misurato in pollici), altezza (in piedi), volume (in piedi cubici di legname utile). Siano $y_i = \log(\text{volume}_i)$, $x_{i2} = \log(\text{diametro}_i)$ e $x_{i3} = \log(\text{altezza}_i)$ $i = 1, \ldots, n$. Si consideri il modello di regressione:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i\,, \qquad i = 1, \ldots, n$$

con ϵ_i variabili casuali indipendenti $N(0, \sigma^2)$ e $x_{i1} = 1$, $i = 1, \ldots, n$.

(a) Si calcoli la statistica del test per la verifica dell'ipotesi $H_0 : \beta_2 = \beta_3 = 0$ utilizzando il coefficiente di determinazione multiplo R^2. Che decisione viene presa nei confronti dell'ipotesi H_0?

(b) Si fornisca la statistica test per la verifica dell'ipotesi $H_0 : \beta_2 = 2, \beta_3 = 1$ e se ne indichi la distribuzione esatta.

(c) Si scrivano i comandi `R` per adattare il modello:

$$Y_i = \beta_1 x_{i1} + 2x_{i2} + x_{i3} + \epsilon_i\,, \qquad i = 1, \ldots, n.$$

(d) Dopo aver adattato il modello di cui al punto (c), si verifichi l'ipotesi $H_0 : \beta_2 = 2, \beta_3 = 1$.

(e) Si calcoli il livello di significatività osservato per il test di cui al punto (d).

Esercizio 4.2. In un ospedale italiano è stato rilevato il peso di neonati in 189 nascite ed è stata fatta una classificazione in base alle abitudine al fumo della madre. Il vettore `peso.fumo` contiene il peso (in grammi) di 74 neonati con madre fumatrice e il vettore `peso.non.fumo` il peso (in grammi) di 115 neonati con madre non fumatrice. Si commenti l'output di `R` di seguito riportato:

```
>  t.test(peso.fumo,peso.non.fumo)

Standard Two-Sample t-Test

data:  peso.fumo and peso.non.fumo
```

```
t = -2.7729, df = 187, p-value = 0.0061
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -413.2819  -69.6875
sample estimates:
 mean of x mean of y
  2824.173  3065.658
```

Nella Figura 1 sono contenuti i Q-Q plot normali per le osservazioni contenute in `peso.fumo` e `peso.non.fumo`. Si commentino i due grafici e si dica in quale modo gli andamenti osservati possono mettere in discussione le assunzioni che sono alla base del test precedente.

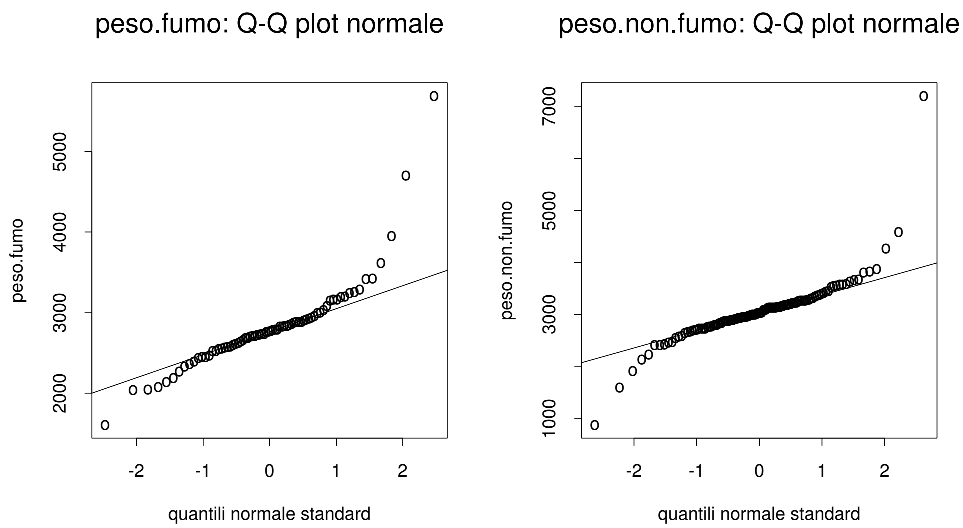

Figura 1: Q-Q plot normale per `peso.fumo` e `peso.non.funo` con retta di riferimento passante per il primo e il terzo quartile.

Esercizio 4.3. Un produttore di sigarette ha inviato ad un laboratorio tre campioni di tabacco. Il laboratorio ha effettuato per ciascun campione cinque osservazioni sul contenuto di nicotina in *mg/10g*, contenute in `s1`, `s2` e `s3`:

```
> s1<-c(24,27,26,21,24)
> s2<-c(27,28,23,31,6)
> s3<-c(28,30,29,27,21)
```

Assumendo che i contenuti di nicotina seguano tre distribuzioni normali indipendenti con medie μ_1, μ_2, μ_3, rispettivamente, e con la stessa ignota varianza, per verificare l'ipotesi $H_0 : \mu_1 = \mu_2 = \mu_3$ il laboratorio intende applicare l'analisi della varianza.

(a) Si scriva la matrice dei dati, dopo aver opportunamente codificato il fattore relativo ai 3 gruppi.

(b) Mediante l'utilizzo di `R`, il laboratorio ha calcolato le seguenti quantità:

$$\sum_{j=1}^{3}\sum_{i=1}^{5}(y_{ij}-\bar{y}_j)^2 = 465.20, \qquad 5\sum_{j=1}^{3}(\bar{y}_j-\bar{y})^2 = 41.20.$$

Si dica se il laboratorio accetta o rifiuta l'ipotesi H_0 al livello 0.05.

(c) Si supponga che lo strumento utilizzato per misurare la nicotina fosse erroneamente tarato e che abbia aggiunto a tutte le misurazioni 1 milligrammo. Correggendo le misurazioni, cambiano le conclusioni della verifica di ipotesi?