Modelli Statistici 1

A.A. 2002-2003 **Homework 3**

Esercizio 3.1. Si consideri il seguente modello di regressione lineare:

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

con $\epsilon_1, \ldots, \epsilon_n$ variabili casuali indipendenti $N(0, \sigma^2)$. La stima di massima verosimiglianza di $(\beta_1, \beta_2, \beta_3)$ si ottiene minimizzando $SQ(\beta_1, \beta_2, \beta_3) = \sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_{i2} - \beta_3 x_{i3})^2$ rispetto a $(\beta_1, \beta_2, \beta_3)$. Si verifichi che la condizione necessaria per un minimo, ossia l'annullamento delle derivate parziali prime di $SQ(\beta_1, \beta_2, \beta_3)$ rispetto a β_1, β_2 e β_3, dà le equazioni normali dei minimi quadrati. Si mostri che il punto stazionario è un minimo.

Esercizio 3.2. Nella città di New York, per 111 giorni consecutivi, sono state rilevate le variabili concentrazione di ozono (in parti per bilione), y_i, radiazione solare, x_{i2}, temperatura (in gradi Fahrenheit), x_{i3}, e velocità del vento (in miglia orarie), x_{i4}, $i = 1, \ldots, 111$. Si assuma il modello di regressione lineare multipla, ovvero che $(y_1, \ldots, y_{111})$ siano realizzazioni delle variabili casuali

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i\,,$$

con ϵ_i i.i.d. $N(0, \sigma^2)$, $i = 1, \ldots, 111$. La stima di massima verosimiglianza della varianza dell'errore è risultata pari a $\hat{\sigma}^2 = 0.2862$. Per verificare la significatività dell'influenza della radiazione solare (x_{i2}), si è poi adattato un modello ridotto con $\beta_2 = 0$, ottenendo una stima di massima verosimiglianza pari a $\tilde{\sigma}^2 = 0.2874$.

1 Si valuti l'attendibilità dell'ipotesi che la radiazione solare non abbia influenza sulla concentrazione di ozono, a parità di temperatura e velocità del vento.

2 È noto che $(X^T X)^{-1}_{22} = 0.12 \times 10^{-5}$. Si determini un intervallo di confidenza per β_2 di livello esatto 0.95. Si commenti il risultato ottenuto.

Esercizio 3.3. Si consideri il modello di regressione lineare multipla $Y = X\beta + \varepsilon$, con X matrice $n \times p$ di costanti note con rango p, β vettore $p \times 1$ di parametri ignoti e $\varepsilon \sim N_n(0, \sigma^2 I_n)$, con $\sigma^2 > 0$. Il vettore $\hat{y}$ dei valori stimati è $\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y$, ossia è $\hat{y} = Py$ con $P = X(X^T X)^{-1} X^T$ matrice di proiezione ortogonale nel sottospazio vettoriale $\mathcal{V}$ di $\mathbb{R}^n$ generato dalle colonne di X.

1 Si mostri che la matrice P è simmetrica.

2 Si mostri che $P^2 = P\,P = P$, ossia che la matrice P è idempotente.

3 Si mostri che anche la matrice $I_n - P$ è simmetrica e idempotente.

4 Si spieghi perché, se $z \in \mathcal{V}$, deve valere $Pz = z$.

5 Sia X_0 la matrice $n \times p_0$, $p_0 < p$, formata dalle prime p_0 colonne di X. Che cosa rappresenta la matrice $P_0 = X_0(X_0^T X_0)^{-1} X_0^T$? Si mostri che la matrice P_0 è simmetrica e idempotente.

6 Si mostri che anche la matrice $P - P_0$ è simmetrica e idempotente.