

Corsi di Laurea Specialistica in Scienze Ambientali
 Corso di Metodi probabilistici e statistici per l'analisi dei dati a.a. 2005-2006
 Prova scritta del 22/09/2006

1. I dati contenuti in N02 (nel sito del corso) si riferiscono a 500 delle osservazioni originate dallo studio dell'inquinamento di una strada, causato dal traffico, in relazione alla situazione meteorologica (*Norwegian Public Roads Administration*). La variabile risposta (y) si riferisce a valori orari del logaritmo della concentrazione di NO_2 , nel periodo 10/2001 - 8/2003. Le variabili predittrici sono i logaritmi di: numero di auto all'ora (x_1), temperatura a due metri dal suolo in gradi (x_2), velocità del vento (metri/sec, x_3), differenza di temperatura tra 25 e due metri dal suolo (x_4), direzione del vento (in gradi, x_5), ora del giorno (x_6), giorno del periodo (x_7),

y	x_1	x_2	x_3	x_4	x_5	x_6	x_7
3.71844	7.6912	9.2	4.8	-0.1	74.4	20	600
3.10009	7.69894	6.4	3.5	-0.3	56	14	196
3.31419	4.81218	-3.7	0.9	-0.1	281.3	4	513
4.38826	6.95177	-7.2	1.7	1.2	74	23	143
\vdots							

Fare un'analisi di regressione lineare per la dipendenza di concentrazione di NO_2 dalle emissioni: determinare una stima del modello e commentare sulla bontà del modello con gli strumenti noti. Stimare il valore di concentrazione di NO_2 per valori delle variabili predittrici di (6.2, 6.1, 1.1, 0.8, 100, 22 560). Valutare infine l'ipotesi che l'ora del giorno non influenzi la concentrazione di NO_2 .

2. Si considerino le $n = 35$ osservazioni relative all'insetto *Leptoconops Torrens* (Tabella T6-15.DAT) relative alla lunghezza ed alla larghezza dell'ala (variabili x_1, x_2).
- (a) Fare il diagramma di dispersione per il campione della popolazione e calcolare la matrice di correlazione. Commentare
 - (b) Valutare la normalità dei dati con le tecniche note ed eventualmente usare trasformazioni per migliorare la normalità.
 - (c) Determinare la regione (ellisse) di confidenza al 95% per la *media* delle due variabili. Determinare inoltre gli intervalli simultanei (T^2) di confidenza al 95% per la media delle due variabili.
 I valori $\mu^T = (98, 44)$ sono accettabili come valori medi della lunghezza e larghezza dell'ala?
 - (d) Si consideri quindi anche il campione di $n = 35$ individui della famiglia *L. carteri* (seconda famiglia). Fare un test di uguaglianza delle medie, con un livello di significatività del 5%. Determinare la combinazione lineare di massima differenza.