Regression and Prediction

H.M. James Hung

An Application Problem

Companies considering the purchase of a computer must first assess their future needs in order to determine the proper equipment. A computer scientist collected data from seven similar company sites so that a forecast equation of computer-hardware requirements for inventory management could be developed.

The data are given in next slide.

z_1	z_2	y
123.5	2.108	141.5
146.1	9.213	168.9
133.9	1.905	154.8
128.5	0.815	146.5
151.5	1.061	172.8
136.2	8.603	160.1
92.0	1.125	108.5

y = CPU time (in hours)

z_1 = Orders (in thousands)

z_2 = Add-delete item count (in thousands)

Source: Data taken from H.P. Artis, Forecasting Computer Requirements: A Forecaster's Dilemma (Piscataway, NJ: Bell Laboratories, 1979) from "Applied Multivariate Statistical Analysis, by Johnson and Wichern, 6th Edition", p.380

A linear regression model is fitted to y on z_1 and z_2.
Use SAS to obtain the OLS estimates of regression coefficients are

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (Z'Z)^{-1}Z'Y = \begin{pmatrix} 8.42 \\ 1.08 \\ 0.42 \end{pmatrix}$$

$$\hat{y} = 8.42 + 1.08z_1 + 0.42z_2$$

$$(Z'Z)^{-1} = \begin{bmatrix} 8.18 & -0.064 & 0.08831 \\ -0.064 & 0.00052 & -0.00107 \\ 0.08831 & -0.00107 & 0.01440 \end{bmatrix}$$

$\hat{\sigma}^2 = \sum_{i=1}^{7}(y_i - \hat{y}_i)^2/(7-3)$ $\hat{\sigma}$ = 1.204

The 95% CI for the mean CPU time at (z_1, z_2) = (130, 7.5) is

$$(8.42 + 1.08 \times 130 + 0.42 \times 7.5) \pm t_{0.025,4} \times \hat{\sigma}\sqrt{A} \quad ,$$

where $t_{0.025,4}$ = 2.776, and

$$A = (1,\ 130,\ 7.5)(Z'Z)^{-1}(1,\ 130,\ 7.5)'$$

However, the 95% prediction interval for the CPU time at (z_1, z_2) = (130, 7.5) is

$$(8.42 + 1.08 \times 130 + 0.42 \times 7.5) \pm t_{0.025,4} \times \hat{\sigma}\sqrt{1 + A} \quad ,$$

where $t_{0.025,4}$ = 2.776, and

$$A = (1,\ 130,\ 7.5)(Z'Z)^{-1}(1,\ 130,\ 7.5)'$$

In the regression fit, calculation of CI and prediction interval, what are the assumptions I made?