Multiple Linear Regression Model – Part VI

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 4 – Lecture 4D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Simultaneous CIs on Regression Coefficients

The CIs for individual regression coefficients discussed previously are one-at-a-time intervals, that is, the confidence coefficient $(1 - \alpha)$ indicates the proportion of correct statements that results when repeated random samples are selected and the appropriate interval estimate is constructed for each sample.

Simultaneous CIs on Regression Coefficients

Some problems require that several confidence intervals be constructed using the same sample data. In these cases, the analyst is usually interested in specifying a confidence coefficient that applies simultaneously to the entire set of interval estimates

Simultaneous CIs on Regression Coefficients

A set of confidence (or prediction) intervals that are all true simultaneously with probability $(1 - \alpha)$ are called simultaneous or joint confidence intervals.

If 95% CIs for two individual regression coefficients are used and independent, then the probability that both statements are correct is $(0.95)^2 = 0.9025$.

Simultaneous CIs on Regression Coefficients

A simple way to construct joint CI:

$$\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{pMS_{\text{Res}}} \sim F_{p,n-p} \quad \Rightarrow \quad P\left\{\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{pMS_{\text{Res}}} \le F_{\alpha,n-p}\right\} = 1-\alpha$$

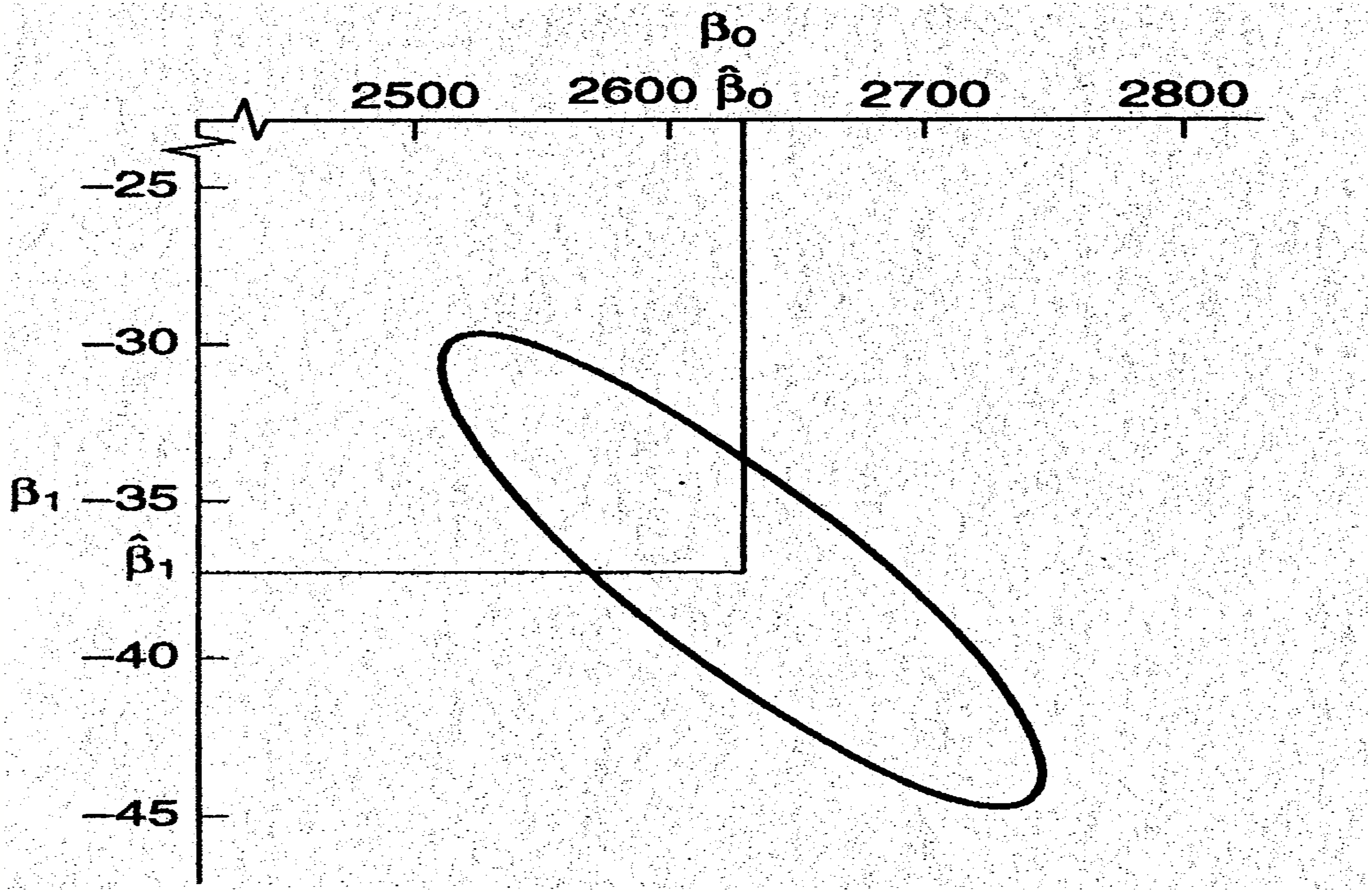
Joint Confidence Region: $$\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{pMS_{\text{Res}}} \le F_{\alpha,p,n-p}$$

The Rocket Propellant Data (Ex 3.10, page 101 of Textbook)

$$\frac{n\left(\hat{\beta}_0-\beta_0\right)^2+2\sum_{i=1}^{n}x_i\left(\hat{\beta}_0-\beta_0\right)\left(\hat{\beta}_1-\beta_1\right)+\sum_{i=1}^{n}x_i^2\left(\hat{\beta}_1-\beta_1\right)^2}{2MS_{\mathrm{Res}}}\leq F_{\alpha,2,n-2}$$

$$\left[20(2627.82-\beta_0)^2+2(267.25)(2627.82-\beta_0)(-37.15-\beta_1)\right.$$
$$\left.+(4677.69)(-37.15-\beta_1)^2\right]/[2(9244.59)]=3.55$$

The Rocket Propellant Data (Ex 3.10, page 101 of Textbook)



Simultaneous CIs on Regression Coefficients

Another procedure is the Bonferroni method applying on s regression coefficients:

change $t_{\alpha/2,n-p}$ to $t_{\alpha/(2s),n-p}$

when constructing individual CIs

The Rocket Propellant Data (Ex 3.11, page 103 of Textbook)

$$\hat{\beta}_0 = 2627.822, \quad \text{se}\left(\hat{\beta}_0\right) = 44.184$$

$$\hat{\beta}_1 = -37.154, \quad \text{se}\left(\hat{\beta}_1\right) = 2.889$$

The joint 95% CI using Bonferroni's are

$$\hat{\beta}_0 - t_{0.0125,18}\,\text{se}\left(\hat{\beta}_0\right) \le \beta_0 \le \hat{\beta}_0 + t_{0.0125,18}\,\text{se}\left(\hat{\beta}_0\right)$$

$$2627.822 - (2.445)(44.184) \le \beta_0 \le 2627.822 + (2.445)(44.184)$$

$$\hat{\beta}_1 - t_{0.0125,18}\,\text{se}\left(\hat{\beta}_1\right) \le \beta_1 \le \hat{\beta}_1 - t_{0.0125,18}\,\text{se}\left(\hat{\beta}_1\right)$$

$$-37.154 - (2.445)(2.889) \le \beta_1 \le -37.154 + (2.445)(2.889)$$

Multicollinearity

Take an example for “exact-linear dependence”:

$$\begin{bmatrix} 10 \\ 13 \\ 15 \\ 16 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 1 & 2 \\ 3 & 6 \\ 4 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$$

$$\mathbf{X'X} = \begin{bmatrix} 30 & 60 \\ 60 & 120 \end{bmatrix} \quad \textbf{not invertible}$$

$$\Rightarrow \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad \text{does not exist}$$

Multicollinearity

"Near-linear dependence":

$$\text{Variance of LS estimator of } \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \sigma^2(\mathbf{X'X})^{-1},$$

but $(\mathbf{X'X})^{-1}$ can be ill-conditioned; the variance of regression coefficient can be very large.

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING