Prediction Intervals

Section 8.6, esp pp. 213-215

11.1

Denote a <u>row</u> of the X matrix by the column vector $\underset{\sim}{x}_i$. Then the regression model could be written $Y_i = \underset{\sim}{x}_i' \beta + \varepsilon_i$

Suppose we have a new vector of IV observations. Book calls it $\underset{\sim}{x}_0$. I would call it x_{n+1}. Then

$$Y_0 = x_0' \beta + \varepsilon_0 \quad , \quad \varepsilon_0 \sim N(0, \sigma^2) \text{ \& }$$

independent of $\varepsilon_1, \ldots, \varepsilon_n$

$E(Y_0) = x_0' \beta$, and we can easily get a <u>CONFIDENCE INTERVAL</u>

$$x_0' \hat{\beta} - x_0 \beta \sim N(0, \sigma^2 x_0' (X'X)^{-1} x_0)$$

$$Z = \frac{x_0' \hat{\beta} - x_0 \beta}{\sqrt{\sigma^2 x_0' (X'X)^{-1} x_0}} \quad , \quad W = \frac{SSE}{\sigma^2} \quad \text{ind}$$

$$T = \frac{Z}{\sqrt{W/(n-k-1)}} = \frac{x_0' \hat{\beta} - x_0 \beta}{\sqrt{\cancel{\sigma^2} x_0' (X'X)^{-1} x_0 \dfrac{SSE}{\cancel{\sigma^2}(n-k-1)}}}$$

11.2

$$T = \frac{x_0'\hat{\beta} - x_0\beta}{\sqrt{x_0'(X'X)^{-1}x_0 \frac{SSE}{n-k-1}}}$$

$$= \frac{x_0'\hat{\beta} - x_0\beta}{s\sqrt{x_0'(X'X)^{-1}x_0}}$$

$1 - \alpha = P(-t_{\alpha/2} < T < t_{\alpha/2})$ etc.

Now suppose we want to PREDICT Y_{n+1} (keep calling it Y_0)

Best guess is surely $x_0'\hat{\beta}$, but we want an INTERVAL around it, a margin of error

CI captures $E(Y_0)$ with prob $1-\alpha$

We want to capture Y_0 - It's a different problem, though related

$Y_0 \sim N(x_0'\beta, \sigma^2)$

and

$x_0'\hat{\beta} \sim N(x_0'\beta, \sigma^2 x_0'(X'X)^{-1}x_0)$

independent, so

11.3

$$Y_0 - x_0'\hat{\beta} \sim N(0, \sigma^2 + \sigma^2 x_0'(X'X)^{-1}x_0)$$

AND

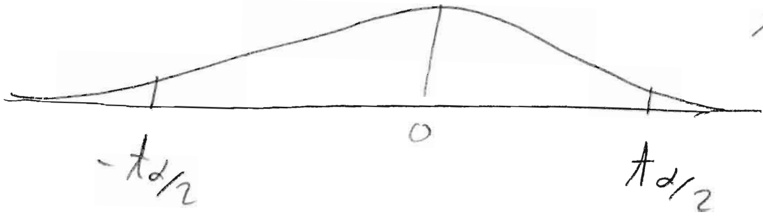
$$Z = \frac{Y_0 - x_0'\hat{\beta}}{\sqrt{\sigma^2(1 + x_0'(X'X)^{-1}x_0)}} \sim N(0,1)$$

Independent of $W = \dfrac{SSE}{\sigma^2}$, so

$$T = \frac{Z}{\sqrt{W/(n-k-1)}} = \frac{Y_0 - x_0'\hat{\beta}}{\sqrt{\sigma^2(1 + x_0'(X'X)^{-1}x_0)}\ \underbrace{\sqrt{\frac{SSE}{\sigma^2}/(n-k-1)}}_{s^2}}$$

$$= \frac{Y_0 - x_0'\hat{\beta}}{s\sqrt{1 + x_0'(X'X)^{-1}x_0}} \sim t(n-k-1)$$

11.4



t dist
with $n-k-1$ df

$$1-\alpha = P(-t_{\alpha/2} < T < t_{\alpha/2})$$

$$= P\left(-t_{\alpha/2} < \frac{Y_0 - x_0'\hat{\beta}}{s\sqrt{1 + x_0 (X'X)^{-1} x_0}} < t_{\alpha/2}\right)$$

$$= P\left(-t_{\alpha/2}\, s\sqrt{\quad} < Y_0 - x_0'\hat{\beta} < t_{\alpha/2}\, s\sqrt{\quad}\right)$$

$$= P\left(x_0'\hat{\beta} - t_{\alpha/2}\, s\sqrt{\quad} < Y_0 < x_0'\hat{\beta} + t_{\alpha/2}\, s\sqrt{\quad}\right)$$ or

$$x_0'\hat{\beta} \pm t_{\alpha/2}\, s\sqrt{1 + x_0 (X'X)^{-1} x_0}$$