620-371: Linear Models

Practice Class 6

7th April, 2009

1. Derive the formula for the confidence interval of the individual parameter β_0 from the formula for the confidence interval of a linear combination of parameters:
$$\mathbf{t}^T\mathbf{b} \pm t_{\alpha/2} s \sqrt{\mathbf{t}^T (X^T X)^{-1} \mathbf{t}}.$$

   Solution: We set
$$\mathbf{t} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
   so that $\mathbf{t}^T\mathbf{b} = b_0$. Then $\mathbf{t}^T(X^TX)^{-1}\mathbf{t}$ is equal to the top left element of $(X^TX)^{-1}$. But this is c_{00} in our notation, so the confidence interval for β_0 is
$$b_0 \pm t_{\alpha/2} s \sqrt{c_{00}}.$$

2. We model the energy consumption of a household in terms of the household income. The data we collect is:

Income ($ k)	Energy consumption (×10 Btu/yr)
20	1.8
30	3.0
40	4.8
55	5.0
60	6.5

   - Find a 95% confidence interval for the average energy consumption of households with yearly income $50,000. You may use $t_{0.025} = 3.182$ for 3 degrees of freedom.

     Solution:
$$X^TX = \begin{bmatrix} 5 & 205 \\ 205 & 9525 \end{bmatrix}, X^T\mathbf{y} = \begin{bmatrix} 21.1 \\ 983.0 \end{bmatrix}$$
$$(X^TX)^{-1} = \begin{bmatrix} 1.701 & -0.036 \\ -0.036 & 0.00089 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -0.096 \\ 0.105 \end{bmatrix}$$

$$s^2 = 0.359$$

The confidence interval is

$$\begin{bmatrix} 1 & 50 \end{bmatrix} \begin{bmatrix} -0.096 \\ 0.105 \end{bmatrix} \pm 3.182\sqrt{0.359}\sqrt{\begin{bmatrix} 1 & 50 \end{bmatrix} \begin{bmatrix} 1.701 & -0.036 \\ -0.036 & 0.00089 \end{bmatrix} \begin{bmatrix} 1 \\ 50 \end{bmatrix}} = (4.17, 6.16).$$

- Find a 95% prediction interval for the energy consumption of a randomly selected household with yearly income \$50,000.

  Solution:

  The prediction interval is

$$\begin{bmatrix} 1 & 50 \end{bmatrix} \begin{bmatrix} -0.096 \\ 0.105 \end{bmatrix} \pm 3.182\sqrt{0.359}\sqrt{1 + \begin{bmatrix} 1 & 50 \end{bmatrix} \begin{bmatrix} 1.701 & -0.036 \\ -0.036 & 0.00089 \end{bmatrix} \begin{bmatrix} 1 \\ 50 \end{bmatrix}} = (3.02, 7.32).$$

3. Prove that

$$\frac{(\mathbf{b} - \boldsymbol{\beta})^T X^T X (\mathbf{b} - \boldsymbol{\beta})}{\sigma^2}$$

   has a χ^2 distribution with p degrees of freedom. *(Hint: you will need a corollary on the distribution of a quadratic form).*

   Solution: $\mathbf{b} - \boldsymbol{\beta}$ is a $p \times 1$ normal random vector with mean $\mathbf{0}$ and variance $V = (X^T X)^{-1}\sigma^2$. Therefore $(\mathbf{b} - \boldsymbol{\beta})^T V^{-1} (\mathbf{b} - \boldsymbol{\beta})$ has a χ^2 distribution with p degrees of freedom and noncentrality parameter $\frac{1}{2}\mathbf{0}^T V^{-1} \mathbf{0} = 0$. But $V^{-1} = \frac{1}{\sigma^2} X^T X$, so the result follows.

4. Using the data from question 2, find a joint 95% confidence region for the two parameters β_0 and β_1. You may keep your answer as an implicit inequality, and use $f_{0.05} = 9.55$ for 2 and 3 degrees of freedom.

   Solution:

$$\begin{aligned} (\mathbf{b} - \boldsymbol{\beta})^T X^T X (\mathbf{b} - \boldsymbol{\beta}) &\leq ps^2 f_\alpha \\ \begin{bmatrix} -0.096 - \beta_0 & 0.105 - \beta_1 \end{bmatrix} \begin{bmatrix} 5 & 205 \\ 205 & 9525 \end{bmatrix} \begin{bmatrix} -0.096 - \beta_0 \\ 0.105 - \beta_1 \end{bmatrix} &\leq 2 \times 0.359 \times 9.55 \\ 5\beta_0^2 + 9525\beta_1^2 + 410\beta_0\beta_1 - 42.1\beta_0 - 1960.9\beta_1 + 100.9 &\leq 6.86. \end{aligned}$$

5. We can write SS_{Reg} and SS_{Res} as

$$SS_{Reg} = \mathbf{y}^T X (X^T X)^{-1} X^T \mathbf{y}$$

   and

$$SS_{Res} = \mathbf{y}^T [I - X(X^T X)^{-1} X^T] \mathbf{y}.$$

Show that these two quadratic forms are independent.

Solution: From the theorem for independence of quadratic forms,

$$
\begin{aligned}
AVB &= X(X^TX)^{-1}X^T\sigma^2 I[I - X(X^TX)^{-1}X^T] \\
&= \sigma^2 \left(X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T \right) \\
&= \sigma^2 \left(X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^T \right) \\
&= 0.
\end{aligned}
$$

Therefore the quadratic forms are independent.

6. Using the data from question 2, test the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

Solution:

$$SS_{Res} = 1.08, SS_{Reg} = 101.5.$$

Our test statistic is

$$\frac{SS_{Reg}/p}{SS_{Res}/(n-p)} = 141.3 > 9.55.$$

Therefore, we reject the model hypothesis — the model is adequate.