620-371: Linear Models

Practice Class 11

19th May, 2009

1. Let $H_0 : C\boldsymbol{\beta} = \mathbf{0}$ be testable. Show that $C(X^TX)^cC^T$ is unique, i.e. does not depend on the choice of conditional inverse for X^TX. (*Hint: Let $(X^TX)^c_1$ be another conditional inverse for X^TX and show that $C(X^TX)^cC^T = C(X^TX)^c_1C^T$. Remember that since H_0 is testable, $C(X^TX)^cX^TX = C$ for any conditional inverse of X^TX*).

   Solution:

$$\begin{aligned} C(X^TX)^cC^T &= C(X^TX)^c_1X^TX(X^TX)^cC^T \\ &= C(X^TX)^c_1[C(X^TX)^cX^TX]^T \\ &= C(X^TX)^c_1C^T. \end{aligned}$$

2. An industrial psychologist is investigating absenteeism among production-line workers, based on different types of work hours: (1) 4-day week with a 10-hour day, (2) 5-day week with a flexible 8-hour day, and (3) 5-day week with a structured 8-hour day. A study was conducted and the following data obtained of the average number of days missed:

	Work plan		
	1	2	3
Mean	9	6.2	10.1
Number	100	85	90

   We also have $s^2 = 110.15$. Test the hypothesis that the work plan has no effect on the absenteeism. The 95% critical value for an F distribution with 2 and 272 degrees of freedom is 3.03.

   Solution:

$$\begin{aligned} C &= \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \\ C\mathbf{b} &= \begin{bmatrix} 2.8 \\ -3.9 \end{bmatrix} \\ C(X^TX)^cC^T &= \begin{bmatrix} 0.0218 & -0.0118 \\ -0.0118 & 0.0229 \end{bmatrix} \\ \frac{(C\mathbf{b})^T[C(X^TX)^cC^T]^{-1}C\mathbf{b}/2}{s^2} &= 3.2. \end{aligned}$$

   Therefore we reject the null hypothesis.

3. For the one-way classification model, given the expressions for $X^T\mathbf{y}$ and $(X^TX)^c$ in the lecture slides, show that

$$SS_{Reg} = \sum_{i=1}^{k} (\bar{y}_i)^2 n_i.$$

Solution:

$$\begin{aligned}
SS_{Reg} &= \mathbf{y}^T X (X^T X)^c X^T \mathbf{y} \\
&= (X^T\mathbf{y})^T \mathbf{b} \\
&= \begin{bmatrix} \sum_{ij} y_{ij} & \sum_j y_{1j} & \sum_j y_{2j} & \dots & \sum_j y_{kj} \end{bmatrix} \begin{bmatrix} 0 \\ \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_k \end{bmatrix} \\
&= \sum_i \left(\sum_j y_{ij} \bar{y}_i \right) \\
&= \sum_i (\bar{y}_i)^2 n_i.
\end{aligned}$$

4. We are interested in looking at the effect of breed and diet on the milk yield of cows. A study is conducted and the following data obtained:

	Diet		
Breed	1	2	3
1	18.8	16.7	19.8
	21.2		23.9
2	22.3	15.9	21.8
		19.2	

(a) Express this as a two-factor model with no interaction in matrix form.
Solution: $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{y} = \begin{bmatrix} 18.8 \\ 21.2 \\ 16.7 \\ 19.8 \\ 23.9 \\ 22.3 \\ 15.9 \\ 19.2 \\ 21.8 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

and $\boldsymbol{\varepsilon}$ is as expected.

(b) Express this as a two-factor model with interaction in matrix form.

Solution: $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{y} = \begin{bmatrix} 18.8 \\ 21.2 \\ 16.7 \\ 19.8 \\ 23.9 \\ 22.3 \\ 15.9 \\ 19.2 \\ 21.8 \end{bmatrix}, X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \xi_{11} \\ \xi_{12} \\ \xi_{13} \\ \xi_{21} \\ \xi_{22} \\ \xi_{23} \end{bmatrix}$$

and $\boldsymbol{\varepsilon}$ is as expected.

(c) Express the hypothesis that there is no interaction in terms of your parameters. Eliminate any redundancies.

Solution: We know that we require $(a-1)(b-1) = 2$ hypotheses, so we take the obviously non-redundant hypotheses

$$\begin{aligned} (\xi_{11} - \xi_{12}) - (\xi_{21} - \xi_{22}) &= 0 \\ (\xi_{11} - \xi_{13}) - (\xi_{31} - \xi_{33}) &= 0. \end{aligned}$$