STAT 269 – Winter 2009

Homework Assignment 8

Assignment 8 — due Friday, March 20 (in the class or in the mail-box for office 511 during the day)

1. Show that the *residual sum of squares* in the simple linear regression model can be found by any of the following formulas:

$$S_{res} = S_{yy} - \hat{\beta} S_{xy},$$

and

$$S_{res} = S_{yy} - \hat{\beta}^2 S_{xx}.$$

2. Consider the multiple linear regression model in the matrix form

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e},$$

with unknown parameters $\beta_0, \beta_1, \beta_2$ and n responses, where the components of $\mathbf{e}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$. Assume that the column vectors $\mathbf{X}_i$ of the matrix $\mathbf{X}$ are *orthogonal*, that is $\mathbf{X}_i'\mathbf{X}_j = 0$ and $\mathbf{X}_i'\mathbf{X}_i = n$, for any $i \neq j$.

a) Find the *least squares estimates* $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$.

b) Determine the *variance-covariance matrix* of the least squares estimates $\hat{\beta}$.

c) Suppose the goal is to predict the expected future response μ to the stimuli variables x_1 and x_2. Give an unbiased estimator of μ and find its variance.

d) Give an example of a matrix $\mathbf{X}$ satisfying the above condition with $n = 4$.

3. Three objects are located on a line at points $p_1 < p_2 < p_3$. A surveyor makes the following measurements:

1) First he measures the three distances from the origin to p_1, p_2, p_3. Let these measurements be denoted by Y_1, Y_2, Y_3.

2) He measures the distance from p_1 to p_2 and p_3. Let these measurements be denoted by Y_4, Y_5.

3) He measures the distances from p_2 to p_3. Denote this measurement by Y_6.

All measurements are subject to independent errors having normal distribution $\mathcal{N}(0, \sigma^2)$.

a) Set up a linear model and describe explicitly the matrix $\mathbf{X}$.

b) Determine the *variance-covariance matrix* of the least squares estimates.

4. Two objects of unknown weights w_1 and w_2 are weighed on an error-prone pan balance in the following way: (1) object 1 is weighed by itself, and the measurement is 3g; (2) object 2 is weighed by itself, and the result is 3g; (3) the difference of the weights $w_1 - w_2$ is measured by placing the objects in different pans, and the result is 1g; (4) the sum of the weights is measured as 7g. The problem is to estimate the true weights of the objects from these measurements.

   a) Set up a linear model, $\mathbf{Y} = \mathbf{X}\mathbf{w} + \mathbf{e}$.

   b) Find the least squares estimates of w_1 and w_2.

   c) Find the estimate of σ^2.

   d) Find the estimated standard errors of the least squares estimates of part b).

   e) Estimate $w_1 - w_2$ and its standard error.

   f) Test the null hypothesis $H_0 : w_1 = w_2$, at the significance level $\alpha = 0.05$.

5. Consider fitting the curve $y = \beta_1 x + \beta_2 x^2$ to the points (x_i, y_i), where $i = 1, ..., n$.

   a) Use the matrix approach to find expressions for the least squares estimates of β_1 and β_2.

   b) Find the variance-covariance matrix of these estimates.

6. Let X and Y be two normal random variables, with given means μ_x and μ_y, variances $\sigma_x^2 > 0$ and $\sigma_y^2 > 0$, and covariance σ_{xy}.

   a) Find the best linear predictor $\alpha + \beta X$ of Y, with the smallest expected error

   $$E(Y - \alpha - \beta X)^2.$$

   b) Let $\hat{\alpha} + \hat{\beta} X$ be the best such predictor. Find its expected error.

   c) Show that $Y - \hat{\alpha} - \hat{\beta} X$ and X are uncorrelated.

   d) **(optional)** Conclude that $\hat{\alpha} + \hat{\beta} X = E(Y|X)$.