STAT 408/608 Homework 2 Solutions: Written Section

February 6, 2015

1. Show that $Var(Y_i) = Var(e_i)$ in the simple linear regression model. (Yes, this should be that simple.) What did you assume?

   For simple linear regression model, $Var(Y_i) = Var(\beta_0 + \beta_1 x_i + e_i) = Var(e_i)$, where $e_i \sim N(0, \sigma^2)$, and $\beta_0 + \beta_1 x_i$ is a constant.

2. Define in words only the least squares criterion.

   Least squares criteria is a statistical approach used to provide the most accurate estimate of relationships between sets of variables in sample data. It is a measure of the differences between the observed data and the calculated data point by summing the squares of the differences between them.

3. Show that the least squares criterion applied to the "intercept-only" model, i.e.

   $$y_i = \beta_0 + e_i, i = 1, 2, \ldots, n$$

   results in the least squares estimator of $\beta_0 : \hat{\beta}_0 = \bar{y}$.

   Apply least squares criterion to the "intercept-only" model: $RSS = \sum(y_i - \hat{\beta}_0)^2$.

   (a) Yes. It is the special case of general form when the design matrix:
   $$X = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

   (b) $\hat{\beta}_0 = (X'X)^{-1}X'y = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$

4. Page 40 in our textbook.

   (a) The design matrix:
   $$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

   (b) $\hat{\beta} = (X'X)^{-1}X'y = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$

5. Using $\hat{\beta} = (X'X)^{-1}X'y$, finish our algebra from class and show that $\hat{\beta}_0 = \bar{y} - \dfrac{SXY}{SXX}\bar{x}$ for the simple linear regression case. Give a few more algebraic details than are on page 133.

$$\begin{aligned} SXX &= \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = n \cdot (\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2) \end{aligned}$$

$$\begin{aligned} SXY &= \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n}(x_i y_i - \bar{x}y_i - x_i\bar{y} + \bar{x}\bar{y}) \\ &= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y \end{aligned}$$

See textbook page 133, and we have $\hat{\beta}_0 = \bar{y} - \dfrac{SXY}{SXX}\bar{x}$, $\hat{\beta}_1 = \dfrac{SXY}{SXX}$.

6. Show that for the usual regression model $y = X\beta + e$, where the usual regression assumptions from question 4 apply, $\mathrm{Var}(a'\hat{\beta}|X) = \sigma^2 a'(X'X)^{-1}a$, where a is a constant vector. (We'll use this fact later.)

$$\begin{aligned} Var(a'\hat{\beta} \mid X) &= Var(a'(X'X)^{-1}X'Y \mid X) \\ &= a'(X'X)^{-1}X'Var(Y \mid X)X(X'X)^{-1}a \\ &= a'(X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1}a \\ &= \sigma^2 a'(X'X)^{-1}a \end{aligned}$$

7. Page 42 in our textbook.

Prediction intervals includes extra variations, and are always wider than the confidence intervals. See the formulas in the textbook. Hence, it is completely possible that the observed Y values fall outside of 95% confidence interval.

8. The figure below shows a scatterplot of some data together with a line that purports to have been fitted by least squares. The averages of the x and y values are 4.4 and 9.9 respectively. The line in the figure cannot be the least squares line. Say why not AND provide a justification for your answer.

The least square line for a given data set must pass through the point $(\bar{x}, \bar{y})$. In this plot, this point is off the line.