Stat 354 Fall 2018
Assignment #1

This assignment is due at the beginning of class on Friday, September 28, 2018. Your solutions will be graded based on both correctness *and* exposition. In particular, neatness and grammar count. You must write out solutions using full sentences (including capital letters to start sentences and periods to end them) and no abbreviations. That is, symbols such as $\therefore$ and $\Rightarrow$ are forbidden; write out the full words *therefore* and *implies* in their place.

1. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \ldots, n$, where $\epsilon_1, \ldots, \epsilon_n$ are independent with common mean $\mathbb{E}(\epsilon_i) = 0$ and common variance $\mathrm{Var}(\epsilon_i) = \sigma^2$ for all i. Define

$$\overline{x} = \frac{1}{n}\sum x_i, \ \ \overline{y} = \frac{1}{n}\sum y_i, \ \ s_{yy} = \sum (y_i - \overline{y})^2, \ \ s_{xx} = \sum (x_i - \overline{x})^2, \ \ s_{xy} = \sum (x_i - \overline{x})(y_i - \overline{y})$$

so that the least squares estimators of β_1, β_0 as derived in class are

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}} \quad \text{and} \quad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x},$$

respectively. Finally, set $\mu_0 = \beta_0 + \beta_1 x_0$ and $\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.

(a) Verify that $\mathbb{E}(\hat{\mu}_0) = \mu_0$.

(b) Verify that

$$\mathrm{Var}(\hat{\mu}_0) = \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_{xx}} \right) \sigma^2.$$

(c) Verify that

$$\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = s_{yy} - \hat{\beta}_1^2 s_{xx} = s_{yy} - \frac{s_{xy}^2}{s_{xx}}.$$

2. Consider the simple linear regression model defined in Problem 1. The purpose of this problem is to prove that if

$$\hat{\sigma}^2 = \frac{1}{n}\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2,$$

then

$$\mathbb{E}(\hat{\sigma}^2) = \left(\frac{n-2}{n} \right) \sigma^2.$$

(a) Recall that if z is a random variable, then $\mathbb{E}(z^2) = \mathrm{Var}(z) + [\mathbb{E}(z)]^2$. Use this fact to prove the following identities:

(i) $\mathbb{E}(y_i^2) = \sigma^2 + (\beta_0 + \beta_1 x_i)^2$,

(ii) $\mathbb{E}(\overline{y}^2) = \dfrac{\sigma^2}{n} + (\beta_0 + \beta_1 \overline{x})^2$,

(iiii) $\mathbb{E}(\hat{\beta}_1^2) = \dfrac{\sigma^2}{s_{xx}} + \beta_1^2$.

(b) Verify that $\mathbb{E}(s_{yy}) = (n-1)\sigma^2 + \beta_1^2 s_{xx}$.

(continued)

(c) Use part (c) of the previous problem to verify

$$\mathbb{E}\left[\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\right] = (n-2)\sigma^2.$$

(d) Deduce

$$\mathbb{E}(\hat{\sigma}^2) = \left(\frac{n-2}{n}\right)\sigma^2.$$

3. (Scale Invariance of the SLR model.) Consider the simple linear regression model defined in Problem 1. Suppose that we replace x_i by kx_i where $k \neq 0$ is a constant so that we have the new model

$$y_i = \beta_0^{\text{new}} + \beta_1^{\text{new}}(kx_i) + \epsilon_i, \quad i = 1, \ldots, n.$$

(a) Carefully verify that $\hat{\beta}_1^{\text{new}} = k^{-1}\hat{\beta}_1$.

(b) Carefully verify that $\hat{\beta}_0^{\text{new}} = \hat{\beta}_0$.

4. The purpose of this problem is to derive a model of regression through the origin; i.e., where it is known a priori that the intercept is zero.

Suppose that we observe data (x_i, y_i), $i = 1, \ldots, n$, and postulate that it is appropriate to describe the relationship between x and y by the regression model $y = \beta x + \epsilon$. In particular, this assumes that

$$y_i = \beta x_i + \epsilon_i, \ \ i = 1, \ldots, n,$$

where β is a parameter and $\epsilon_1, \ldots, \epsilon_n$ are independent with common mean $\mathbb{E}(\epsilon_i) = 0$ and common variance $\text{Var}(\epsilon_i) = \sigma^2$.

(a) Let $S(\beta) = \sum \epsilon_i^2 = \sum(y_i - \beta x_i)^2$. Prove that the minimum of $S(\beta)$ occurs at

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}.$$

(b) Verify that $\hat{\beta}$ is an unbiased estimator of β.

(c) Verify that

$$\text{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2}.$$