Linear Statistical Models — Quiz 01 — Full Name ______________________

1. Suppose we have covariate data $x_1, \ldots, x_n$ and response data $y_1, \ldots, y_n$.

(i) (4) Write down the simple linear statistical model for y_i, making sure to define each and every term you specify, including all assumptions. Use b_0 and b_1 as notation for the regression coefficients.

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

$$\varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

b_0, b_1, σ^2 unknown parameters

(ii) (3) Write down the formula for the sum of squares criterion.

$$rss(b_0^*, b_1^*) = \sum_{i=1}^{n} (y_i - b_0^* - b_1^* x_i)^2$$

(iii) (3) Write down the formulas for the estimates of the two regression coefficients. Define all terms

$$\hat{b}_1 = Sxy/Sxx$$

$$Sxx = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$Sxy = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$$

(iv) (3) In terms of the estimates $\widehat{b}_0$ and $\widehat{b}_1$, write down the formulas for the fitted values, the residuals, and $\widehat{\sigma}^2$ (the variance estimate).

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_i$$

$$\hat{e}_i = y_i - \hat{y}_i$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{e}_i^2$$

2. (3) Explain the difference between an estimate and an estimator.

estimate is a function of the data.

estimator is a R.V., the same function of data but with Y_i replacing y_i

3. (3) Name two desirable properties of estimators, and explain why they are desirable.

Unbiased: sampling distribution is centered on the estimand

small variance: sampling distribution tightly centered around estimand

4. (3) In the notes, we derived the following:

$$\text{Var}(\widehat{B}_0) = \sigma^2 \left(\frac{1}{n} + \frac{(\overline{x})^2}{sxx} \right), \quad \text{Var}(\widehat{B}_1) = \frac{\sigma^2}{sxx}, \quad \text{Cov}(\widehat{B}_0, \widehat{B}_1) = -\frac{\sigma^2 \overline{x}}{sxx}$$

Derive a simplified expression for the prediction variance $\text{Var}(\widehat{B}_0 + \widehat{B}_1 x_i)$.

$$\text{Var}(\hat{B}_0 + \hat{B}_1 x_i) = \text{Var}(\hat{B}_0) + x_i^2 \text{Var}(\hat{B}_1) + 2x_i \text{Cov}(\hat{B}_0, \hat{B}_1)$$

$$\sigma^2 \left(\frac{(\bar{x})^2}{sxx} + \frac{1}{n} + \frac{x_i^2}{sxx} - \frac{2x_i\bar{x}}{sxx} \right)$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{sxx} \right)$$

5. (5) The "normal equations" result from taking partial derivatives of the sum of squares criterion and setting them to zero. Here are the equations:

$$0 = \sum_{i=1}^{n}(y_i - \widehat{b}_0 - \widehat{b}_1 x_i) \qquad \text{and} \qquad 0 = \sum_{i=1}^{n}(y_i - \widehat{b}_0 - \widehat{b}_1 x_i)x_i$$

Solve these equations to derive expressions for $\widehat{b}_0$ and $\widehat{b}_1$. You must show your work for full credit.

$$0 = \sum_{i=1}^{n}(y_i - \hat{b}_0 - \hat{b}_1 x_i)$$

$$n\hat{b}_0 = n\bar{y} - \hat{b}_1 n\bar{x} \Rightarrow \hat{b}_0 = \bar{y} - \hat{b}_1\bar{x}$$

$$0 = \sum_{i=1}^{n}(y_i - (\bar{y} - \hat{b}_1\bar{x}) - \hat{b}_1 x_i)x_i$$

$$0 = \sum(y_i - \bar{y})x_i - \hat{b}_1\sum(x_i - \bar{x})x_i$$

$$\sum(y_i - \bar{y})x_i = \hat{b}_1\sum(x_i - \bar{x})x_i$$

$$\sum(y_i - \bar{y})x_i - \sum(y_i - \bar{y})\bar{x} = \hat{b}_1\left[\sum(x_i - \bar{x})x_i - \sum(x_i - \bar{x}\right.$$

$$\sum(y_i - \bar{y})(x_i - \bar{x}) = \hat{b}_1\sum(x_i - \bar{x})^2$$

$$\hat{b}_1 = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$