Problem Set 9

You are strongly encouraged to solve the following exercises before next week's tutorial:

- Starting on page 230 (end of Chapter 13): 3, 6, 7
- Starting on page 591 (end of Chapter 14): 2, 3, 7, 11, 21.

Additional Exercise:

In class we proved that the variance of the least squares estimator of the linear regression slope was

$$\mathrm{Var}[\widehat{\beta}_1] = \frac{\sigma^2}{\sum_j (x_j - \overline{x})^2}.$$

We also managed to express

$$\widehat{\beta}_1 = \frac{\sum_i (x_i - \overline{x}) y_i}{\sum_j (x_j - \overline{x})^2} \qquad \text{and} \qquad \widehat{\beta}_0 = \sum_i \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_j (x_j - \overline{x})^2} \right] y_i.$$

(a) Show that $\widehat{\beta}_0$ is an unbiased estimator of β_0.

(b) Calculate the variance of $\widehat{\beta}_0$.

(c) Calculate $\mathrm{Cov}\left(\widehat{\beta}_0, \widehat{\beta}_1\right)$.

(d) Show that $\widehat{\beta}_1 = r_{XY} \dfrac{S_Y}{S_X}$.

Solution:

(a) First note that since $\mathbb{E}[\varepsilon_i] = 0 \ \forall i$,

$$\mathbb{E}[\overline{y}] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[y_i] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[\beta_0 + \beta_1 x_i + \varepsilon_i] = \beta_0 + \beta_1 \overline{x}.$$

Now,

$$\mathbb{E}[\widehat{\beta}_0] = \mathbb{E}[\overline{y} - \widehat{\beta}_1 \overline{x}] = \mathbb{E}[\overline{y}] - \overline{x}\mathbb{E}[\widehat{\beta}_1] = \beta_0 + \beta_1 \overline{x} - \beta_1 \overline{x} = \beta_0.$$

(b)
$$\mathrm{Var}[\widehat{\beta}_0] = \sum_i \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_j (x_j - \overline{x})^2}\right]^2 \mathrm{Var}[y_i] = \sigma^2 \sum_i \left[\frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{\sum_j (x_j - \overline{x})^2}\right]^2$$

$$= \sigma^2 \sum_i \left[\frac{1}{n^2} - \frac{2\overline{x}(x_i - \overline{x})}{n\sum_j (x_j - \overline{x})^2} + \frac{\overline{x}^2 (x_i - \overline{x})^2}{\left\{\sum_j (x_j - \overline{x})^2\right\}^2}\right]$$

$$= \sigma^2 \left[\frac{1}{n} - \underbrace{\frac{2\overline{x}\sum_i (x_i - \overline{x})}{n\sum_j (x_j - \overline{x})^2}}_{0} + \frac{\overline{x}^2 \sum_i (x_i - \overline{x})^2}{\left\{\sum_j (x_j - \overline{x})^2\right\}^2}\right] = \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_j (x_j - \overline{x})^2}\right].$$

(c)
$$\mathrm{Cov}\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \mathrm{Cov}\left(\overline{y} - \overline{x}\widehat{\beta}_1, \widehat{\beta}_1\right) = \mathrm{Cov}\left(\overline{y}, \widehat{\beta}_1\right) - \overline{x}\mathrm{Var}[\widehat{\beta}_1]$$

$$= \mathrm{Cov}\left(\frac{1}{n}\sum_k y_k, \frac{\sum_i (x_i - \overline{x})}{\sum_j (x_j - \overline{x})^2} y_i\right) - \frac{\sigma^2 \overline{x}}{\sum_j (x_j - \overline{x})^2}$$

$$= \frac{1}{n\sum_j (x_j - \overline{x})^2} \sum_k \sum_i (x_i - \overline{x}) \mathrm{Cov}\left(y_k, y_i\right) - \frac{\sigma^2 \overline{x}}{\sum_j (x_j - \overline{x})^2}$$

$$= \frac{1}{n\sum_j (x_j - \overline{x})^2} \sum_i (x_i - \overline{x}) \mathrm{Cov}\left(y_i, y_i\right) - \frac{\sigma^2 \overline{x}}{\sum_j (x_j - \overline{x})^2}$$

$$= \frac{\sigma^2}{n\sum_j (x_j - \overline{x})^2} \underbrace{\sum_i (x_i - \overline{x})}_{0} - \frac{\sigma^2 \overline{x}}{\sum_j (x_j - \overline{x})^2}$$

$$= -\frac{\sigma^2 \overline{x}}{\sum_j (x_j - \overline{x})^2}.$$

(d)
$$\widehat{\beta}_1 = \frac{S_{XY}}{S_X^2} = \frac{S_{XY}}{S_X S_X} \cdot \frac{S_Y}{S_Y} = \frac{S_{XY}}{S_X S_Y} \cdot \frac{S_Y}{S_X} = r_{XY} \frac{S_Y}{S_X}.$$