EEB 581, Problem Set One: Solutions

1 : Data was measured on 50 individuals for arm size (x) and brain size (y), with the following results:

$$\overline{x} = 10, \quad \overline{y} = 50, \quad \sum_{i=1}^{50} (x_i - \overline{x})^2 = 100, \quad \sum_{i=1}^{50} (y_i - \overline{y})^2 = 400, \quad \sum_{i=1}^{50} (x_i - \overline{x})(y_i - \overline{y}) = 175$$

(a) Compute the variances of x and y, their covariance, and their correlation.

$$\textbf{Var}(x) = \frac{100}{49} = 2.04, \quad \textbf{Var}(y) = \frac{400}{49} = 8.16, \quad \textbf{Cov}(x,y) = \frac{175}{49} = 3.57$$

$$\textbf{Corr}(x,y) = \frac{3.57}{\sqrt{2.04} \cdot \sqrt{8.16}} = 0.88$$

(b) What the best linear regression of arm size (x) on brain size (y)?

$$b_{x\,|\,y} = \frac{3.57}{8.16} = 0.44, \qquad a = \overline{x} - b_{x\,|\,y}\,\overline{y} = 10 - 0.44 \cdot 50 = -11.88$$

Hence, the regression is (Arm size) = -11.88+ 0.44(Brain size)

(c) What the best linear regression of brain size (y) on arm size (x)?

$$b_{y\,|\,x} = \frac{3.57}{2.04} = 1.75, \qquad a = \overline{y} - b_{y\,|\,x}\,\overline{x} = 50 - 1.75 \cdot 10 = 32.58$$

Hence, the regression is (Brain size) = 32.50 + 1.75(Arm size)

(d) What fraction of the total variance in brain size does the regression account for?

Fraction of the total variance explained by the regression is the squared correlation, or $0.88^2 = 0.766$

2 : What is the covariance between a particular data point (x_i) and the sample mean $\overline{x}$?

$$\sigma\left(x_i, \frac{1}{n}\sum_{j=1}^{n} x_j\right) = \frac{1}{n}\sum_{j=1}^{n} \sigma(x_i, x_j) = \frac{\sigma(x_i, x_i)}{n} + \frac{1}{n}\sum_{j \neq i}^{n} \sigma(x_i, x_j) = \frac{\sigma^2(x_i)}{n}$$