# PSTAT 126 - Assignment 1

Fall 2022

Due: Tuesday, October 4 at 11:59 pm on Canvas

*Note: If you use `Rmd` format to work on your assignment, use the same indentation level as **Solution** markers to write your solutions. Improper indentation will break your document.*

1. Let $X$ and $Y$ be random variables, with $\mu_X = E(X), \mu_Y = E(Y), \sigma_X^2 = \mathrm{Var}(X)$ and $\sigma_Y^2 = \mathrm{Var}(Y)$

   a. Prove the following property regarding $\mathrm{Cov}(X, Y)$. For a fixed real number $b$, show that

   $$\mathrm{Cov}(b + X, Y) = \mathrm{Cov}(X, Y) \quad \text{and} \quad \mathrm{Cov}(bX, Y) = b\,\mathrm{Cov}(X, Y)$$

   **Solution**:

   b. Prove the following property regarding $\mathrm{Corr}(X, Y)$. For any fixed real numbers $a$ and $b$, and $c, d > 0$, show that

   $$\mathrm{Corr}(a + cX, b + dY) = \mathrm{Corr}(X, Y).$$

   **Solution**:

   c. For this part only, assume $\mathrm{Cov}(X, Y) > 0$. According to lecture, this implies that there exists a real number $b_1$ such that $\mathrm{Var}\,(Y - b_1 X) < \sigma_Y^2$. Show that this is true whenever $0 < b_1 < 2\beta_1$, where $\beta_1 = \mathrm{Cov}(X, Y)/\sigma_X^2$.
   *Hint:* $\mathrm{Var}\,(Y - b_1 X)$ *is a quadratic function of* $b_1$. *Use what you know about parabolas to answer this question.*

   **Solution**:

   d. Let $\beta_1$ be as in part c) and set $\beta_0 = \mu_Y - \beta_1 \mu_X$. Show that the minimal MSE (mean squared error) is

   $$\mathrm{MSE}\,(\beta_0, \beta_1) = E\left[(Y - \beta_0 - \beta_1 X)^2\right] = \sigma_Y^2 - \frac{\mathrm{Cov}^2(X, Y)}{\sigma_X^2}.$$

   To do this, complete the follwing steps.

   1) First, show that $E\,(Y - \beta_0 - \beta_1 X) = 0$ so that

   $$\mathrm{MSE}\,(\beta_0, \beta_1) = \mathrm{Var}\,(Y - \beta_0 - \beta_1 X)\,.$$

   **Solution**:

   2) Use property 6) from lecture about the variance of the difference between random variables to show that

   $$\mathrm{Var}\,(Y - \beta_0 - \beta_1 X) = \sigma_Y^2 + \beta_1^2 \sigma_X^2 - 2\beta_1\,\mathrm{Cov}(X, Y).$$

   **Solution**:

   3) Use that fact that $\beta_1 = \mathrm{Cov}(X, Y)/\sigma_X^2$ to simplify this last expression.

   **Solution**:

2. We are now given data on $n$ observations $(x_i, Y_i)$, $i = 1, \ldots, n$. Assume we have a linear model, so that $E(Y_i) = \beta_0 + \beta_1 x_i$, and let $\hat{\beta}_1 = S_{XY}/S_{XX}$ and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$ be the least squares estimates given in lecture.

   a. Show that $E(S_{XY}) = \beta_1 S_{XX}$ and $E(\bar{Y}) = \beta_0 + \beta_1 \bar{x}$, and use this to conclude that $E\left(\hat{\beta}_1\right) = \beta_1$ and $E\left(\hat{\beta}_0\right) = \beta_0$. In other words, these are unbiased estimators.

      **Solution**:

   b. The fitted values $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ are used as estimates of $E(Y_i)$, and the residuals $e_i = Y_i - \hat{Y}_i$ are used as surrogates for the unobservable errors $\varepsilon_i = Y_i - E(Y_i)$. By assumption, $E(\varepsilon_i) = 0$. Show that the residuals satisfy a similar property, namely

$$\sum_{i=1}^{n} e_i = 0$$

      **Solution**: