

## HW 23

### Problem 1

For  $n = 1040$  male college soccer players, the correlation between height and weight is about  $r = 0.75$ . The sample means for heights and weights are about  $\bar{x} = 71$  in and  $\bar{y} = 166$  lbs, and the sample standard deviations are about  $s_x = 2.5$  in and  $s_y = 16$  lbs.

- (a) Find the least squares regression line for predicting weight from height. What proportion of the variability in weights is explained by a linear fit on height?
- (b) Find the fitted weight for a 66 inch player and for a 76 inch player. Explain how these fitted values illustrate the regression towards the mean effect in an answer that involves standard deviations relative to the respective means.
- (c) Use the sample correlation and standard deviation of the weights to find the root mean squared error for the simple regression model. Explain what this number represents in this context.

## Problem 2

Consider the no-intercept linear regression model

$$Y_i \mid X_i = x_i \sim N(\beta x_i, \sigma^2), \quad i = 1, \dots, n.$$

We should include an intercept in the model even if we believe the mean response when  $x = 0$  should be 0, however working with the no-intercept model can help understand the more complicated model since here  $\beta$  is a scalar rather than a vector.

- (a) Show that the least squares estimate for  $\beta$  is  $\hat{\beta} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ , where  $\mathbf{X}$  is the  $n \times 1$  matrix (vector) of  $x_i$  values and  $\mathbf{Y}$  is the  $n \times 1$  vector of  $Y_i$  values.
- (b) Write the joint log-likelihood of  $(\beta, \sigma^2)$  and explain why the MLE for  $\beta$  is the same as the least squares estimate for  $\beta$ .
- (c) Find the mean and variance of  $\hat{\beta}$ .

### Problem 3

A simple exponential decay model says that the concentration,  $C_{(t)}$  of a pesticide remaining after time  $t$  is  $C_{(t)} = C_0 e^{-\gamma t}$  for  $t > 0$  where  $C_0$  is the initial concentration and  $\gamma$  is a constant that determines the rate of decay.

- (a) Show how taking the natural log of both sides of the equation above results in a linear model for  $Y = \log(C_{(t)})$  on  $t$ . What are the slope and intercept?
- (b) If you have data on concentrations at  $n$  different times,  $t_i$ , you could estimate  $\gamma$  by fitting a SLR of  $Y_i$  on  $t_i$ . This implicitly assumes an additive error term  $\epsilon_i$  that is approximately normally distributed. Write out the implied model for  $C_{(t)}$  and describe how error enters this model.