- Due: October 11, Wednesday, 11:59PM
- How to submit: via Blackboard. If you have multiple files, upload a zipped file
- Submission link will disappear after 48 hours
- Homework solution is not required to be typed, but must be legible.

Problem 1 A student fitted a linear regression function for a class assignment. The student plotted the residuals e_i against y_i, and found a positive relation. When the residuals were plotted against the fitted values $\hat{y}_i$, the student found no relation. How could this difference arise? Which is the more meaningful plot? Explain why.

Problem 2 $\hat{y}_0$ is the fitted value or predicted value of the response when the predictor is x_0.

1) Derive the variance of $\hat{y}_0$, $\sigma^2\{\hat{y}_0\}$, in a Simple Linear Regression setting when there is only one predictor.

2) Derive the variance of $\hat{y}_0$, $\sigma^2\{\hat{y}_0\}$, in a Multiple Linear Regression setting when there are p predictors.

3) Derive the variance of the prediction error, $\sigma^2\{y_0 - \hat{y}_0\}$, in a Multiple Linear Regression setting when there are p predictors.

Problem 3 For a multiple linear regression model with 4 variables,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$, what is the test statistic for each of the following tests? and what is the probability distribution for each test statistic? Assume there are n data points.

1) testing whether or not $\beta_4 = 0$
2) testing whether or not $\beta_2 = \beta_3 = 0$
3) testing if all slopes are zero

Problem 4 For multiple linear regression with n data points and p predictors, prove that $\sum\limits_{i=1}^{n} Var(\hat{y}_i) = (p+1)\sigma^2$, where σ^2 is the common variance of the error term in the MLR model.

Problem 5 Let H be the hat matrix for multiple linear regression with p predictors and n data points. The notation h_{ij} represents $H[i][j]$, the element of H at the i-th row and the j-th column.

1) Show that $h_{ii} \in [0, 1]$.
2) Show that $\sum_{i=1}^{n} h_{ii} = p + 1$.
3) Plot h_{ii} vs. $|x_i - \bar{x}|$ (absolute value) for $p = 1$, and plot h_{ii} vs $\|x_i - \bar{x}\|_2$ (2-norm) for $p > 1$. Use the data generated by your own code, and let $n = 100$, $p = 4$.