## Regression Analysis - STAT 482 - Exam 1: Due December 2, 2021

## Alex Towell (atowell@siue.edu)

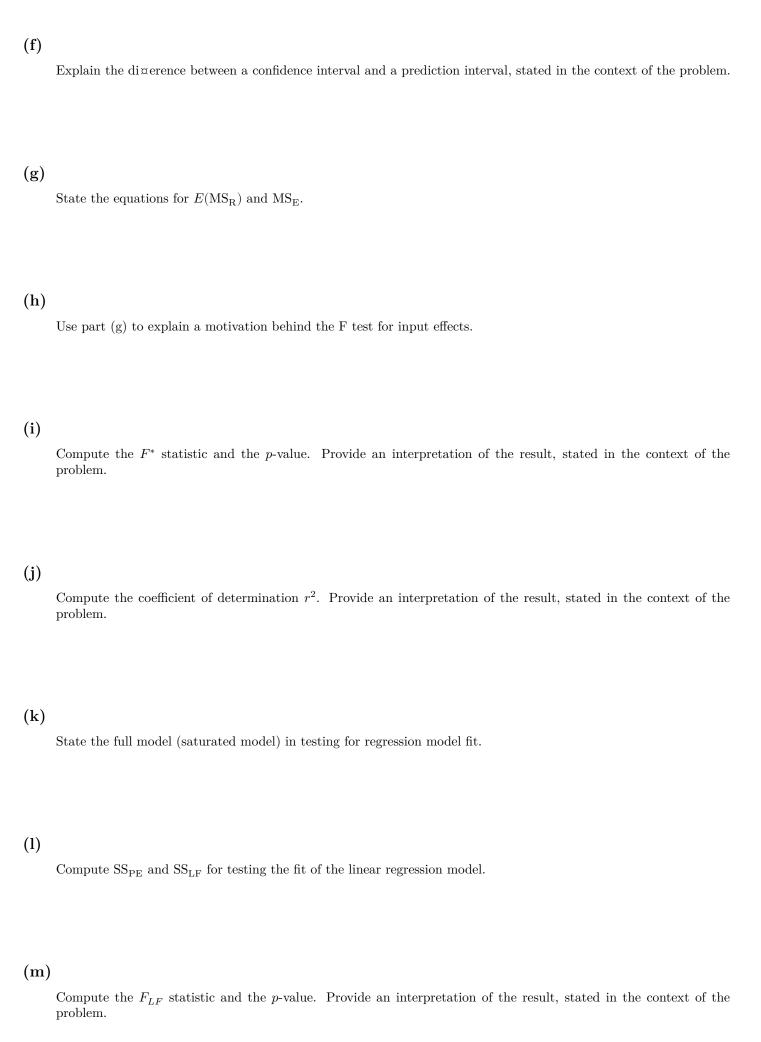
## Problem 1

Experience with a certain type of plastic indicates that a relationship exists between the hardness (measured in Brinell units) of items molded from the plastic (response variable y) and the elapsed time (measured in hours) since the termination of the molding process (input variable x). Sixteen batches of the plastic were made, and from each batch one test item was molded. Each test item was randomly assigned to one of the four predetermined time levels. The data is available on Blackboard as a csv file.

(a)				
	State the s	imple line	ar regression	model

- (b)

  Provide an interpretation for the regression parameter
- (c) Compute an interval estimate for  $\beta_1$ , stated in the context of the problem.
- (d) Compute a confidence interval for h at  $x_h = 40$  hours.



(n)	Create a plot comparing the fitted values from the regression model with the fitted values from the saturated model.
(o)	Explain an advantage of using the regression estimate $\hat{\mu}_{40}$ instead of the sample mean $\bar{y}_{40}$ and explain when this advantage will lead to a more accurate estimator.
Pro	oblem 2
	Measurements were made on men involved in a physical fitness course. The input variables under consideration are age (in years), weight (in kgs), time to run 1.5 miles (in minutes), heart rate while resting (in bpm), heart rate while running, and maximum heart rate. The goal is to determine which input variables are needed to model the oxygen uptake rate (ml/kg body weight per minute). The data is available on Blackboard as a csv file.
(a)	Describe the goal of the discrepancy function approach to model selection. How is the best model defined? What are the two sources of model error?
(b)	Plot the $C_p$ statistic against the model dimension. Plot the $C_p$ statistic against the candidate models. Which variables are included in the selected model?
(c)	Test the selected model against its best competitor having fewer parameters. (Compute the test statistic and the $p$ -value.) How does discrepancy based model selection compare to $p$ -value based selection?