Directions: Any problem marked with asterisk (*) should be completed using IPython Notebook (Jupyter) and can be uploaded at https://dropitto.me/m50f15 .

1*. **Grade point average.** The director of admissions of a small college selected 120 students at random from the new freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the ACT test score (X). The results of the study follow (see data link below). Assume that first-order regression model is appropriate.

Data source: https://netfiles.umn.edu/users/nacht001/www/nachtsheim/Kutner/Chapter%20
%201%20Data%20Sets/CH01PR19.txt

This data can also be accessed using the function read_tb_data(ch,pr) *with ch=1 and pr=19.*

(a) Write a python function to set up ANOVA table for the above data. The function should take parameters X and Y, and print complete ANOVA table for the given X and Y with five columns:

$$| \ Source \ | \ SS \ | \ df \ | \ MS \ | \ E\{MS\} \ |$$

(b) What is estimated by MSR in your ANOVA table? by MSE? Under what condition do MSR and MSE estimate the same quantity?

(c) What is the absolute magnitude of the reduction in the variation of Y when X is introduced into the regression model? What is the relative reduction? What is the name of the latter measure?

(d) Plot the data, with the least squares regression line for ACT scores between 20 and 30 superimposed.

Reference: *Kutner et. al. "Applied Linear Regression Models", Ed. 5 (problem no 2.23 on pages 93).*

2. The normal error regression model is assumed to be applicable.

(a) When testing $H_0 : \beta_1 = 5$ versus $H_a : \beta_1 \neq 5$ by means of a general linear test, what is the reduced model? What are the degrees of freedom df_R?

1

(b) When testing $H_0 : \beta_0 = 2, \beta_1 = 5$ versus H_a : not both $\beta_0 = 2$ and $\beta_1 = 5$ by means of a general linear test, what is the reduced model? What are the degrees of freedom df_R?

3. In developing empirically a cost function from observed data on a complex chemical experiment, an analyst employed normal error regression model. β_0 was interpreted here as the cost of setting up the experiment. The analyst hypothesized that this cost should be $7.5 thousand and wished to test the hypothesis by means of a general linear test.

(a). Indicate the alternative conclusions for the test.

(b) Specify the full and reduced models.

(c) Without additional information, can you tell what the quantity $df_R - df_F$ will equal in the analyst's test? Explain.