**Applied Statistics and Design of Experiments**

**Exercise 6**

**February 20, 2019**

*For these exercises you may work individually or in groups. There are no uploads required for this exercise.*

**Question 1.**

**Data set: ‘fev.csv’**

1.1. Use two separate linear regression models to assess (i) the association between FEV and age, and (ii) the association between FEV and height (ht).

1.2. Then fit a linear regression model with age and height included in the same model. Interpret the results of this model and compare with the results of the two models in the previous question.

1.3. Check the X’X-inverse matrix. Is the off-diagonal element corresponding to age and height equal to 0 or not. Explain this using a scatterplot, correlation and a regression analysis to show the association between age and height.

**Question 2. Data set “pilot.csv”**

Your job is to design experiments to test the effect of temperature on the output of a manufacturing process in two different scenarios. The goal of the experiments is to estimate the effect of temperature on output using a linear regression model with the result expressed using the regression coefficient for temperature, which has interpretation as the difference in mean output per degree of temperature. You may assume that all of the assumptions of the linear regression model will hold for this experiment. The data from a pilot study is available (“pilot.csv”).

Scenario 1. The client would like you to use the same values of temperature (deg C) as in the pilot study: 10, 20, 30, and 40. Your job is to determine the sample size needed to estimate the regression coefficient with a standard error of at most 0.01.

Scenario 2. In this case you can use any values of temperature within the range from 10 to 40 deg C. However, this experiment will be limited to at most 50 observations. Which values of temperature should you use to minimize the standard error of the estimated regression coefficient?

**Question 3.**

Fit a regression model to the fev data with age, male gender, height, and the smoking indicator. Assess the assumptions of this model.