**Applied Statistics and Design of Experiments**

**Exercise 5**

**February 13, 2019**

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

**Question 1.**

Data set: ‘wine.csv’

The data are from an experiment designed to assess the effects of alcohol content and pH on the quality of red wines. Alcohol content was set at 10%, 12%, or 14% and pH was set at 2.8, 3.2, or 3.6. The experiment was run as a full factorial design with each of the 9 combinations of alcohol content and pH being used. A batch of harvested grapes was divided into 54 homogeneous portions, and the portions were randomly assigned to the 9 conditions (combination of alcohol content and pH) with 6 portions per treatment combination. For each of the 54 portions, a batch of wine was made with the assigned alcohol and pH values. The 54 batches of wine were stored in a cellar with position in the cellar randomly assigned for 18 months. Then a sample of wine from each barrel was removed and the quality of the wine was assessed on a scale of 0-100.

The variables are:

alc = alcohol content of wine (10, 12, or 14%)

ph = pH of wine (3.2, 3.4, or 3.6)

quality = quality rating of wine

1.1. Use 1-way ANOVA to test the effects of alcohol on quality (use all of the data but ignore the ph variable for this analsyis). Interpret the results.

1.2. Use 1-way ANOVA to test the effects of ph on quality (use all of the data but ignore the alcohol variable for this analysis). Interpret the results.

1.3. Now use linear regression to test the effect of alcohol on quality (again use all of the data but ignore the ph variable for this analysis). Plot the fitted regression line. Interpret the results. Compare results of 1-way ANOVA and linear regression.

1.4. Now use linear regression to test the effect of ph on quality (again use all of the data but ignore the alcohol variable for this analysis). Plot the fitted regression line. Interpret the results. Compare results of 1-way ANOVA and linear regression.

**Question 2. (continuation of analysis of wine data)**

2.1. Use 2-way ANOVA without interaction to test the effects of alcohol and ph on quality. Compare results to the two 1-way ANOVAs from Q1.1 and Q1.2.

2.2. Use a linear regression model without interaction to assess the effects of alcohol and ph on quality. Compare results to the two linear regression models in Q1.3 and Q1.4. Compare results of 2-way ANOVA and regression.

2.3. Add an interaction between alcohol and ph to the ANOVA model. Interpret the results.

2.4. Add an interaction between alcohol and ph to the linear regression model. Compare results of ANOVA and linear regression for the interaction. Make a plot of the fitted linear regression model with interaction.

**Question 3.**

**Data set: ‘fev.csv’**

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

3.2. Then fit a linear regression model with age and the smoking variable included in the same model. Interpret the results of this model and compare with the results of the two models in Q3.1.