Investigation 3: Smoking and lung capacity? (assigned on Thur Oct 8; due on Thur Oct 15)
You may work with in a group of as many as three students on this assignment, handing in one report with all names, provided that you all contribute to the work. Word-processed reports are preferred to handwritten ones. Integrate computer output into your report as appropriate.

The Minitab worksheet FEV.mtw (available from the “data files” link on our course webpage) contains data on lung capacity as measured by forced expiratory volume (FEV), the amount of air an individual can exhale in the first second of forceful breath (in liters), for a group of 654 people. The dataset also includes whether or not the person smokes and the person’s age.

a) Examine (and submit) boxplots of FEV, comparing smokers and non-smokers. Also calculate (and submit) the mean and SD, along with the five-number summary, of FEV, again comparing smokers and non-smokers. [Hints: Use Graph> Boxplot, One Y, With Groups. And use Stat> Basic Statistics> Display Descriptive Statistics.]

b) Comment on what the boxplots and statistics reveal about whether smokers and non-smokers differ with regard to FEV. Are you surprised by these findings? Explain.

c) Examine (and submit) a scatterplot of FEV vs. age [Hint: Use Graph> Scatterplot.] Comment on the direction, strength, and form of the association between FEV and age.

d) Now examine (and submit) a scatterplot of FEV vs. age, using different symbols for smokers and non-smokers. [Hint: Use Graph> Scatterplot, With Groups.] Comment on what additional information this graph reveals that the previous one did not.

e) Now add regression lines, separately for smokers and non-smokers, to the scatterplot (and submit this graph). [Hint: Use Graph> Scatterplot, With Regression and Groups.] Comment on how the two lines compare, particularly with regard to slope. Also explain what the comparison of slopes means in this context.

f) Now add different symbols and regression lines for sex as well as smoking status to the graph. [Hints: Include sex as a second variable under “categorical variables for grouping.” This should produce four separate regression lines.] Comment on what this graph reveals about comparing the relationship between FEV and age across the four sex-smoking possibilities.