Homework #5

1. (5 points) Smoke or Coffee?
Two research reports are published about a disease risk of living habits. The first report shows that the smoker proportion is higher in the diseased persons than in the health persons, with a p-value of 0.011. The second report shows that the coffee drinker proportion is higher in the diseased persons than in the health persons, with a p-value of 0.021 The first study uses a sample of 500 diseased persons and 300 healthy persons. The second study uses a sample of 1324 diseased persons and 2850 healthy persons.
Based on these two reports, which factor do you believe that caused the disease: smoking or drinking coffee?

2. (5 points) If there are only two groups, would the results of a one-way ANOVA analysis be the same as those of a two-sample t-test?

3. (15 points) Do exercise12.4.8.
Replace part (d) in textbook by:
(d) Before carrying out the study, we expected that the cholesterol levels for people with no diseases (4^{th} group) should be lower than people with disease. Therefore, to check if this is true, we conduct a hypothesis test for H_0: $\mu_1+\mu_2+\mu_3-3\mu_4=0$ by t-test. Does the conclusion changes with Bonferroni correction? Does it change with Schéffe's method? Which procedure should be used here (Bonferroni, Schéffe, other correction method, or no correction at all)?
Add a part (e)
(e) It is clear from the data that the group with intermittent claudication has much higher LDL cholesterol levels than other groups. After seeing the data, we want to test H_0: $\mu_1-(\mu_2+\mu_3+\mu_4)/3=0$ by t-test. Carry out the test. Is any multiple testing correction procedure needed? If so, which one? If not, why not?

4. (10 points)
For the data set *airquality* contained in the R base package (there is no need to input it, see the example in Lab1), we want to know for which pairs of months the mean Ozone measurements are different. (You can use "**?airquality**" to check the description of the data set.)
(a) Do the multiple pairwise comparison test at 0.05 level, using Bonferroni, Tukey, and fdr corrections respectively. What are your conclusions?
(b) Are the results for those three correction methods the same? If not, which you think is the more appropriate conclusion here? Why?

5. (5 points)
For the data set *lowbwt* used before, systolic blood pressure measurements for 100 low birth weight infants are saved under variable sbp, gender is saved under variable sex and Preeclampsia (formerly called toxemia, a complication of pregnancy for the child's mother) is saved under the variable "preeclampsia". (For the lowbwt.csv from the website of the 3rd edition of the textbook, the variable is coded as "preeclampsia". If you use the data file from previous edition, that variable is coded as "tox".)
We wish to see whether mean systolic blood pressure is the same for low birth weight boys and girls.
(a) Use R to produce the ANOVA tables: one with preeclampsia status as the blocking variable and one without any blocking.
(b) What is the p-value for gender effect on systolic blood pressure with the blocking? What is the p-value for gender effect on systolic blood pressure without the blocking?

6. (10 points) Mini-Project: measuring your response time.
Using the webpage in Lab 1 to collect 30 response times each with all four sizes of boxes: "small", "medium", "large" and "xlarge".
(a) Conduct an ANOVA to see if there is a difference between your response times for different box sizes.
(b) Conduct the pairwise comparisons between groups using HSD and LSD.