STATISTICAL METHODS II
ASSIGNMENT 03
DUE: 1 FEBRUARY 2011

This homework assignment deals with problems concerning comparing means of multiple groups. Please make sure you read the questions thoroughly and think about them *before* you begin your answer. Two of the three questions use real data. As always, you will need to use `R` to answer it. Download the data sets from the web site. The filenames are given in the individual problems.

Your answers to the questions must be nicely typed. The answer should be at least paragraph in length and should follow the same pattern in what information is included:

- State the problem.
- State the null and alternative hypotheses in words.
- State the test you will use, its assumptions, and why you chose this test.
- In your answer, include the value of the test statistic, the degrees of freedom (if applicable), and the calculated p-value.
- Clearly draw the appropriate conclusion.

When you hand in this assignment, attach your `R` script to the back of the pages and include graphs immediately after (or with) the problem.

If you have any questions or issues, let me know as soon as possible.

Good luck!

Problem 03.1 [[3]]

There is a ubiquitous dataset in Statistics. It was compiled by Edgar Anderson at the request of Sir R. A. Fisher (a name you will hear quite often). This dataset has been used to illustrate topics such as discriminant analysis, categorical analysis, classification, and simple regression. We will use the data set to illustrate Analysis of Variance.

There are three species of iris in the dataset (*Iris setosa*, *Iris virginica* and *Iris versicolor*), with four features measured for each of the 50 flowers (the length and the width of both the sepal and the petal, in centimeters). The species name is a categorical variable (grouping variable). The four measurements are continuous variables. The species name will be the independent variable. The measurements will be dependent variables.

For now, I just want to know if the petal widths are the same across the species; that is,

$$H_0 : \mu_s = \mu_v = \mu_c$$

where μ_s is the expected petal width for the *Iris setosa* variety; μ_v, for the *Iris virginica* variety; and μ_c, for the *Iris versicolor* variety.

Before you actually perform the analysis, you will need to create a boxplot (properly annotated) and test the primary assumption of ANOVA. If the assumption is violated by this data, then you will still use ANOVA, but you will also need to use the Kruskal-Wallis test. Note that the dataset comes with `R`, so you merely need to add the `data=iris` parameter to your functions.

Problem 03.2 [[3]]

A few weeks back, a professor I know made the statement that corruption is killing Africa. The level of honesty in government in Africa is lower than in any other region in the world. This seems to be the common wisdom. However, it is true? Using the dataset `gdpcap`, determine if common wisdom is correct; that is, determine if the Africa region has a significantly lower level of honesty in government (`hig`) than the other regions.

Again, make sure you produce a nice boxplot and test the major assumption of ANOVA before you perform the test. If the assumption is violated in this dataset, use the Kruskal-Wallis test.

Problem 03.3 ⟦4⟧

The dataset `chickwts` also comes with `R`. It consists of two variables: `feed`, which is a categorical variable (factor) giving the type of feed given to the chickens, and `weight`, which gives the weight of each of the 71 chickens.

The chicken ranchers want to know if there is any statistical difference in the six feed types in terms of producing heavier chickens.

Produce an appropriately-labelled boxplot, determine the null hypothesis, and test that hypothesis. Come to the appropriate conclusion (appropriate in terms of using the correct test for the correct reason). Remember, since the dataset comes with `R`, just add the `data=chickwts` parameter to your functions.