STATISTICAL METHODS II
ASSIGNMENT 06
DUE: 22 FEBRUARY 2011
(REVISED)

This homework assignment deals with both new and old. We are beyond the point of me specifying which test you need to use and to the point that you select the best test yourself. This assignment gives you practice doing that. As usual, please make sure you read the questions thoroughly and think about them *before* you begin your answer. The first two questions use pseudo-data; the last, real data. As always, you will need to use `R` to answer it. Download the data from the web site (or link to them in your script).

Your answers to the questions must be nicely typed. The answers should be as long as they need to be in order to prove your point. Think of these as reports to clients. As such, this assignment also has you create reports that look better—presentation counts more here than previously.

When you hand in this assignment, attach your `R` script to the back of the pages as an appendix. The graphs need to be woven in your narrative; that is, meaningfully refer to them in the text, explain what the graph tells us, and number the graphs. You can still include them all at the end of the homework if you wish, or you can put them in the body of your assignment. Include statistics appropriately, making sure that you explain the statistic and what it means, rather than just providing it. In fact, put your statistics and degrees of freedom and p-values in parentheses. Finally, make sure you provide the name for the test, not the `R` function. The only place I should see anything from `R` is in the appendix.

By the way, the test's name is the Shapiro-Wilk test, not the Shapiro test. Also be careful about the difference between the Kruskal-Wallis test and the Kruskal test.

Finally, as usual, if you have any questions or issues, let me know as soon as possible.

Good luck!

Problem 06.2 ⟦5⟧

The biological half-life of a drug is the amount of time it takes for one half of the drug to leave a body. Thus, if the half-life of Drug A is 3 hours, one half of the original amount exists after 3 hours; one-quarter after 6 hours; one-eighth after nine hours. Unlike radioactive half-life, there are various factors that affect the biological half-life of a drug—including other drugs.

The biological half-life of caffeine is five hours. Because it would be beneficial to have a longer half-life for caffeine (fewer trips to Starbucks in the mornings), I would like to find a drug that interacts with the caffeine, lengthening its biological half-life. I hit upon a cheap drug that might help prolong the effects of caffeine—acetylsalicylic acid (aspirin).

To test whether aspirin extends the effects of caffeine on the body, I made the trek to the OSU institutional review board (IRB), which makes sure my experiment does not violate the rights or welfare of my subjects (humans). They must have been drunk (or enjoyed my donation to their "coffee fund"), because they approved my experiment. With approval in hand, I posted fliers around campus announcing this exciting drug trial. With this advertising (and some monetary offers), I had 500 students in my trial group.

For a week, I gave each person an aspirin tablet (500mg). Five minutes later, I gave each person 125mg of caffeine dissolved in liquid. I then took blood samples at one-hour intervals. Running an assay on the blood, I was able to determine the amount of caffeine remaining in the blood at the various times (at time zero, there was 62.5g). I then tabulated the results, which are given below. The table contains the total amount of caffeine in all of the students' bloodstreams at the given time interval. It also contains the expected total amount of caffeine in all of the students' bloodstreams under the null hypothesis that the biological half-life is five hours.

Did taking the aspirin help extend the effects of caffeine?

Time (hr):	1	2	3	4	5	6	7	8	9
Expected (g):	51	42	34	28	23	19	15	13	10
Observed (g):	49	38	30	23	18	14	11	8	7

Note. *You can do this by hand, including looking up the p-value in a table in the back of your book. You may also use this as an opportunity to write an* R *function. If you want me to look at your algorithm (Step 4) or your code (Step 5), I will be more than happy to do so, giving much feedback. If you create a function that works, attach it to your assignment and get a point extra credit. If you want to use the function I wrote, email me and pay one point from this assignment.*

Problem 06.3 ⟦5⟧

In January 2010, there was a presidential election in the island state of Sri Lanka. The election pitted incumbent president Mahinda Rajapaksa against former general Sarath Fonseka. As both were responsible for ending the insurgency that had ravaged their state for a score years, the opinion polls had them in a dead heat.

The official counts showed that the dead heat never materialized; Rajapaksa won 58 percent of the vote, while Fonseka won 40 percent. Of course both sides accused the other of electoral fraud, but only Fonseka found himself in prison facing the death penalty for treason. The interesting thing is that the election returns strongly suggest electoral fraud. Here is how:

In a free and fair election, the proportion of the vote declared invalid is independent of the proportion of the vote for any major candidate. In other words, if the proportion of the vote declared invalid is different in provinces supporting different candidates, there is something improper taking place. This is exactly what we see in this election.

The proportion of the vote declared invalid (`REJECTED/TOTAL`) is much higher in provinces supporting Fonseka than in provinces supporting Rajapaksa.

Using the `sri2010pres` dataset, support this last statement with graphs and statistics.