STATISTICAL METHODS II
MIDTERM EXAMINATION I
DUE: 1 MARCH 2011

This examination covers the most important things we covered during the first third of the course. I will not tell you which test to use; use the correct one(s). I will not tell you to provide a graph; provide a well-labeled graph.

Make sure your graphs are well-labeled. The axes need to be labeled. The units need to be labeled. The graph needs to be titled (`main`). The axis values (`las`) need to be horizontal. Make sure the margins for your graphs look good and do not cut off axis labels.

Your examimation answers must be nicely typed. The answers should be as long as they need to be in order to fully answer the question. Grammar counts. Explain in detail what test you are using, what it tests, what the null hypothesis of the test is, what the conclusion of the test is (with test statistic, degrees of freedom, and p-value in parentheses). Explain a lot. Re-read your answers and make sure they logically answer the question posed.

In your answers, include statistics appropriately. Finally, make sure you provide the name for the test, not the `R` function. The only place I should see anything in `R` is the appendix.

When you turn in this examination on Tuesday, 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 (before the `R` Appendix), or you can put them in the body of your assignment.

Start each answer on a new page. Only print on one side of the paper.

Finally, as usual, if you have any questions or issues, let me know as soon as possible. The worst I can do is not answer your question. You have access to all non-living sources.

Problem 1 ⟦5⟧

Let us start out with something rather easy. You have heard me say "We want to use a t-test (or an analysis of variance procedure). Let us see if we can." What is the main strength of parametric tests? What is the main weakness?

Problem 2 [[5]]

Power is important in statistical tests. We spent two days discussing (and testing) the power of a test. What is the relationship between power and the Type I error rate? What is the relationship between power and the Type II error rate? If I increase the α from 0.05 to 0.10, will the power of the test increase or decrease? If I increase the variance in the two samples, will the power of the test increase or decrease? Finally, what is power?

Problem 3 ⟦5⟧

Tom has a dataset consisting of measurements on three groups. The measurement is the number of minutes after sunrise that the goose flock flies away from the lake. The measurements range from -90 to 84 minutes. His research question concerns what affects the early flight time of goose flocks. The three groups are three locations.

Tom performs the necessary tests to determine if he can use an analysis of variance procedure. He discovers that the Shipiro-Wilk test fails, so he cannot use that procedure. Mary tells Tom to square all of his measurements (y-values) and see if the transformed data still fails the tests. Tom does so and finds that he is now able to perform analysis of variance.

Tom, however, is worried that the conclusions based on the transformed data are not valid for the untransformed data. Tom may be right.

Under what circumstances will the conclusions based on the transformed data be valid for the untransformed data? Under what cirsumstances will the conclusions not necessarily give appropriate results?

Problem 4 [[5]]

Timmy was faced with a dataset containing six groups. In order to determine if any of the six groups had a median that was different from the others, he used 15 pairwise Mann-Whitney tests. The results of the tests are provided below in Table 1.

What appropriate conclusions should Timmy draw regarding the differences in the groups? Make sure you explain your reasoning.

Group	1	2	3	4	5	6
1		0.56	0.06	0.07	0.65	0.11
2			0.04	0.03	0.08	0.05
3				0.04	0.06	0.06
4					0.25	0.66
5						0.55
6						

Table 1. *Table of p-values for the pair-wise comparisons between the six groups.*

Problem 5 [[5]]

Given a standard 2×2 table, how does one calculate the accuracy of the test? Write a function in `R` that calculates the accuracy when given a 2×2 table.

To help solidify things, let us suppose that I have a medical test that performs as follows:

	Actual	
	Diseased	Healthy
Predict: Diseased	210	30
Healthy	45	167

Think: What are the steps to writing a function? How should I calculate the accuracy of the test? How do I translate that calculation into `R`?

PROBLEM 6 ⟦20⟧

Let us revisit an old dataset that we have already worked with—the Sri Lankan 2010 Presidenital election (`sri2010pres`). The dataset contains the number of votes for Mahinda RAJAPAKSA, Sarath FONSEKA, along with the number of votes declared invalid (REJECTED), the TOTAL number of votes cast, the number of people REGISTERED in the electoral DIVISION, as well as whether or not Rajapaksa won the specified province (WONPROVINCE). In addition to these count data, the PROVINCE, DISTRICT, and electoral DIVISION are provided (akin to our state, county, and precinct).

(1) First, drop all "Displace" and "Postal" DIVISIONS. Why would that be necessary? (Think about assumptions and calculations and what happens if you do not do this removal.)

(2) Second, create a boxplot, by PROVINCE, comparing the relative support for Mahinda Rajapaksa (proportion of the vote for Mahinda Rajapaksa in that province). There should be nine boxes in this plot. Make sure this graph is labeled correctly (*per* the directions). Also make sure you explain in your narrative what the boxplot is showing.

(3) Third, determine whether (and to what extent) these nine provinces can be grouped into fewer province-groups. In other words, determine which provinces are statistically different (and similar). Group those that are statistically alike.

(4) Fourth, if this cannot be done with the data, explain why.

Make sure you explain everything the statistics and boxplot tell us.

Problem 7 ⟦20⟧

One test frequently used (mis-used, really) in detecting electoral fraud is the First-Digit Benford Test (BENFORD-1). Under the assumption that there is no electoral fraud in the counting of the ballots, the first digit of the counts has a distinct distribution (See Table 2). Thus, if there are 1000 counts, then we would expect 301 of the leading digits to be a '1', 176 to be a '2', etc. If my actual counts were 143, 1555, 645, 323, and 14, then I would observe three leading '1's, one leading '3', and one leading '6'. For a sample of size 5, I would expect $5 \times 0.301 = 1.505$ leading '1' digits, $5 \times 0.176 = 0.880$ leading '2' digits, etc.

Now, using the Sri Lankan 2010 Presidential election dataset (`sri2010pres`), determine if the counts at the electoral DIVISION level fail the BENFORD-1 test. Make sure you supply an appropriate graph and appropriate tables, as well as appropriate statistics and p-values.

Leading Digit	1	2	3	4	5	6	7	8	9
Expected RF	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Table 2. *Table of Expected relative frequencies (RF) for each possible leading digit in the BENFORD-1 test.*

PROBLEM 8 ⟦20⟧

The t-test was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin, Ireland. He actually returned to school to study statistics because he discovered that his usual tests that assume Normality were not rejecting the bad beer at the correct level—too much was being wasted. The reason the Normal test was not the correct was because he had to estimate the variance of the population using the sample. A Normal test is only proper if the variance of the population is known.

With that bit of history in mind, let us do some tests on beer in honor of Dr. Gosset, who wrote under the *nom de plume* of 'Student.'

⋆⋆⋆

Believing that my favorite beer is the best-tasting beer around, I decide to perform a taste test among the faculty. I select five brands of beer (comparable to my favorite) and buy several kegs of each. I then went to all of the faculty meetings on campus over the course of a month, giving each professor a single glass of beer (randomly selected by me, one keg per meeting). Each professor would then rate the beer from 1 to 5, where 5 indicated 'The Best Beer of All Time!' and 1 indicated 'dog water.'

After traveling to each of the departments on campus once, I tallied my results, and placed the data in the `beer1` dataset.

(1) What conclusions should I draw, regarding the most-preferred beer?
(2) Comment on the strength of the experimental method.
(3) Of course, supply graphs and statistics as appropriate.