Statistical Methods II
Final Examination

Due: May 5, 2011
10:00am

This examination covers the most important things we covered during 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. I will not tell you that you need to analyze every aspect of the information; analyze it, especially when the answer looks too easy.

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` 'speak' is the appendix.

When you turn in this examination on Thursday, 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.

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: A Flighty Problem [[5]]

In the United States (and with US airlines), in the 10-year period 1965–1975 prior to Congress passing the Airline Deregulation Act of 1978 (PL95-504), the number of crashes per year was 3 (with a variance of 1). In the 10-year period 1985–1995 following airline deregulation, the average umber of crashes per year was 7 (with a variance of 2).

The research question is "Did deregularion cause the annual average number of airline crashes to increase?"

Please answer the research question.

Problem 2: Where's Old MacDonald? [[5]]

Farmer Jim currently uses Gold Start© fertilizer on his wheat, getting an average annual yield of 76 bushels per acre (with a variance of 7). His friend, Farmer Bob, uses Diamond© liquid plant food on his wheat, getting an annual yield of 84 bushels per acre (with a variance of 5). Diamond© costs 5% more than Gold Start©.

Should Farmer Jim change fertilizer?

Problem 3: Airport Fodder [[5]]

A recent poll discussed on CNN showed that 70% of Americans (± 2 at a 95% confidence inteval) think the country is moving in the wrong direction. This is larger than six months ago, when only 60% said the same thing (± 2 at a 95% confidence interval).

The commentator said this meant President Obama is in real trouble for the 2012 election, as 70% of the electorate no longer supports his agenda.

Is this conclusion appropriate?

Problem 4: A Power Defect [[5]]

Kip, a manager of a production line that produces 5000 units per hour, wants to determine if last week's power outage adversely affected the ability of the primary machine to create acceptable parts.

He gathered a random sample of three parts from before the power outage and discovered that none of them were defective. He collected a random sample of four from after the power outage and discovered that none were defective. He concluded that the power outage had no adverse effect on production quality.

Is his conclusion appropriate?

Problem 5: The Full Monte [20]

Our social science theory tells us that the population has a characteristic that is distributed uniformly between 0 and 100, with an average of 50. You take a ramdom sample of size $n = 35$ people and measure this characteristic in each person. You then perform a t-test of the null hypothesis $H_0 : \mu = 50$. You realize that the t-test assumes the measurements are distriuted Normally; however, you still use the t-test.

Is $n = 35$ a large enough sample to ensure that the p-values are meaningful for a t-test on this population?

Problem 6: Hit the Upper V [[20]]

I taught high school for eight years. By the fourth year, I noticed that there appeared to be a difference in the classroom performance of the students based on their primary sport. To put this observation to the test, I kept track of the grades of students and their sport for my Honors Pre-Calculus class. In 2004, the ten soccer players got 5 As, 4 Bs, and 1 C. The ten basketball players got 3 As, 6 Bs, and 1 D. The softball players got 1 A, 4 Bs, and 3 Cs. Luckily, there was no overlap among the three groups, so everything is independent (as we like it).

My research question is "Is the sport of choice for the high school student independent of that student's grade in my Honors Pre-Calculus class?"

Answer the research question for these three sports.

Problem 7: It's in the Gbagbo [[20]]

On 28 November 2010, Côte d'Ivoire held a long-awaited presidential election (its first attempt at a free and fair election in your lifetime). The official results from the Independent Electoral Commission (IEC) showed that Alassane Ouattara received 54.1% of the votes cast, whereas President Laurent Gbagbo received only 45.9%. Gbagbo refused to accept the decision of the IEC, as the Constitutional Court (CC) stated Gbagbo received 51.45% of the votes (with 48.55% to Ouattara).

With Gbagbo refusing to step down, Côte d'Ivoire was thrown, once more, into civil war (or a short crisis). Eventually, Ouattara won the battle and currently sits as president of Côte d'Ivoire.

Unfortunately, the Constitutional Court did not post the actual vote counts, so we cannot analyze *their* counts; however, the Independent Electoral Commission did. Using the `cdi2010pres2` data, do the following:

- Create a boxplot of the proportion of votes rejected according to who won the *region*.
- As the two distributions significantly different?
- Create a scatterplot of proportion of the vote rejected against the proportion of the vote in favor of Ouattara.
- Plot the regression curve (prediction curve). Include 95% confidence bands.
- Are these two variables independent?

Problem 8: Friends, Romans, Countrymen! ⟦20⟧

I am currently involved in a project with a researcher in DC who wants to determine (among other things) if receiving academic earmarks helps or hurts a university's chances of receiving peer-reviewed funding.

On each budget, the US Congress (both the House and the Senate) add earmarks, which provide direct federal funding to certain non-Federal agencies to allow them to carry out their duties. Earmarks destined for universities are called Education Earmarks. In 2003, Oklahoma State University received 15 earmarks totalling $12,348,204. The vast majority of the earmarks (in number and value) were for the Department of Defense.

A second option for funding research is to apply to the National Science Foundation (NSF), the National Institutes of Health (NIH), or the National Institutes for Mental Health (NIMH) for peer-reviewed funding. Peer-reviewed funding requires other scientists to approve the research. In 2003, Oklahoma State University received $52,859,000 in peer-reviewed funding from the NSF.

The Carnegie classification system divides all schools of higher education into at least a dozen categories. The two categories of interest for us are the RU/H and the RU/VH schools. 'RU' stands for 'Research university,' with 'H' meaning high level and 'VH' meaning very high level. OSU is RU/H (as is OU).

Using the `earmarks` data, find an appropriate model that predicts peer-reveiwed funding (in dollars) given earmarks received (in dollars) and the Carnegie classification.

Plot the peer-reviewed funding received (in dollars) against the earmarks received (in dollars). Plot a prediction line for the RU/H schools. Plot a prediction line for the RU/VH schools. Will you use different colors? Will you provide a legend? I hope so.

Now, we know that Oklahoma State University received $400,000 in earmarks in 2010. We also know that Oklahoma State University is an RU/H school. Predict the value of the peer-reviewed funding we received (from the NSF). Comment on the accuracy of your estimate.