

# Statistical Methods II

## Assignment 14

Optional: April 26, 2011

And here we are at long last. I must admit I will miss each of you; it has been an enjoyable course. I wish it could continue for several more weeks, but I think we need a brain-break to allow what we have done a chance to settle in our minds. I hope y'all have learned as much as I hope you have.

Think of this final (optional) homework assignment as a practice test for this past third of the course. I will post solutions and the script on Tuesday as class begins. That will give you time to work on this assignment as if it were real, thus showing you where some weaknesses exist from this third part of the course.

\*\*\*

Do not expect the final to look exactly like this homework; it will be a sampling of the *entire* scope and sequence we covered in this course. I will be in DC this weekend (Friday to Sunday). I will be back Sunday late, so answers to your questions may have to wait until Monday. With that said, email me any questions and I may be able to answer the questions sooner than expected.

Ciao!

## PROBLEM 1: THE LAND OF ENCHANTMENT

[[4]]

Go back through the script from April 21, 2011. In that script, we predicted the probability that the SSM ballot measure would pass. During the prediction, I rambled on about why we needed to subtract 1998 from the year to create a new variable.

- (1) Redo the analysis, but use the *actual year* instead of the modified year.
- (2) Make a prediction of the proportion of the vote in support of the SSM ballot measure for New Mexico.
- (3) Now, using this model, what is the probability that the ballot measure will pass in New Mexico?
- (4) Create a histogram and interpret it.

## PROBLEM 2: LASSO THE GATERS!

[[4]]

Use the `ncaa2009football` data to do the following:

- (1) Which team scored the most points on average? Is the average statistically larger than that of the other teams?
- (2) Which team won by the most points, on average? Is this average statistically larger than that of the other teams?
- (3) If Oklahoma State University were to play the University of Florida, what is the probability that OSU would win?

## PROBLEM 3: A NOBEL PRIZE FOR DR. PAULING

[[2]]

Use the `vitaminC` data to do the following:

- (1) Using `R` and only `R`, what proportion of the sample had a cold? What proportion took Vitamin C supplements?
- (2) How many of the cold sufferers took Vitamin C supplements? Assuming this experiment was properly randomized (a complete random design), did the Vitamin C supplements help the people from catching a cold?
- (3) If I take Vitamin C supplements, what is the probability that I will catch cold, all things being equal? If I decide to not take Vitamin C supplements, what is my probability of catching a cold? Is this difference statistically significant?

## PROBLEM 4: THE DISTRIBUTION IS FISHY!

[[4]]

Use the `fakepoisson` data to do the following:

- (1) Create a nicely labeled histogram for the variable `y`. Is `y` distributed Normally? Prove it!
- (2) Theory tells me that `y` comes from a Poisson distribution with a mean of 20. What is the mean of the `y` sample?
- (3) Does it appear as though `y` is distributed Poisson with a mean 20? Use a t-test.
- (4) Is the sample size large enough for the t-test to give meaningful p-values? Evidence? (Hint: April 19. Use 100,000 trials. Make sure the sample size of your random distribution is the same as that for your real sample. The R function for the distribution is `rpois(n, lambda= m)`, where `m` is the hypothesized mean — the mean under the null hypothesis.)