Statisticians,

The second intellectual festival (AKA second exam) will cover all of the material in modules eight through fifteen and will have questions from the following list:

1) ~15 multiple choice questions. These tend to be related to specific facts (e.g., definitions & symbols) or very short calculations.
2) Three or four short answer (paragraph-length) questions from among the following:
   a. What are the two major goals of regression? Give a specific example to illustrate each goal.
   b. What are the two major assumptions of regression? Draw plots that illustrate situations where these assumptions are (a) both met, (b) one is violated, and (c) the other is violated.
   c. Completely describe all differences between a population and a sampling distribution.
   d. Completely describe the difference between a standard deviation and standard error.
   e. What are the two major methods or techniques for performing statistical inference? Completely describe the differences between the objectives for the two methods?
   f. Describe choices that you, as a researcher, can make to reduce β. Which is the best choice to make and why?
   g. Describe choices that you, as a researcher, can make to reduce the margin-of-error. Which is the best choice to make and why?
   h. What is the difference between "statistical significance" and "practical significance"? Explain why is it important to make a distinction between these two forms of significance?
3) Percentage table calculations from a two-way frequency table (e.g., row, column, table percentages). [*module 8*]
4) A bivariate EDA for quantitative data. [*module 9*]
5) Linear regression questions (some subset of the 12 typical questions that I ask). This will require you to examine a fitPlot() with a best-fit line superimposed with the best-fit equation and r^2 included. [*module 10*]
6) "What is the probability …" questions. This will require you to use distrib() in R. [*modules 11 & 12*]
7) Form statistical hypotheses from a research hypothesis. [*module 13, class exercise 1*]
8) Compute a p-value and make a decision. [*module 13, class exercises 2&3; module 15*]
9) Identify a Type I and Type II error for a situation and assess which one would be more egregious. [*module 13, class exercise 4*]
10) Calculate statistical power. [*This will be an extra credit question*] [*module 13, class exercise 5*]
11) Calculate confidence regions. [*module 14, class exercises 2&3; module 15*]
12) Calculate required sample size. [*module 13, class exercise 4*]

The following R help will be provided to you at the exam.

---

library(NCStats)
distrib(x,mean=##,sd=##,lower.tail=XXXXX,type="X")

where **x** is replaced with the value of the quantitative variable or the area

      mean=## has ## replaced by the value of the mean

      sd=## has ## replaced by the value of the standard deviation

      lower.tail=XXXXX has XXXXX replaced with TRUE (the default) for a "left-of" and FALSE for a "right-of" calculation

      type="X" has X replaced with p (the default) for a forward and q for a reverse question

---

The exam is closed book, closed notes, etc. You should bring a calculator and you MUST use a pencil. Exams written in red ink will not be accepted. In full disclosure -- there will be multiple versions of the exam so please do not embarrass yourself, and earn an "F" for the class, by cheating from your neighbor. Your RStudio must open with either no scripts or a blank script – i.e., the upper left pane must not have any previously entered R code in it. Opening any other software

or previous script will result in an automatic "F" for the class. I will be monitoring computer usage and you will only be allowed to open RStudio.

The exam will start promptly at the beginning of class (on the hour) and will end promptly 1 hour and 50 minutes later. Please let me know ASAP if you have any conflict with staying 20 minutes past our usual class ending time.

We can discuss this more in class, but also let me know via e-mail if you have any questions.