Statisticians,

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

1) ~15 multiple choice questions.
2) Three or four short answer (paragraph-length) questions from among the following: [*NOTE: that "describe" and "explain" mean more than just "list" … you need to be detailed with your answers.*]
    a) What are the two major goals of regression? Give a specific example that illustrates 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 three major differences between a population and a sampling distribution.
    d) Completely describe three major 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 ☐.Which is the best choice to make and why?
    g) Describe choices that you, as a researcher, can make to reduce power. Which is the best choice to make and why?
    h) Describe choices that you, as a researcher, can make to reduce the margin-of-error. Which is the best choice to make and why?
    i) 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) A bivariate EDA for quantitative data. [module 10]
4) Linear regression questions (some subset of the 12 typical questions that I ask). Note that this question will have two sets of results (a figure with the best-fit line, the equation for that line, and the r^2 value) -- one where the response variable is on the y-axis and another where the response variable is on the x-axis. Thus, you will have to correctly identify what the response variable is and then choose the proper set of results to answer the questions. In other words, only one set of the results are valid and should be used [module 11]
5) "What is the probability …" questions. This will require you to use distrib() in R. [modules 13]
6) Form statistical hypotheses from a research hypothesis. [module 14]
7) Compute a p-value and make a decision. This will require you to use distrib() in R. [modules 14 and 16]
8) Identify a Type I and Type II error for a situation and assess which one would be more egregious. [module 14]
9) Calculate β.
10) Calculate and interpret confidence regions. This will require you to use distrib() in R. [modules 15 and 16]
11) Calculate required sample size. This will require you to use distrib() in R. [module 15]

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

```
library(NCStats)
distrib(val,mean=meanval,sd=sdval,lower.tail=FALSE,type="q")
```

where
- **val** is a value of the quant. variable or area (i.e., percentage as a proportion)
- **meanval** is population mean (μ)
- **sdval** is standard deviation (σ) or error (SE)
- **lower.tail=FALSE** is included for "right-of" calculations
- **type="q"** is included for reverse calculations

The exam is closed book, closed notes, etc. You should <mark>bring a calculator</mark> 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. **You must use RStudio on a Northland College computer** (i.e., you will not be allowed to use your personal computer). 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. In addition, I will be monitoring computer usage and you will only be allowed to open RStudio.

The exam will start promptly at 0800 for the first section and 1000 for the second section. The exam will end promptly 1 hour and 50 minutes later. Please let me know ASAP if you have any conflict with starting 20 minutes earlier (for the first section) or ending 20 minutes later (for the second section).

There are many resources available to you to prepare for this exam. Among these are

1. Module readings and videos.
2. Module homeworks (answer keys are on the Resources page).
3. Module class exercises (answer keys are on the Resources page).
4. Module review exercises (many of these are old test questions; answers are linked to).
5. Tutors are available to answer questions (schedule on the syllabus)
6. I am available to answer questions.

I suggest that you actually do the homeworks, class exercises, or review exercises. Simply being familiar with the questions will likely not prepare you well enough to answer the questions on the exam in the amount of time provided. I also suggest preparing for the exam in two hour blocks without distraction (no interruptions from your phone, no music, etc.), so as to best replicate what the conditions at the exam will be like.

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