Statisticians,

The first intellectual festival (AKA first exam) will cover all of the material in the first nine modules and will have questions from the following list:

1) 15-20 multiple choice questions. These tend to be questions related to specific facts (e.g., definitions & symbols) or very short calculations.

2) Three short answer (paragraph-length) questions from among the following:

   a. Thoroughly describe what the two major goals of statistics are and why each is important.
   b. Define natural and sampling variability. Provide a thoughtful example that depicts each type of variability.
   c. Describe two major principles or realities that lead to the importance of statistics in everyday life and scientific research.
   d. Completely describe the underlying philosophical differences in how the mean and median measure center.
   e. Describe how and why you would decide to use either the mean and standard deviation or the median and IQR to measure center and dispersion in a univariate EDA for quantitative data.
   f. Describe the major differences between an observational and experimental study.
   g. Describe the major principles of experimental design and why each is important?
   h. Describe several situations (be specific) where observational studies are valuable (even though strict cause-and-effect statements cannot be made).

3) An IVPPSS. [*module 2*]

5) Experimental design questions (what is response, factors, etc.). [*module 3*]

4) Calculate (by hand, showing your work) the mean/sd for a short (~6) list of numbers. [*module 5*]

5) Calculate (by hand, showing your work) the median/IQR for a short (~20) list of numbers. [*module 5*]

6) A univariate EDA for categorical or quantitative data. This will require you to extract information from R output that I will provide -- including a histogram and summary statistics for the quantitative case and a frequency and percentage table for the categorical case. You will NOT need to use R commands to produce histograms, summary statistics, boxplots, frequency tables, percentage tables, or bar charts. [*modules 6 & 7*]

7) Percentage calculations from a two-way frequency table (e.g., row, column, table percentages). [*module 9*]

8) Three to five normal distribution questions (one background, three to five questions related to that background). This question will require you to use distrib() in R to produce output to answer the questions. I will provide a reminder of the arguments to distrib() that will look like that below (from the R cheatsheet). [*module 8*]

```
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 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. 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.