

Statisticians,

The first intellectual festival (AKA first exam) will cover all of the material in the first seven 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 is each 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.
- 5) Experimental design questions (what is response, factors, etc.).
- 4) Calculate (by hand, showing your work) the mean/sd for a short (~6) list of numbers.
- 5) Calculate (by hand, showing your work) the median/IQR for a short (~20) list of numbers.
- 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.
- 7) 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 the following ....

---

```
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. In addition, I will be monitoring computer usage and you will only be allowed to open RStudio.

The exam will start promptly at 1000 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 (there should be no classes that begin at 1130).

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