Math 58B - Exam 2 Preparation - Spring 2023

**Logistics:**

• Thursday, April 27, 2023 – take exam at any time during the day, use as long as you need.

• Material up through Friday, April 21, 2023

• Two sides of notes are allowed (one piece of paper)

• Bring calculator, no computers

**Overview:**

We have analyzed studies that involve one quantitative variable, where the data are a sample (ideally, a random sample) from a large population. Using one quantitative variable, we also created prediction intervals for individual responses. We have also analyzed studies that involve comparing a quantitative variable across two groups, particularly those coming from a randomized experiment.

We use chi-square analyses to extend the two proportion test to many groups. We use ANOVA to extend the two means test to many groups.

We modeled data measured with two quantitative variables using a linear model. We have now moved from a descriptive model to using the model for inferential statements about a population linear model. We have also described the technical conditions and checking the technical conditions via residual plots. Additionally, we have started to add multiple variables to the regression model.

We continued to use graphical and numerical summaries for understanding the data at hand. We have used inference methods based on simulation/randomization, bootstrapping, and also approximate methods based on a mathematical distribution (we have now seen the following distributions: normal (must specify mean and st dev), t (must specify df), chi-squared (must specify df), and F (must specify two different df)).

We have studied two primary types of statistical inference:

* Statistical significance, where the goal is to assess the degree to which the sample data provide evidence supporting a research conjecture (i.e., test a hypothesis and obtain a p-value);
* Statistical confidence, where the goal is to estimate a population parameter with an interval of plausible values (i.e., create a confidence interval).

We have studied three ways to conduct statistical inference:

* Randomization tests
  + Tactile (e.g., with coins or cards) & Technology (e.g., with applet or infer syntax)
* Bootstrap confidence intervals
  + Tactile (e.g., drawing marbles from a bag) & Technology (e.g., with applet or infer)
* Mathematical modeling
  + When technical conditions are satisfied & with technology (e.g., xpnorm in R)

We have also considered how the scope of conclusions to be drawn depends on how the data were collected. More specifically:

* Random assignment allows for the possibility of drawing cause/effect conclusions.
* Random sampling allows for generalizing to a larger population.

We have examined confidence intervals for different parameters:

* Population mean
* Difference in population means
* Slope coefficient
* (Additionally: prediction intervals for **individuals** (not parameters) are created when the response variable is quantitative.)

We have hypothesis tests for different parameters:

* Difference in population proportions / probabilities across two or more groups (e.g., chi-square)
* One population mean
* Difference in population means
* Equality of many population means
* Slope coefficient

**Outline:**

* Chi-square test of independence (two variables, each ≥ 2 categories)
* Sampling distribution of differences of sample means, Central Limit Theorem for Sample Mean, standard error of sample mean
* t-test and t-interval for inference on one population mean; bootstrap interval for one population mean
* t-interval for predicting individual observations
* Two-sample t-test for difference in population means; randomization test for difference in population means
* Two-sample t-interval for difference in population means; bootstrap interval for difference in two population means
* Power calculations for difference in population means
* Effects of within-group variability (s), sample size (n), difference in group means on: the sampling distribution and power
* ANOVA
* dangers of extrapolation
* sampling distribution of b1, factors influencing SE(b1)
* LINE conditions (what they are and why we need them)
* reasons for transforming data (i.e., to make sure the LINE conditions hold)
* interpreting a linear model with multiple explanatory variables

No Binomial Distribution (section 5.2 in the notes) or Paired t-test (section 6.4 in the notes) on the exam!

**Advice:**

• Organize notes for efficient retrieval of information/formulas

• Don’t plan to use cheat sheet too much

* Prepare as if exam were closed book/notes
* **Focus on understanding, not memorization**

• Expect similar questions to what we answer in class every day, clicker Q, on HW, labs

• Be ready to interpret computer output

* Possibly excluding irrelevant output

• Be prepared to **think/explain/interpret**

* Do not just plug into formulas
* Be ready to explain process of how you would do calculations

• Read carefully

* Be sure to answer the question asked

• Take advantage of information provided

* Perhaps including output

• Relate conclusions to context

• Practice

* Work through solved examples at end of text chapters
* Re-work in-class examples, HW, labs, clicker questions (without looking at solutions)
* Be able to answer all of the reflection questions (in class notes)

• Give yourself enough time!

* Be cognizant of time constraints
* Attempt all questions
* Sleep well & eat well before the exam