Math 247: Review for Final Exam

Suggested review: Redo all old exams.

You will need a calculator (any type is fine) for the exam.

You may have an 8½ x 11 sheet (both sides) with notes and formulas on it. No formulas will be given on the test so make sure you put down any formula that you think you'll need!

The BIG ideas of this course:

Inferential Statistics: Using information from a sample to make inferences (using either a Hypothesis Test or a Confidence Interval) about a population.

What are the steps of hypothesis testing?

What types of hypothesis tests have we studied and what situations do they apply to?

Why do we have to check the Conditions after choosing a type of hypothesis test.

What is a P-value and what does it tell us?

What is a confidence interval and what does it tell us?

How do hypothesis tests and confidence intervals relate to each other?

When can we draw a cause-and-effect conclusion and when can we not?

Populations and samples

- Random sample vs. sample of convenience
- Types of variables: qualitative/categorical vs. quantitative/numeric.
- Types of studies: observational vs. controlled experiments
    - When can we conclude that an association between two variables is actually a cause-and-effect relationship?
- Confounding
- Bias
- Parameters (describe the population) vs. Statistics (estimates of a parameter given by a sample)

Descriptive statistics: Numerical Summaries and Graphical Displays

- Analyze the shape of a histogram in terms of center, spread, and shape
- Use the mean and median to determine whether data is skewed left, skewed right, or symmetric.
- Know what the word "resistant" means (a value that is not very affected by outliers).
    - Know the mean is not resistant and that the median is resistant.
    - Know that standard deviation is not resistant
- Find the standard deviation of a data set by hand, using a table. Know what the term "Sum of the Squared Error" (SSE) means and find this value.
- Find the z-value and interpret it in terms of how many standard deviations a data value is from the mean.

- Know the meaning of "percentile": *k* percent of the data is BELOW the *k*th percentile.
- Know what Quartiles are.
- Know what the Five Number Summary is: Min, Q1, Median, Q3, Max and how it relates to a Boxplot.
- Use a Boxplot to compare sample data sets, in terms of the following:
    - Symmetry, skewness of a data set
    - Spread, variability (using IQR = length of box)
    - outliers (the * symbol is used by StatCrunch to indicate suspected outliers)

Linear Models and Linear Regression

- Interpret a Linear Regression printout:
- Use the equation of the Regression Line to find values such as
    - Slope: Interpret as a rate of change
    - A "predicted" value; determine its meaning in the context of the problem.

- Know the characteristics of the Correlation Coefficient (r-value), and relate it to a scatterplot.
    - $-1 \le r \le 1$ The r-value is between -1 and 1.
    - Perfect positive correlation (upward trend) has an r-value of 1
    - Perfect negative correlation (downward trend) has an r-value of -1
    - A weak linear association has an r-value close to 0.
    - The r-value is **not resistant**; in other words, outliers can seriously affect its value and thus lead to incorrect conclusions.

- **Know that "Correlation is not Causation**"; i.e., just because there is a strong correlation between two variables does NOT mean that one caused the other or even that the variables are related in any way. ("Spurious" correlation.)

- **Know that extrapolation can lead to incorrect conclusions and must be done with caution:** Extrapolation means making predictions beyond the data values.

Probability and Probability Distributions

- Know basic definition of probability and find probability of simple events and compound events (and, or)
- Find probability of repeated events (e.g. flipping a coin twice, three times, etc.)
    - Know the difference between sampling WITH replacement and WITHOUT replacement. Find probabilities of repeated events in both cases.
- Set up a probability distribution table.
- Identify Mutually Exclusive Events
- **Identify Independent or Dependent Events** For Dependent Events, one event happening changes the probability for the other event to happen. There is an association between the events, they're related in some way. If there's no association, the events are Independent.

Chi-Square Tests

- Be able to set up a **Chi-Square Hypothesis Test for Association** and form a conclusion (Interpret) based on StatCrunch output (will be given) for the compute step.

Sampling Distributions:

- Understand what a Sampling Distribution is (distribution of a statistic like p-hat or x-bar) found by drawing repeated samples of the same size from a population.
- Know what "Sampling Variability" is.
- Know what the Central Limit Theorem is and how it relates to Sampling Distributions.

Confidence Intervals

- Know what all of these terms mean and how they relate to a confidence interval: point estimate, margin of error, critical value (t* or z*)
- Use a Table to find a critical t-value (t*) value, based on a given level of confidence, and a given sample size
- Find and interpret a confidence interval for a population mean or a population proportion.
- **Relate a CI to a hypothesis test!**
- Interpret a confidence interval for a comparison of two populations. Know what it means when zero is "captured" by the confidence interval and explain what it means in terms of significant difference in between population proportions or population means.

Hypothesis Testing

- **Determine which type of test** would be appropriate for a given problem and set up the hypotheses. Questions to ask yourself when choosing the test:

  Are you trying to find out something about a mean (average) or about a proportion (some percent of a group that has some characteristic)?

  Use μ for mean tests, use p for proportion tests.

  Is there only 1 sample in this problem or are there 2 samples?

  Choose a 1-sample or 2-sample test accordingly.

  If you're comparing two means from two populations (two samples), are the individuals in the sample

  Dependent? (Such as Before-and-After, or Twin studies): Use the Paired t-Test

  Independent? (the most common situation): Use the Two Sample t-Test
- Know whether a test is right-tailed, left-tailed or two-tailed and how that relates to the area in a distribution graph and the P-value.

- **Be able to interpret a StatCrunch printout for any of the hypothesis tests and Confidence Intervals we've studied.**

One Sample t-Test:

Perform an entire hypothesis test by hand for a **1 mean**

- Use the formula to construct the Test Statistic
- Determine the outcome of the test and write a conclusion. Be able to write the conclusion in Laymen's Terms. Your conclusion should start with "Reject H0" or "Do not reject H0" then be followed up with "Laymen's Terms", i.e., stating what the result means in common language, using the words "significant" or "not significant".
- Illustrate all information on a distribution graph with two axes.
- (Note: You won't have to find the p-value by hand.)

Interpret a StatCrunch printout for a Hypothesis Test, and give a conclusion based on the P-value.

ANOVA (Signal and Noise)

- Know the central idea of ANOVA (finding the ratio of variance BETWEEN groups and variance WITHIN groups).
- Know the hypotheses for ANOVA.
- Know the conditions for using the test and how to check them.
  - Samples are random and independent
  - Underlying populations are normal (check boxplots for outliers, strong skewing)
  - Variances must be equal (check that the largest standard deviation is no more than double the smallest
  - standard deviation or the IQR of boxplots of samples are roughly equal.
- Identify the variables in an ANOVA test: Treatment (aka "Factor"), Treatment Groups (aka "Factor Levels"), and Response
- Know what each of the values in an ANOVA table represent. Particularly know where the variances between and within groups are located.
- Know how to compute the F-value and interpret the p-value from a test printout