STAT 250 STUDY GUIDE

Chapter 1: Collecting Data in Reasonable Ways

- Distinguish between population and sample
    - Explain the difference between a census and a sample
    - Explain the difference between a statistic and a population characteristic
- Distinguish between an observational study and an experiment
    - Understand that the conclusions that can be drawn from a statistical study depend on the way in which the data are collected; causal relationships need a well-designed experiment. The group studied and recruitment method affect to whom the results of the study generalize.
    - Understand the difference between random selection and random assignment
    - Evaluate whether the conclusions drawn from the study are appropriate, given the description of the study
- Describe procedures for drawing a simple random samples and understand what characterizes a simple random sample
- Evaluate the design of an observational study
    - SRS, Stratification, Cluster sampling, systematic sampling, convenience , voluntary response
- Explain why random selection is an important component of a sampling plan.
- Explain why volunteer response samples and convenience samples are unlikely to produce reliable information about a population.
- Evaluate sources of bias in a survey; Selection bias, Measurement (Response bias), nonresponse
- Evaluate the design of a simple comparative experiment
    - Completely Randomized, Block, Matched Pairs
- Understand the limitations of using volunteers as subjects in an experiment
- Explain the purpose of a control group and/or a placebo group in an experiment
- Explain the purpose of blinding in an experiment.
- Explain why random assignment is important when collecting data in an experiment.
- Describe procedures for randomly assigning experimental units to treatment groups

Chapter 2: Graphical Methods for Describing Data Distributions

- Identify the type of data – univariate, bivariate, multivariate

- Identify the type of variable – Categorical (Qualitative), Numeric(Quantitative) – Discrete or Continuous
- Choose an appropriate type of display for a given data/variable type (we had a chart)
- Define distribution, frequency, and relative frequency
- Know how to construct and interpret bar charts, comparative bar charts, and pie charts
- Know how to construct and interpret dot plots, stem and leaf displays, comparative stem and leaf displays, and split stem and leaf displays (center, spread, shape, gaps, outliers)
- Know how to construct and interpret histograms and comparative histograms
- Describe the shape of a distribution from a graphical display (symmetric, skewed left, skewed right, modes). Give examples of variables you would expect to have each shape.
- Construct and interpret a scatterplot (more in Ch 4) and time plot.
- Identify when poor practice produces misleading plots.

Chapter 3: Numerical Methods for Describing Data Distributions

- Choose appropriate measures of center and spread for a distribution based on the shape
- Compute mean and median of a data set
- Understand how outliers affect mean and median. Know where mean and median fall relative to one another in symmetric and skewed distributions.
- Compute the range and interquartile range of a data set
- Define and interpret sample standard deviation
- Determine whether an observation is an outlier using the fences method
- Know which measures are resistant/robust and non-resistant to outliers
- Know how to construct and interpret a boxplot
- Use a comparative box plot to compare/contrast the distribution of a variable for different groups
- Know the definition of a percentile

Chapter 4: Describing Bivariate Numerical Data

- Use a scatterplot to assess strength, direction, and form of a relationship
- Define/interpret the correlation coefficient (r)
- Know the facts about correlation listed in our notes
- Determine the strength of an association (weak/moderate/strong) based on the correlation and the scale in our book

- Be able to arrange scatterplots (on the same scale) by order of correlation. For practice see http://www.istics.net/Correlations/
- Understand how feature such as outliers and grouping can strengthen or weaken a linear association
- Identify potential reasons for observed association (causation, confounding, lurking varaiables)
- Know and interpret the components of a regression line
- Use a regression line to make predictions
- Compute and interpret residuals given observed data points and the equation of a regression line
- Explain why the regression line we use is called the Least Squares regression line
- Define and interpret the coefficient of determination, r^2. Compute r from r^2 and vice versa.
- Define and interpret the standard deviation about the regression line, s_e
- Locate regression values on Excel output (slope, intercept, r, r^2, s_e)
- Use a residual plot to evaluate whether a linear regression model is appropriate for the relationship between two variables
- Know what extrapolation is and why it should be avoided.

Chapter 6 Random Variables and Probability Distributions

- Identify if a random variable is discrete or continuous
- Express the probability distribution of a discrete random variable in a table
- Verify that a discrete distribution satisfies the properties:
    - All probabilities between 0 and 1
    - Probability of all possible outcomes sums to 1
- Compute the mean (expected value) and standard deviation of a discrete random variable using a probability distribution table
- Know that a probability distribution for a continuous random variable is expressed with a density curve
    - The curve always falls at or above zero
    - The probability a random variable falls in a given interval is equal to the area under the curve on that interval
- Compute probabilities of uniform random variables using a rectangular density "curve" (like the trolley waiting time example)
- Find areas under a given normal distribution (left tail, interval, or right tail)
- Find quantiles of any normal distribution (a value in the distribution given an area – inverse normal problems)
- Know how to turn values from any normal distribution into z-scores from the standard normal

- Check for normality in a data set – histogram, empirical rule, interpret a normal probability plot – you are looking for the dots to fall on a 45 degree line if the data set matches up to what you would expect a data set of that size to look like coming from a normal distribution.
- Know the properties of a binomial experiment and identify whether a series of trials satisfies the criteria for a binomial experiment.
- Know the formula for the P(X=x) in a binomial distribution and be able to describe the components of it
- Know what calculator commands to use to find binomial probabilities and cumulative probabilities
- Compute the mean and standard deviation of a binomial distribution
- Know under which conditions the normal distribution can be used to approximate the binomial distribution
- Use the normal distribution to approximate the binomial distribution when appropriate (continuity correction not required) - find approximate probabilities using the normal curve

Chapter 8 – Sampling Variability and Sampling Distributions

- Describe what a sampling variability is and what a sampling distribution is
    - Understand that the value of a sample statistic varies from sample to sample
    - Understand that a sampling distribution describes the sample-to-sample variability of a statistic
    - Understand how the standard deviation of the sampling distribution of a statistic is related to sample size
- Distinguish between a sample proportion and a population proportion and use correct notation for each
- Know the sampling distribution for a single proportion $\hat{p}$
    - Center
    - Spread
    - Shape – when is it approximately normal?
- Use the sampling distribution of $\hat{p}$ to compute probabilities involving $\hat{p}$

Chapter 9 – Estimating a Population Proportion

- Properties of a good estimator
  - Unbiased
  - Precise
- Know what margin of error is (interpretation), how it is computed for a single proportion, and how it is affected by sample size
- Define confidence interval and confidence level
- Know the general formula for a confidence interval
- Know and use the formula for a confidence interval for a single proportion
  - Under what conditions can you use the large sample confidence interval for a population proportion?
- Know how to get the critical value z* for any confidence level
- Be firm in your interpretation of a confidence interval
  - Gives a range of plausible values for the population characteristic
  - The confidence level expresses our confidence in the method. The method will produce an interval that captures the true characteristic 95% of the time and misses it 5% of the time.
  - It is NOT a probability that the characteristic is in the interval. The characteristic is fixed. The sample we chose and interval we computed is what was random
- Know how varying confidence level and sample size affect the width of the interval.
- Compute the sample size needed to achieve a desired margin of error for a population proportion p at a desired level of confidence.

Chapter 10 – Asking and Answering Questions about a Population Proportion

- Translate a research question into hypotheses – identify which population characteristic is of interest and determine the appropriate null and alternative hypotheses
- Interpret the results of a hypothesis test
  - Rejecting the null hypothesis implies strong support for the alternative hypothesis
  - Failing to reject the null hypothesis does NOT imply support for the null. Just means the observed data would be unsurprising if the null were true
- Know when it is appropriate to carry out the large-sample test for a population proportion
- Use all steps to carry out a large-sample test for a population proportion
  - Hypotheses – use appropriate notation

    - Check Criteria – np_0 and $n(1-p_0)$ both at least 10 and appropriate data collection
    - Compute – test statistic, p-value (for one and two-sided tests)
    - Communicate – compare P-value to alpha, state conclusion (reject/fail to reject), give conclusion in context of problem
- Interpret a P-value
- Know the types of errors that may occur in hypothesis testing
    - Be able to describe Type I and Type II errors in the context of the problem
    - Know how significance level of the test affects the probabilities of Type I and Type II errors and power
    - Know how sample size affects the probabilities of Type I and Type II errors and power
- Understand the reasoning used to reach a decision in hypothesis testing – Do the data "agree" with the null?
- Understand the difference between practical significance and statistical significance

Chapter 11 – Asking and Answering Questions about a Difference in Population Proportions

- Know the general properties for the sampling distribution of a difference in two population proportions
    - Center/Mean
    - Spread/Standard Deviation
    - Conditions under which the shape is approximately normal
- Know how and when two implement a large-sample confidence interval for a difference in population proportions.
    - Conditions (data collection and sample size)
    - Formula for confidence interval – be sure you know SE($\widehat{p_1} - \widehat{p_2}$)
    - Interpretation of interval and confidence level
- Know how and when to implement a large-sample hypothesis test for the difference in proportions
    - H – state hypotheses in terms of p_1 and p_2
    - C – check that conditions are met – use numbers of successes and failures in samples to check sample size requirement
    - C – compute test statistic (here we use a combined estimate of p -p_c in the standard error) and p-value
    - C – communicate conclusions: compare P-value to alpha, state conclusion (reject/fail to reject), give conclusion in context of problem

Chapter 12: Asking and Answering Questions about a Population Mean

- Determine the mean and standard deviation of the sampling distribution of $\bar{x}$; understand how the value of the standard deviation is related to sample size.
- Know when the sampling distribution of $\bar{x}$ is approximately normal
- Compute the margin of error when the sample mean $\bar{x}$ is used to estimate μ.
- Compute and interpret a confidence interval for μ. How do sample size, confidence level, and standard deviation affect the width of the CI?
- Know how t distributions relate to standard normal distributions
- Know how critical values from the t-distribution are found. Compute t* values with calculator or Excel.
- Sample size calculations for a desired margin of error to estimate a population mean. What do you use for sigma in the formula?
- Translate a research question or claim about a population mean into null and alternative hypotheses
- Use all steps to conduct a t test of hypotheses about a population mean (hypotheses, conditions, test statistic, p-value, conclusion in context)
- What is the distribution of the test statistic when the null is true?
- Compute p-values using tcdf on calculator or t.dist or t.dist.rt in Excel. Interpretation of p-values and the significance level of the test
- Definitions and conceptual understanding of Type I and Type II errors and power

Chapter 13: Asking and Answering Questions about the Difference Between Two Population Means

- Recognize whether two samples are independent or paired
- Translate a research question or claim about a difference in means or mean difference into null and alternative hypotheses using correct notation.
- Know the conditions for appropriate use of the two-sample t confidence interval and the two-sample t test.
- Carry out a two-sample t test for a difference in population means using unpooled methods. For unpooled, you would not be asked to compute d.f. by hand (use the conservative smaller of two samples sizes minus 1). The pooled method will not be on your exam.
- Compute and interpret a two-sample t confidence interval for a difference in population means (exam will only have unpooled)
- Know the conditions for appropriate use of the paired-samples t confidence interval and the paired-samples t test.
- Carry out a paired samples t test for a mean difference
- Compute and interpret a paired-samples t confidence interval for a population mean difference.

Chapter 15: Learning from Categorical Data

- Determine which Chi-Square test is appropriate for a given situation (Goodness of fit, Homogeneity/Independence)
- Determine appropriate null and alternative hypotheses for Chi-Square tests
- Know the conditions necessary for the Chi-square tests
- Compute expected cell counts
- Compute contribution of a single cell to the chi-square statistic
- Use software to compute a chi-square statistic for a list or table of observed cell counts (not for exam)
- Know what degrees of freedom to use in each Chi-square test
- Use calculator or Excel to compute a P-value for a given a Chi-Square test statistic
- Interpret the results of a Chi-Square test in context.