STAT 250 Midterm 1 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
* 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, r2. Compute r from r2 and vice versa.
* Define and interpret the standard deviation about the regression line, se
* Locate regression values on Excel output (slope, intercept, r, r2, se)
* 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) 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
* Use your calculator or Excel 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