Exam 1 Study Guide

Chapter 1:

- What is statistics?
- What is data?
  - What is a dataset?
  - What are elements?
  - What are variables?
  - What are observations?
- Measurement Scales
  - Nominal, Ordinal, Interval, Ratio
- Types of variables
  - Categorical (numeric or nonnumeric), Quantitative (numeric)
- Types of data
  - Cross-sectional
  - Times-series
  - Panel (cross-sectional time-series)

Chapter 2:

- What is a population? What is a sample?
- What is statistical inferencing?
- What is a parameter?
- What is a sample statistics?
- What is an estimator? What is a point estimate?
- How do we describe a data series?
- Summarizing categorical data
  - Tabular
    - Frequency, relative frequency, and percent frequency tables
  - Graphical
    - Bar chart
      - Drawn based on frequency tables
    - Pie chart
      - Drawn based on percent frequency table
- Summarizing quantitative data
  - Tabular
    - Frequency, relative frequency, and percent frequency tables
    - Cumulative frequency, cumulative relative frequency, and cumulative percent frequency tables
  - Graphical
    - Histograms
      - Know how to draw a histogram
      - Skewness and shape of distribution

- Summarizing the relationship between two variables
  - Tabular
    - Cross-tabulation
      - Row relative/percent frequency
      - Column relative/percent frequency
      - Total relative/percent frequency
  - Graphical
    - Scatter plots
      - How a scatter plot is drawn?
      - What does a scatter plot tell us about the relationship?
    - trend-lines
      - Slope of the trend-line tells us about the direction of the relationship
      - Fit of the trend-line to the data – How close the line is to the dots on a scatter plot tells us about the strength of the relationship

Chapter 3:

- Measures of location
  - Mean
    - Arithmetic
    - Weighted
  - Median – the middle of the distribution
    - If the number of observations is odd: pick the middle one
    - If the number of observations is even: take the average of the two middle ones
  - Mode (Most frequent value)
    - Less useful in flat distribution (where all values are equally frequent)
  - Percentile
    - Quartiles
      - 25^{th} percentile = 1^{st} quartile
      - 50^{th} percentile = 2^{nd} quartile
      - 75^{th} percentile = 3^{rd} quartile
    - Quantiles
      - 20^{th} percentile = 1^{st} quintile
      - 40^{th} percentile = 2^{nd} quintile
      - 60^{th} percentile = 3^{rd} quintile
      - 80^{th} percentile = 4^{th} quintile

- Measures of variability
  - Range – difference between the maximum and minimum values
  - Interquartile range – the range for the middle 50% of the observations
  - Variance
  - Standard deviation (square root of variance)
    - In some sense a measure of the average distance of observations from the mean of the distribution

- Measures of distribution
    - Histograms and skewness
        - Symmetric distribution (skewness = 0)
        - Skewed to the right (skewness > 0)
        - Skewed to the left (skewness < 0)
- Measures of association between two variables
    - Covariance
        - A static that determine the direction (positive or negative) of the relationship between two variables x and y
        - Covariance tells us nothing about the strength of the relationship between the two variables of x and y
    - Correlation coefficient
        - A statistic between -1 and +1, that determines the direction and the strength of the linear association between two variables x and y
        - Correlation coefficient is different from slope of the trend-line
    - Correlation coefficient is about the linear relationship between two variables
- Five number summary
    - Box-plot