Chapter 1. Notes

Main concepts:

Statistics: The science of planning studies and experiments, obtaining data, and organizing, summarizing, presenting, analyzing, and interpreting those data and then drawing conclusions based on them.

Population: The complete collection of all measurements or data that are being considered. Typically, a population is the complete collection of data that we would like to make inferences about.

Census: The collection of data from every member of a population.
Sample: A subcollection of members selected from a population

Parameter: a numerical measurement describing some characteristic of a population

Statistic: a numerical measurement describing some characteristic of a sample

Quantitative (or numerical) **data**: consists of numbers representing counts or measurements.

Categorical (or qualitative or attribute) **data**: consists of names or labels (not numbers that represent counts or measurements).

Quantitative data can be further described by distinguishing between discrete and continuous types.

Discrete data: result when the data values are quantitative and the number of values is finite, or "countable."

Continuous (numerical) **data**: result from infinitely many possible quantitative values, where the collection of values is not countable

Levels of Measurement: Another way of classifying data is to use four levels of measurement: nominal, ordinal, interval, and ratio.

Nominal level of measurement: characterized by data that consist of names, labels, or categories only, and the data cannot be arranged in some order (such as low to high).

Example: Survey responses of yes, no, and undecided

Ordinal level of measurement: involves data that can be arranged in some order, but differences (obtained by subtraction) between data values either cannot be determined or are meaningless.

Example: Course grades A, B, C, D, or F

Interval level of measurement: involves data that can be arranged in order, and the differences between data values can be found and are meaningful. However, there is no natural zero starting point at which none of the quantity is present.

Example: Years 1000, 2000, 1776, and 1492

Ratio level of measurement: data can be arranged in order, differences can be found and are meaningful, and there is a natural zero starting point (where zero indicates that none of the quantity is present). Differences and ratios are both meaningful.

Example: Class times of 50 minutes and 100 minutes

Summary - Levels of Measurement:

Nominal - categories only.
Ordinal - categories with some order.
Interval - differences but no natural zero point.
Ratio - differences and a natural zero point.

Basics of Collecting Data

Statistical methods are driven by the data that we collect. We typically obtain data from two distinct sources: observational studies and experiments.

Experiment: apply some treatment and then proceed to observe its effects on the individuals. (The individuals in experiments are called experimental units, and they are often called subjects when they are people.)

Observational study: observing and measuring specific characteristics without attempting to modify the individuals being studied

Simple Random Sample: A sample of n subjects is selected in such a way that every possible sample of the same size n has the same chance of being chosen.

Sampling Methods:

1. Systematic Sampling: Select some starting point and then select every kth element in the population.
2. Convenience Sampling: Use data that are very easy to get.
3. Stratified Sampling: Subdivide the population into at least two different subgroups (or strata) so that the subjects within the same subgroup share the same characteristics. Then draw a sample from each subgroup (or stratum).
4. Cluster Sampling: Divide the population area into sections (or clusters), then randomly select some of those clusters, and choose all the members from those selected clusters.

Types of Observational Studies:

Cross-sectional study: Data are observed, measured, and collected at one point in time, not over a period of time.
Retrospective (or case control) study. Data are collected from a past time period by going back in time (through examination of records, interviews, and so on).
Prospective (or longitudinal or cohort) study: Data are collected in the future from groups sharing common factors (called **cohorts**).