

MSIT / STAT 3000 Gilbert/Holmes/Werner Test # 1

Tuesday

August 13th

Review Syllabus / Clicker Registration

Thursday

August 15th

Chapter 1 – Statistics and Variation

Statistics is a way of reasoning designed to help us understand the world.

The process of statistics consists of:

1. Formulating a research question which drives statistics to find answers
2. Collecting data using sound methods
3. Describing data and analyzing data
4. Drawing conclusions or generalizations from data

Variation – a common concept that drives statistics and our ability to answer questions and draw conclusions. Variation is all around us. Diversity around us drives interest for statisticians.

Descriptive Statistics –uses numerical and graphical methods to look for patterns, to summarize, and to present data quantitatively in a convenient form. It is the way we investigate available information.

Inferential Statistics – uses sample data to make predictions and estimates about a larger set of data. We make broader generalizations and measure their reliability.

Chapter 2

2.1 What are data?

- Data values or observations are information collected regarding some subject.
- Data can be numbers or names or anything that tells us “Who” and “What.”
 - The “who” are not only people. It is what the data is about or who the data is from.
 - The “what” is the variable that has been measured.
- Data is useless without context.
- Data are often organized into a data table.
 - The “who” is usually the first column.
 - The “what” is the remaining columns.
 - The rows of a data table correspond to individual cases about whom we record some characteristics.
- These characteristics may be collected on or about:
 - Respondents – individuals who answer a survey
 - Subjects or participants - people in an experiment
 - Experimental units – animals, plants, websites, or other inanimate objects
- Variables – the characteristics recorded about each individual or case
 - Refer to the “what”

2.3 Data Sources – Where, How, and When

- When data are collected can be important. Data that is a decade old is less reliable than data collected last year.
- Where data is collected can also be important. Data collected about Mexico differs from data collected about the United States.
- How data are collected can make a difference between insight and nonsense.
 - Data collected through Internet surveys are almost always worthless because, typically, only those who extremely agree or disagree will respond.
 - Data collected by agencies or businesses online are usually extremely useful because they are monitored.
 - Good statistical method is foundational for collecting any data.
- Data can be found:
 - By performing an experiment and actively manipulating or changing variables
 - Taking a sample or survey
 - In information collected by public or private agencies
 - On internet sites

2.2 Variable Types

- Categorical variable –places an individual into one of several groups or categories
 - Statisticians may have to convert a category to numbers through proportions.
 - Arithmetic operations are NOT meaningful for data resulting from categorical variables.
 - Example: colors used in an advertising campaign (red, blue, green, ...)

- Quantitative variable – takes numerical values for which arithmetic operations such as adding and averaging make sense
 - These variables can be categorized to help with process
 - Numbers can represent another categorization. Example: Area codes (404, 678, 706) are categorical variables
 - Example: the number of female executives employed (0, 1, 2, ...)

Chapter 3

3.3 Populations and Parameters

- Population – the entire collection of individuals or instances about which information is sought
 - This represents the entire group under study, NOT the same as sample.
 - Example: all faculty members at the college
- Sample – a subset of a population, examined in hope of learning about the population
 - This is not a perfect representation of the population.
 - It resembles the population but does not mirror it.
 - Example: 10 randomly selected faculty members at the college
- Parameter – a quantity that describes an attribute of the population under study.
 - This represents the characteristics we are interested in.
 - It can be a mean or proportion or slope.
 - Example: the average age of all faculty members at the college
- Statistic – a value calculated from sampled data, particularly one that corresponds to, and thus estimates a population parameter.
 - Example: the average age of the 10 randomly selected faculty members at the college

- Individuals
 - Example: faculty members at the college
- Variable
 - Example: the age of each faculty member at the college
- Data are always an imperfect view of the parent population.
 - What I see (in the data) is not quite the way it (the population) really is.
 - Inference is what I think is likely to be happening back in the population. These inferences are based on sample data.
- Population aligns with the parameter.
- Sample aligns with the statistic.

Chapter 3

- Three ways to gather data:
 1. Take a census of the entire population
 - We take a census every 10 years because it is written in the Constitution of the United States

Tuesday

August 20th

2. Conduct an experiment
 - Impose treatment onto the subjects by manipulating the variable for some and then manipulating differently for others creating the observed effect
 - It is important to use sound statistical methods.

- The implication of the experiment is that the results would work on most people or the average number of people.
 - We use experiments to conclude causation.
3. Take a sample (often using a survey)
- Many times statisticians must use observational studies because it would be unethical to do an experiment. For example, we cannot conduct an experiment about the effect of parents smoking cigarettes around their children.
 - We conduct observational studies by finding information that exists or by gathering information as we go along.
 - The strongest conclusion one can make from an observational study is a relationship or an association.
- Researchers cannot take a census or question everyone in the world, but they make generalizations about the world everyday through statistical research and application.
 - Data and the population are not mirror images.
 - Variation exists when gathering all data.
 - The idea of random needs to be applied when gathering data in an unbiased and unflawed way.

3.1 Three Ideas of Sampling

- Ideas of Sampling
 - Examine a Part of the Whole
 - We use samples to hopefully make sound conclusions about the entire population. Sometimes these sampling methods are considered biased if they over or under-emphasize some characteristics of the population.

- The tool used to gather the data, not the actual data, is flawed. With biased sampling methods, the summaries drawn from a sample may differ from the actual population. Statisticians try to select samples at random.
- Randomize
 - Randomizing gives us the best chance at obtaining a representative sample that resembles the population we are trying to learn about.
- The Sample Size is What Matters
 - The size of the sample determines what we can conclude from the data regardless of the size of the population. Some samples need to be very large while others can be very small. For example, when predicting what political candidate a certain city might vote on, a small sample of 15 voters will produce a fairly accurate decision if the sample is actually chosen at random.
- Sampling frame – list of individuals from which the sample is drawn. We choose the sample from the frame by assigning a sequential number to each individual and then picking numbers at random.
 - Example: choose from names in a phone book or a class roster
- Sampling variability/Sampling error – sample-to-sample differences in the values for the variables we measure
 - Example: If Dr. Gilbert is trying to determine the average distance her students travel from home to get to UGA, each of her three classes will have a different average distance.