Key Terms from the Presentation

You should be able to explain each of these in your own words, and be able to recognize situations where each one applies:

descriptive statistics
inferential statistics
histograms
statistical significance
mode
mean
median
standard deviation

How do Descriptive Statistics Help Researchers Understand Data?

After the initial collection of data, the researcher must organize it. A first step is to list all the possible responses and how often each one occurred. This is a frequency graph, or frequency distribution. A frequency distribution is usually depicted in a graph such as a histogram, or a polygon. Here, for instance, is a histogram of weights on a football team. Each bar represents the number of people who fall within a certain weight range.

Measure of Central Tendency and Variability

A measure of central tendency is a single number that people use to represent an entire data set. Usually it's considered to be the "middle" of the data set. There are three measures of central tendency that are used to express the central point of a distribution of numbers:

- the mode: the most commonly occurring response(s)
- the median: the halfway point between the lowest and highest number
- the mean: the "average," or all of the numbers added up and divided by the total number of observations. The mean is the measure that most people think of when they hear the word "average."

Researchers also need a measure of variability, or how much variety there is in the data. This is especially useful to researchers when they want to make inferences about a whole population based on a sample. There are three measures of variability:
• The **range** shows the spread of scores in a distribution, or the difference between the highest and lowest number.

• The **standard deviation** is the average distance from the mean. The higher the standard deviation, the more variability, or variety, there is in the data.

• The **variance** is simply the square of the standard deviation. Sometimes in calculations it's easier to use the variance than the standard deviation.

**The Normal Distribution, z-Scores, Percentiles**

Frequency plots of distributions come in many shapes. Sometimes they're skewed, with a peak on one side and a long tail off to the other side. Sometimes a distribution has more than one peak. Such **multi-modal** distributions are sometimes actually collections of different distributions that were combined.

One of the most common shapes is a bell-shaped curve with the mean in the center. This is the normal distribution, and it's a common shape for distributions where there's an equal chance that someone can fall above or below the mean. Here's a normal distribution of test scores with a mean of 65 and a standard deviation of 10:

![Normal Distribution Graph](image)

Test scores for traits such as intelligence are often normally distributed. Two important properties of the normal distribution are:
• the mean, median, and mode are all the same
• 68% of all observations (e.g., test scores) fall within one standard deviation of the mean, 95% of all observations (e.g., test scores) fall within two standard deviations, and 99% within three standard deviations.
Inferential Statistics: Using a Sample to Learn About a Population

So if a researcher finds something out about a sample, then what? How can it be determined whether or not it says something about the whole population?

Inferential statistics is the process of using samples to draw conclusions about a population. Using inferential statistics, you can learn a lot about a population from relatively little information. For example, a research scientist can make a reasonable estimate about whether the results of an experiment prove a new discovery. Or a pollster can predict the outcome of a national election by sampling only .001% of the population. But the sample must be random and the study must be sound. An inference is only as good as the sample data it's based on.

How do Researchers Interpret Data?

There are two basic functions of inferential statistics:

- Provide estimations of population parameters. For instance, say you wanted to know whether most Americans are introverts or extroverts, as defined by a score on a particular personality test. With a properly drawn sample, you can determine how likely it is that the results for your sample are valid for the entire population.

- Hypothesis testing. Often when conducting a study, a researcher wants to support a hypothesis such as "television viewing is negatively correlated with scores on a test of scholastic aptitude." In designing the study, there will be a null hypothesis, which is that there's no correlation, and an alternative hypothesis, which is that there is a correlation. If a properly drawn sample shows a strong enough correlation, then you can reject the null hypothesis and say that your study supports the alternative hypothesis.
Hypothesis testing involves calculating the likelihood that you’d get your results due to random chance, as opposed to getting your results because they reflect a true correlation or some other real phenomenon in the population. If a properly drawn sample shows a characteristic such as a correlation, and the likelihood of getting your results due to random chance is low enough, then you can reject the null hypothesis.

This likelihood is called the p-value, or probability value. It's calculated based on factors including the size of your sample (the larger the sample, the lower the p-value) and the standard deviation (the lower the standard deviation, the lower the p-value). In other words, you can tell (infer) more from a large sample with little variability than from a small sample with a lot of variability. Of course, this all assumes that you've drawn the sample properly. The p-value is also affected by how large of a difference you observe. For instance, if a group of people taking an antidepressant drug is remarkably happier (as opposed to barely happier) than people who did not take the drug, then there's less of a likelihood that the difference is due to random chance.

The p-value relates to an important concept called **sampling error**. This error is a measure of how different your sample is from your population. Ideally you want the sampling error to be zero, but because a sample is drawn randomly there's always a chance that the sample will represent unusual (non-representative) cases instead of common ones. However, if you draw a large sample and you keep the sampling bias as low as possible, the possibility of a large sampling error becomes very small.

If the p-value (for results due only to random chance) is low enough, then the results are said to be **statistically significant**. The maximum p-value is set ahead of time – usually at .05 or 5%. So, if there's a 5% or less chance that you could get your results from chance alone, then you can consider your results to be statistically significant.

Sometimes the null hypothesis is that there's no significant difference between two treatments. Say, for instance, that one group received a new antidepressant drug and the other group received a placebo (a pill that does nothing). At the end of the trial period, you give each group a questionnaire to rate their mood. If:

- the group that received the drug (treatment group) has a mean score close enough to the group that received the placebo (control group) and,
- the sample is small enough and the standard deviation high enough that the p-value is higher than .05,

then the null cannot be rejected. In other words, the study supports the null hypothesis that that new drug has no statistically significant effect.

Note that you don't **prove** hypotheses with hypothesis testing. Rather, you accept (support) or reject (disprove) them.