Module A: The Foundations of Statistics

Slide Deck A1:

An Introduction to Statistics

The section in which we start our journey to learning about the world around us. Here, we will be introduced to some introductory terminology that will be used throughout the term.

Today's Objectives

By the end of this slidedeck, you should

- know the two main branches of statistics
- know the two types of statistical analyses
- know the purposes of confidence intervals and hypothesis tests
- know how to present scientific results in a major paper

Introduction to Statistics

The goal of the discipline called **statistics** is to provide information so that *informed* decisions can be made. The goal of this course is to enable you to filter through the statistics you encounter so that you can be better prepared for the decisions you make in your daily life.

So, what is statistics?

- **Statistics** is the science of gathering, describing, and analyzing data.
- **Statistics** are the actual numerical descriptions of sample data.

Context is important for determining which meaning is being used.

Branches of Statistics

There are (at least) two branches of statistics:

- **Descriptive Statistics**:
  - This branch focuses on displaying statistics in a meaningful manner. It focuses on gathering and sorting data so that one can meaningfully summarize and display it.
- **Inferential Statistics**:
  - This branch involves using statistics to *estimate* population parameters.

Note that inferential statistics depends on descriptive statistics in two ways. First, it requires descriptive statistics to perform the estimation. Second, it uses descriptive statistics to "tell the story of the data."

Descriptive Statistics

This branch focuses on displaying statistics in a meaningful manner. It focuses on gathering and sorting data so that one can meaningfully summarize and display it.

In this course, the second learning module is dedicated to descriptive statistics. In that module, you will learn not only how to calculate statistics and display data, you will learn which statistics and graphics can be used to tell the story of the data. Here, you will learn

- **univariate statistics** like the mean, median, mode, standard deviation, variance, interquartile range, coefficient of variation
- **bivariate statistics** like the correlation
- **univariate graphics** like the histogram, the box-and-whiskers plot (boxplot), pie chart, the bar chart
- **bivariate graphics** like the side-by-side boxplot, the side-by-side barplot, the stacked box chart, the mosaic plot, the scatter plot

Inferential Statistics

This branch involves using statistics to draw conclusions about *population* parameters.

In this course, we will cover inferential statistics in the fourth and fifth learning modules. The fourth introduces you to this topic, and the fifth takes it to the next level. Together, the two modules cover a variety of estimation procedures and tests. These include the following

- one-sample t-procedure, Wilcoxon test, Binomial test, chi-square test, non-parametric bootstrap
- two-sample t-test, Mann-Whitney test, proportions test, F-test
- chi-square goodness-of-fit test, analysis of variance (ANOVA), chi-square test of independence, correlation test
- linear regression

Goals of Inferential Statistics

There are (at least) two types of analyses (goals) for inferential statistics:

- **Exploratory Analysis**:
  - Exploratory analysis uses data to estimate parameters.
- **Confirmatory Analysis**:
  - Confirmatory analysis uses statistics to test claims about reality (hypotheses).

Note that a "proper" analysis will use both (if possible). The latter determines if the claim about the parameter is reasonable. If it is not, then the former will provide a set of reasonable values for the parameter.

Exploratory Analyses

Exploratory analysis uses data to **estimate parameters**. This is frequently done using **confidence intervals**. These confidence intervals are sets of values that are "reasonable" for the parameter.

Note:

- confidence intervals are dependent on the data
- they provide values for the parameter that are *reasonable*

Confirmatory Analysis

Confirmatory analysis uses statistics to **test claims about reality**. When performing confirmatory analysis, one must have something in need of confirming. This "thing" is called a hypothesis. Once the hypothesis exists, data are collected and used to determine if the hypothesis is reasonable, *given this data*.

Note:

- hypothesis testing requires the claim before the data are analyzed
- it uses the data to determine if the claim is reasonable

Presentation of Results

Sharing the results of your analyses is important; it allows others to learn from your analysis and to build upon your findings. Thus, you will need to be able to *properly report your findings*.

There is nothing special about the presentation; it follows the typical structure of any argumentative essay, albeit tuned to the proper use of data:

- **Introduction**: gives purpose and summarizes the report, ends with the claim
- **Literature Review**: provides logic behind why your claim (hypothesis) is reasonable
- **Data and Methods**: provides the logic of your data, how it was collected, its structure; the logic of your statistical method, why it is appropriate, whether its requirements are met
- **Results and Discussion**: tells what the data tell you about the claim, what this means regarding the bigger question
- **Conclusion**: summarizes your report, provides weaknesses in your analysis, ends with the results

Today's Objectives

Now that we have concluded this lecture, you should be able to

- know the two main branches of statistics
- know the two types of statistical analyses
- know the purposes of confidence intervals and hypothesis tests
- know how to present scientific results in a major paper

Supplemental Readings

The following may be of interest to you in terms of the topics covered today:

- Hawkes Learning: Section 1.1
- Intro to Modern Statistics: Chapter 1
- R for Starters: Nothing