Topic 1.1

Introducing Statistics: What Can We Learn from Data?

What Can We Learn from Data?

Unit 1 introduces the concept of data and describes how data can vary in the real world. In certain circumstances, this variation may suggest certain conclusions about data. However, not all variation is meaningful. The study of statistics will help us understand and make sense of uncertainty and variation.

In this unit, we will learn about categorical and quantitative variables, and how to represent them appropriately. We will also learn to describe and compare distributions of data that consist of a single variable (one-variable datasets) using tables, graphical representations, summary statistics, or a combination of these methods. These statistical methods will help us assess claims about a particular data point or about groups of data as in a sample. Later in the unit, we will be introduced to the normal distribution. The application of the normal distribution will be our first step in understanding how some distributions of sample data can be described using theoretical models for populations. This brief introduction to the normal distribution is a building block for later units, where we learn to use probabilistic modeling to make statistical inferences.

To prepare for the AP exam, we suggest focusing on developing the following skills:

- For categorical variables be able to:
  - Represent and describe variables using frequency tables, relative frequency tables, and bar charts.
- For quantitative variables be able to:
  - Represent and describe variables using histograms, stem-and-leaf plots, dotplots, and boxplots (including the five-number summary).
  - Describe the distribution in terms of shape, center, and variability (spread), as well as any unusual features such as outliers, gaps, clusters, or multiple peaks.
  - Recognize skewness (left and right) and how the mean relates to the median in skewed distributions.
  - Calculate the main measures of center, position, and spread.
  - Understand and apply the empirical rule for normal distributions.
  - Determine proportions and percentiles from a normal distribution (using *z*-scores).
  - Use percentiles and *z*-scores to compare the relative positions of points within a dataset.