Basic Terminology

Reading

Section 1.1

Notes

- The science of **statistics** deals with the collection, analysis, interpretation, and presentation of **data**.

- **Descriptive Statistics** deals with organizing and summarizing of the data

- **Inferential Statistics** deals with drawing conclusions from sample data about the larger population they represent.

- Key Terms:

  Population A collection of "things" we want to study

  Sample A *manageable* portion/subset of the population. We collect information only for the sample, and this gives us our data.

  Statistic A statistic is a number that represents a property of the sample data. For instance if we have a sample of students, it could be their average gpa.

  Parameter A number that is a property of the population. Typically unknown to us. But we use the statistic to try to estimate it.

  Variable A characteristic of each individual/case in the population. For each individual and each variable there is a corresponding well-defined and unique value.

  Data The actual values of the variable, one for each individual.

- **Variables** are arranged in two main types:

  Scalar Also called **numerical** or **quantitative** variables. Their values are numbers in a scale, in some specific unit of measurement. A defining characteristic of scalar variables is that it makes sense to form averages. Examples: GPA, height in feet, income in USD.

  Categorical Also called **qualitative**. They classify the individuals in groups. Examples: Gender, Grade. Categorical variables are further divided into **Nominal** and **Ordinal**, depending on whether the different categories have a natural order to them or not. Gender would thus be nominal, Grade would be ordinal.

- Some times it may be hard to determine if some data is scalar or ordinal. The main difference is the types of analyses we do and questions we ask. In general, if the distance between the numbers is important, that tends to make the values scalar. If on the other hand there are two few different possible values, that is a reason to treat the variable as ordinal.

- Activity: List a number of different variables that you could measure on students. Then assign a type to each variable.