Statistics, Individuals and Variables

Statistics is the collecting, organizing and interpreting of information (data). Individuals are the objects described by a set of data.

- Population is all individuals of interest.
  - Inferential Statistics: Assume, or infer, something about the population based on data.
- Sample is a subset of the population.
  - Descriptive Statistics: Describing your sample when interpreting your data.

Variables are characteristics of individuals.

When your data is provided as a list of information, you can identify the individuals of the story problem situation by organizing the data in a spreadsheet format.

In a spreadsheet (or table) format:
- rows are the individuals
- columns are the variables

Example: A spreadsheet for employees of a company.

<table>
<thead>
<tr>
<th>Name</th>
<th>SS#</th>
<th>Gender*</th>
<th>Title</th>
<th>Salary in $</th>
<th>Hours worked per week</th>
<th>Vaccinated for H1N1</th>
<th>Damage Reports Filed</th>
<th>State born in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrera, Ricardo</td>
<td>111-11-1111</td>
<td>0</td>
<td>Clerk</td>
<td>25000</td>
<td>40</td>
<td>Y</td>
<td>24</td>
<td>FL</td>
</tr>
<tr>
<td>Murphy, Patrick</td>
<td>999-99-9999</td>
<td>0</td>
<td>Manager</td>
<td>45000</td>
<td>50</td>
<td>N</td>
<td>13</td>
<td>OH</td>
</tr>
<tr>
<td>Zion, Mary</td>
<td>555-55-5555</td>
<td>1</td>
<td>Manager</td>
<td>50000</td>
<td>50</td>
<td>Y</td>
<td>20</td>
<td>CA</td>
</tr>
</tbody>
</table>

*For Gender, 0 = Male, 1 = Female

TYPES OF VARIABLES

In a spreadsheet, consider the entries of the variable to determine which type the variable is.

1. Quantitative: Data is described using numbers such that the values of the numbers are used in calculations. For example, calculating the average test score. They can also have units of measure attached to them, such as miles/hour, gallons, inches, seconds, degrees, etc.
   a. GRAPHS for organizing include Dotplot, Histogram, Stemplot, Boxplot, etc.
2. Categorical: Data described usually with words but can be numbers if the values are not taken into consideration, such as with SS#, Telephone Number, Zip Code, ISBN#, Driver’s License, etc.
   a. GRAPHS for organizing include Bar Graph, Pie Chart, etc.

We want to know the Distribution of a variable: what values does the variable have and how often do these values occur. For a quantitative variable

minX is the smallest number in a list of data
maxX is the largest number in the list of data
Spread is a descriptive phrase, “Spread is from minX to maxX”
Range is a number, Range = maxX – minX

“Frequency” is the “number of individuals”. The total number of individuals in a sample is denoted as n.
“Relative Frequency” is the “percent of individuals”.