Math 243 Lecture Notes

CHAPTER 1: INTRO TO DATA

Define data:
- A collection of numbers, characters, images, or other items that provide information about something
- Raw information

Define statistics:
Statistics (as a field of study) is a way of reasoning, along with the tools and methods designed to help us understand the world.
When we refer to statistics (as things) we're referring to particular calculations that have been made from data.

Example 1. Draw a picture representing the concepts of sample and population. Then list two specific examples.

Population: Students enrolled at PCC Fall 2016
Sample: Students enrolled in a MIT class Fall 2016
Sample: Students at the Sylvania campus
Sample: Students aged 30 and above
Sample: Students in this class
Some more vocabulary:
- Individuals who answer a survey are called respondents.
- People on whom we experiment are subjects.
- Experimental units are inanimate subjects such as animals, plants, websites, etc.
- Records are rows of a database; the more generic term is case.

The 6 W’s:  Who, What, Where, When, Why, How

Example 2. For each data description below, identify who and what are being investigated, as well as the population of interest.

(a) Ian Walker, a psychologist at the University of Bath, wondered whether drivers treat bicycle riders differently when they wear helmets. He rigged his bicycle with an electronic sensor that could measure how close each car was that passed him. He then rode on alternating days with and without a helmet. Out of 2,500 cars passing him, he found that when he wore his helmet, motorists passed 3.35 inches closer to him, on average, than when his head was bare.

Who: The 2,500 cars passing him
What: The distance of each car from his bicycle
Population: All cars (in Bath? In England? Everywhere?)

(b) Coffee stations in offices often just ask users to leave money in a tray to pay for their coffee, but many people cheat. Researchers at Newcastle University alternately taped two posters over the coffee station. During one week, it was a picture of flowers; during the other, it was a pair of eyes. They found that the average contribution was significantly higher when the eyes poster was up than when the flowers were there. Apparently, the mere feeling of being watched—even by eyes that were not real—was enough to encourage people to behave honestly.

Who: Customers at the Newcastle Univ. coffee vending
What: Honesty, measured by the amount paid (or not!)
Population: Humans ... or coffee drinkers?

(c) A look at 474 participants in the San Antonio Longitudinal Study of Aging found that participants who drank two or more diet sodas a day “experienced waist size increases six times greater than those of people who didn’t drink diet soda.”

Who: The 474 participants
What: Waist size and also amount of soda consumed
Population: All people (in the US?)
A **variable** is a characteristic recorded about an individual. A variable that describes a group or category is called a **categorical** or **qualitative variable**. A variable containing measured numerical values (with measurement units) is called a **quantitative variable**.

**Example 3.** Create a list of examples of qualitative variables and quantitative variables.

<table>
<thead>
<tr>
<th>Categorical (Qualitative) Variables</th>
<th>Quantitative Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>breed of a dog</td>
<td>age of dog</td>
</tr>
<tr>
<td>color of clothing</td>
<td>number of shirts in a closet</td>
</tr>
<tr>
<td>size of clothing (S, M, L, etc.)</td>
<td>size of shoe (6, 7, 8, ...)</td>
</tr>
<tr>
<td>temperature (cold, cool, warm, hot)</td>
<td>temperature (in °F, for example)</td>
</tr>
<tr>
<td>coins (quarters, nickels, etc.)</td>
<td>coin value ($.25, $.05, etc.)</td>
</tr>
<tr>
<td>exports</td>
<td>GDP</td>
</tr>
<tr>
<td>type of car</td>
<td>mpg</td>
</tr>
<tr>
<td>relatives in home</td>
<td># of family members in home</td>
</tr>
</tbody>
</table>
Example 4. Identify if the following are categorical or quantitative:

(a) Zip codes

Generally, this is just categorical

Originally, lower numbers were used on the east coast and got larger as you went west.

(b) Area codes

Generally, this is just categorical

Originally, large cities had the lowest area codes

An ordinal variable is one that reports order without natural units.

Example 5. An example of an ordinal variable is a variable measured with the Likert scale.

(a) Rate how much you like _____________ in the Google Spreadsheet (see your MyPCC email to access this). Record the frequencies of ratings 1, 2, 3, 4, and 5 below.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

(b) Explain how this rating could be considered either a quantitative variable OR a qualitative variable.

Categorical: (not tired at all, ... extremely tired)

Quantitative: 1, 2, 3, 4, 5

An average of 3 could be obtained from a class of all 3s, or 1/2 1s and 1/2 5s.