

Chapter 1 & 2 - Data

August 24, 2009

Statistics - a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data. (www.m-w.com)

Statistics is about **variation**. Data will vary since we cannot measure everything and what we do measure will not be measured perfectly. Data we look at and base decisions on provides an imperfect picture of the world.

Steps in doing Statistics right: (p. 4)

1. **Think** - Know where you are heading and why.

Context -

WHO was measured? (Student's or parent's incomes)

WHAT was measured? (Student's height or weight)

HOW was data collected? (Student's weight - was it self reported or were each weighed and recorded. We will find out later the problems with voluntary surveys.)

WHERE was the data collected? (Incomes in Murray or New York City)

WHEN was study performed? (Value of Stocks at end of 2007 or 2008)

WHY was the study performed?

2. **Show** - Calculating statistics (Numerical calculations on the data) and making displays of the data.

3. **Tell** - Explain your results, so someone else can understand.

Data - Systematically recorded information, which can be numbers or labels, together with its context.

Individual (cases) - the objects data describes. Individuals who answer a survey are called **Respondents**. Individuals involved in an experiment are called **Subjects** or **Participants**. Animals, plants and other inanimate objects are often called **Experimental Units**.

Often the cases are a **sample** selected from a larger **population**.

Variables - Characteristics recorded about each individual. There are two types.

Categorical - Variables that name categories. Can be words or numbers.

Quantitative - Variables with numerical values for which arithmetic operations make sense. Always have units. [This pig weighs 450. Does that mean pounds, grams, kilograms, etc?]

Units - A quantity or amount adopted as a standard of measurement, such as dollars, hours, feet, or grams. (Base Units on p. 10)

Some variables can be either categorical or quantitative depending on what we want to learn from the variable.

Example: Final grades are numbers like 72, 75, 83, 92, 96, 99. An average can be performed on these grades and possibly used to compare with another class to see which class did better. (Categorical or Quantitative?) We use the numbers to tell how many students received A's, B's, C's, D's, or F's. (Categorical or Quantitative?)

Identifier Variables are categorical variables with exactly one individual in each category. These are used to identify the individuals. Examples: Student #, Social Security #, Account #, etc.