

Remember

to Record

Essentially, there are two ways to obtain data:  
**observation** and **controlled experiments**.

The data collection method is related to the nature of the problem to be solved and the ethical and practical constraints of collecting data in some environment.

### **Caution**

The conclusions suggested by statistics can be no stronger than the quality of the measurements which produced the statistical evidence. Fuzzy or confounded measurements must produce fragile conclusions.

## Definition

A **response variable** measures the outcome of interest in a study.

## Definition

An **explanatory variable** causes or explains changes in a response variable.

Suppose a new species of tomato has been genetically engineered to increase yields. How can we measure whether the new species produces higher yields?

## Example 2.1.2

Does an SAT preparation course improve performance on the SAT?

**Solution**

## Definition

A **placebo** is a fake treatment that has the potential to cause a response.

## Definition

**Double-blind studies** are used to counteract the placebo effect. In a double-blind study the subjects are not told whether they are members of the treatment group or the control group. The evaluators (the persons that measure the response variable) are not told whether their subjects are members of the control or experimental group.

## Definition

Data in which the observations are restricted to a set of values (such as 1, 2, 3, 4) that possess gaps is called **discrete**.

## Definition

Data that can take on any value within some interval is called **continuous**.

## Definition

**Qualitative** data is measured on a nominal or ordinal scale.

## Definition

Data that represents whether a variable possesses some characteristic is called **nominal**.

## Definition

**Ordinal data** represents categories that have some associated order.



## Definition

**Quantitative** data is measured on an interval or ratio scale.

## Definition

If the data can be ordered and the arithmetic difference is meaningful, the data is **interval**.

## Definition

**Ratio data** is similar to interval data, except that it has a meaningful zero value.

Specify whether the following variables are well-defined or not. Justify your answer.

**a.** Height

**c.** Hot

**e.** Beauty

**b.** Weight

**d.** Temperature

The Nurse's Health Study conducted on 87,245 women at Boston's Brigham and Women's Hospital revealed that women who eat a cup of beta carotene-rich food a day have 40 percent fewer strokes and 22 percent fewer heart attacks than those who consume a quarter of a cupful per day.<sup>3</sup>

- a.** Which phase of the Scientific Method best describes this study?
- b.** Is this an observational study or a controlled experiment?
- c.** What is the response variable?
- d.** What is the explanatory variable?
- e.** Which group is the treatment group?
- f.** Which group is the control group?

Determine the level of measurement for each of the following variables.

- a.** Golf score in relation to par
- b.** SAT score
- c.** Rating from 1 to 5 of quality of service in a restaurant
- d.** Make and model of a vehicle
- e.** The number of students with a business major

Determine the level of measurement (nominal, ordinal, interval, or ratio) for each of the following variables.

- a.** The temperature (in degrees Fahrenheit) of patients with pneumonia.
- b.** The age at which the average male marries.
- c.** Client satisfaction survey responses: Poor, Average, Good, and Excellent.
- d.** The region of the U.S. in which an individual lives: North, South, East, or West.
- e.** The number of people with a Type A personality.

The results of a study investigating the nutritional status of mid-nineteenth century Americans were reported in “The Height and Weight of West Point Cadets: Dietary Changes in Antebellum America,” in the *Journal of Economic History*. The data is based upon physical examination lists for West Point applicants from 1843 to 1894. Some of the information obtained from each cadet were his height, weight, the state from which the cadet was appointed, the occupation of the father, the income of the parents, and the type of home residence (city, town, or rural) of the cadet.

- a.** List the different variables measured on the cadets.
- b.** Which variables are quantitative and which are qualitative?
- c.** Give the levels of measurement for these variables.