

1.2: Basic Definitions

Data : The numbers we collect. (Note the word data is plural. Datum is singular.) Data may be grouped into sets, hence *data set*.

Variable : A mathematical term used to denote something that can take on a range of *values*. There are important two types of variables :

- i. **Independent variable (IV)** : You set the value, a.k.a. *explanatory variable*.
- ii. **Dependent variable (DV)** : Value set (generally caused) by the independent variable, a.k.a. *outcome variable*. See Figure 1.4.

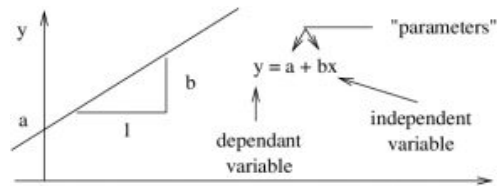


Figure 1.4: In the equation of a line, y is the dependent variable, x is the independent variable.

Random Variable : A dependent variable with random noise added. Value given by a *stochastic process*. We will only refer to random variables when discussing the theoretical relationship between probability distributions. Random variables, which we will denote with capital letters like X , are defined by their probability distribution. A stochastic process produces values that form a probability distribution if you allow the process that generates their values run for long enough.

Note : Data are frequently called “variables” in anticipation of how they will be used. The software program SPSS uses that convention.

1.2.1 Types of Data (important!)

Qualitative variable : described by a word, e.g. gender with “values” male or female. Qualitative variables are converted to *discrete quantitative variables* before analysis (e.g. male = 1, female = 2). In SPSS, you need to assign discrete numbers to qualitative variables in the “Values” column in the “Variable View” screen.

Quantitative variable : two types :

- i. **Discrete variable** : integer valued. In mathematical symbols $x \in \mathbb{Z}$ (read “the variable x belongs to [the set symbol \in means “belongs to”] the set of integers \mathbb{Z} ”). e.g. -2, -1, 0, 1, 2, 3, etc.
- ii. **Continuous variable** : real valued (essentially any number). In mathematical symbols $x \in \mathbb{R}$ (read “the variable x belongs to the set of real numbers \mathbb{R} ”). Geometrically, \mathbb{R} is the number line.

Note : Continuous variables can be converted to discrete variables by *grouping* :

heights ≤ 5 ft = “short” (group value = 1)

heights > 5 ft = “tall” (group value = 2)

Groups are also known as *classes*. We will be spending time defining classes in Chapter 2. Identifying what type of variable your data is will be the best way for you to decide what statistical test you need after you have learned and understood a number of different tests.

1.2.2 Measurement Scales (avoid this!)

Some texts, and the SPSS helper program (although I have never tried it), attempt to classify data into “scales” that try to go somewhat beyond the integers and real numbers. I don’t think such classification is particularly useful and recommend that you avoid such classification. Nevertheless, it exists, so we will take a very quick look at such scales. (There is no agreement about their definitions from source to source.)

One textbook that I used for a Univariate Statistics class for many years^[1] lists 4 types of scales :

- i. nominal : discrete categories with no order (e.g. profession or gender) – qualitative.
- ii. ordinal : discrete categories with order (e.g. grades, A, B, C . . .) – qualitative.

- iii. interval : quantitative measure but no zero: ratios make no sense (e.g. temperature – makes no sense to say that one day was twice as hot as another day).
- iv. ratio : has zero, and hence ratios have meaning – quantitative.

SPSS uses :

- i. nominal.
- ii. ordinal.
- iii. scale : this scale is equivalent to the ordinal and ration scales listed above combined — as best as I can make out.

SPSS lets you specify a measurement scale under the “Measure” column in the “Variable View” screen. My recommendation is to leave it at “Unknown” or set it to “Scale”, otherwise it will try to restrict the statistical tests you can do when you don’t want it to. Measurement scales were invented to guide you to an appropriate statistical test but it doesn’t work that well. Instead, consider if your variable is continuous or discrete and then think about your situation.

1.2.3 Kinds of Sampling and Studies

This material properly belongs to a course on research methods and experimental design, but we will take a very quick look here. Ultimately your data need to be selected from the population at random. All mathematical statistical tests assume *random sampling*. The probability distributions that are used are *defined* by random sampling (the randomness — probability distribution relationship is pretty much a tautology). The real world is nit ideal, however, and you may be forced to deal with bias introduced by the following sampling schemes :

1. Random Sampling : Samples selected from the population at random.
2. Systematic Sampling : The population is ordered somehow (e.g. by house address or by phone number) and there is a rule for selecting samples (e.g. every 4th house or every 10th phone number).
3. Stratified Sampling : The population is, or can be, ordered into groups and sampling is done at random from the groups.
4. Cluster Sampling : Restrict sampling to a few groups of the population (a few strata).

And, depending on the control you have over your independent variable, studies may be classified as :

1. **Observational Study** : Just watch. You have no control over the independent variables.
2. **Experimental Study** : Control some variables to isolate other variables. The object is to manipulate the independent variable.

Astronomy is a passion of mine; observing stars and planets through a telescope is an example of an observational study. Experimental studies can be affected (knowingly or unknowingly) by *confound variables*. These are causes (independent variables) that you are not interested in but which affect the outcome (dependent variables) and can lead to data *bias* that you need to account for. Such issues are beyond the scope of an introductory statistics course.

-
1. Bluman AG, Elementary Statistics: A Step-by-Step Approach, numerous editions, McGraw-Hill Ryerson, circa 2005. ↩
-

This page titled [1.2: Basic Definitions](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Gordon E. Sarty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.