

VARIATION AND CRITICAL EVALUATION*

Ananda Mahto

Based on *Sampling and Data: Variation and Critical Evaluation*[†] by

Susan Dean

Barbara Illowsky, Ph.D.

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 3.0[‡]

Abstract

This module discusses statistical variability within data and samples. Students will be given the opportunity to see this variability in action through participation in an optional classroom exercise. This module also has a section that discusses Critical Evaluation.

1 Variation in Data

Variation is present in any set of data. For example, 16-ounce cans of beverage may contain more or less than 16 ounces of liquid. In one study, eight 16 ounce cans were measured and produced the following amount (in ounces) of beverage:

15.8, 16.1, 15.2, 14.8, 15.8, 15.9, 16.0, 15.5

Measurements of the amount of beverage in a 16-ounce can may vary because different people make the measurements or because the exact amount, 16 ounces of liquid, was not put into the cans. Manufacturers regularly run tests to determine if the amount of beverage in a 16-ounce can falls within the desired range.

Be aware that as you take data, your data may vary somewhat from the data someone else is taking for the same purpose. This is completely natural. However, if two or more of you are taking the same data and get very different results, it is time for you and the others to reevaluate your data-taking methods and your accuracy.

2 Variation in Samples

It was mentioned previously that two or more **samples** from the same **population** and having the same characteristics as the population may be different from each other. Suppose Doreen and Jung both decide to study the average amount of time students sleep each night and use all students at their college as the population. Doreen uses systematic sampling and Jung uses cluster sampling. Doreen's sample will be different from Jung's sample even though both samples have the characteristics of the population. Even if

*Version 1.2: Sep 12, 2012 2:08 am -0500

[†]<http://cnx.org/content/m16021/1.14/>

[‡]<http://creativecommons.org/licenses/by/3.0/>

Doreen and Jung used the same sampling method, in all likelihood their samples would be different. Neither would be wrong, however.

Think about what contributes to making Doreen's and Jung's samples different.

If Doreen and Jung took larger samples (i.e. the number of data values is increased), their sample results (the average amount of time a student sleeps) would be closer to the actual population average. But still, their samples would be, in all likelihood, different from each other. This **variability in samples** cannot be stressed enough.

2.1 Size of a Sample

The size of a sample (often called the number of observations) is important. The examples you have seen in this book so far have been small. Samples of only a few hundred observations, or even smaller, are sufficient for many purposes. In polling, samples that are from 1200 to 1500 observations are considered large enough and good enough if the survey is random and is well done. You will learn why when you study confidence intervals.

2.2 Optional Collaborative Classroom Exercise

Exercise 1

Divide into groups of two, three, or four. Your instructor will give each group one 6-sided die. **Try this experiment twice.** Roll one fair die (6-sided) 20 times. Record the number of ones, twos, threes, fours, fives, and sixes you get below ("frequency" is the number of times a particular face of the die occurs):

First Experiment (20 rolls)

Face on Die	Frequency
1	
2	
3	
4	
5	
6	

Table 1

Second Experiment (20 rolls)

Face on Die	Frequency
1	
2	
3	
4	
5	
6	

Table 2

Did the two experiments have the same results? Probably not. If you did the experiment a third time, do you expect the results to be identical to the first or second experiment? (Answer yes or no.) Why or why not?

Which experiment had the correct results? They both did. The job of the statistician is to see through the variability and draw appropriate conclusions.

2.3 Running this experiment in R

While it is much more interesting to experiment using a real die, we can also simulate it in R. To do this, we first create an object in R with the numbers 1 through 6 (representing the six faces on a die). After that, we can use the `sample()` function to take as many samples as we require, adding, of course, `replace = TRUE` to use sampling with replacement.

```
# Six faces on a die
die.face = c(1:6)
set.seed(123)
aa = sample(die.face, 20, replace = TRUE)
# Create a frequency table to analyze the
# results. The first row shows the number on the
# die face. The second row shows the frequency
# at which that number was rolled.
table(aa)
## aa
## 1 2 3 4 5 6
## 3 3 3 3 2 6
bb = sample(die.face, 20, replace = T)
table(bb)
## bb
## 1 2 3 4 5 6
## 2 4 1 4 5 4
```

Notice the difference in the results of table `aa` and table `bb`. You can expect similar variation in your results when manually rolling the die.

3 Critical Evaluation

We need to critically evaluate the statistical studies we read about and analyze before accepting the results of the study. Common problems to be aware of include

- **Problems with Samples:** A sample should be representative of the population. A sample that is not representative of the population is biased. Biased samples that are not representative of the population give results that are inaccurate and not valid.
- **Self-Selected Samples:** Responses only by people who choose to respond, such as call-in surveys are often unreliable.
- **Sample Size Issues:** Samples that are too small may be unreliable. Larger samples are better if possible. In some situations, small samples are unavoidable and can still be used to draw conclusions, even though larger samples are better. Examples: Crash testing cars, medical testing for rare conditions.
- **Undue influence:** Collecting data or asking questions in a way that influences the response.

- **Non-response or refusal of subject to participate:** The collected responses may no longer be representative of the population. Often, people with strong positive or negative opinions may answer surveys, which can affect the results.
- **Causality:** A relationship between two variables does not mean that one causes the other to occur. They may both be related (correlated) because of their relationship through a different variable.
- **Self-Funded or Self-Interest Studies:** A study performed by a person or organization in order to support their claim. Is the study impartial? Read the study carefully to evaluate the work. Do not automatically assume that the study is good but do not automatically assume the study is bad either. Evaluate it on its merits and the work done.
- **Misleading Use of Data:** Improperly displayed graphs, incomplete data, lack of context.
- **Confounding:** When the effects of multiple factors on a response cannot be separated. Confounding makes it difficult or impossible to draw valid conclusions about the effect of each factor.

Glossary

Definition 1: Population

The collection, or set, of all individuals, objects, or measurements whose properties are being studied.

Definition 2: Sample

A portion of the population under study. A sample is representative if it characterizes the population being studied.