

1.4: Central Tendency and Variability

Central Tendency

Central tendency refers to statistical measures that summarize the typical or average value within a group of numbers. These measures include the mean, mode, and median.

- **Mean:** The mean is the most commonly used measure of central tendency. It is calculated by summing up all values in a dataset and dividing the sum by the total number of cases. The mean has the advantageous mathematical property of minimizing variance.
- **Median:** The median represents the middle score or measurement in ranked scores or measurements. It divides the distribution into two halves. If the number of scores is even, the median is the average of the two middle scores.
- **Mode:** The mode is the most frequent score in a dataset. It represents the value that occurs most often among the data points (Vogt & Johnson, 2011).

Variability

Variability refers to the extent to which individual scores in a dataset differ. It measures the dispersion or spread of scores around a central tendency, such as the mean. Two commonly used measures of variability are variance and the standard deviation.

Variance: Variance quantifies the spread of scores in a distribution. A larger variance indicates that individual scores are more spread out from the mean, while a smaller variance indicates that scores are closer to the mean. It is calculated as the average of the squared deviations from the mean, representing the average squared distance of each score from the mean.

Standard Deviation: The standard deviation is the square root of the variance. It provides a measure of variability in the original units of measurement. By taking the square root of the variance, we obtain a measure that is more interpretable and easier to understand (Vogt & Johnson, 2011).

We can compute the central tendency and variability measures using R. We will use the gapminder database, a well-known dataset used in data analysis and visualization. It contains socio-economic indicators for countries around the world over several decades. We will use R data packages to get this data to download the dataset.

Gapminder Data Package

You can easily install and load data packages in R using the `library()` function when you want to access the data. Let's use the gapminder data package as an example. Created by Jennifer Bryan for educational purposes, the gapminder data package provides a simplified version of the original Gapminder database found at gapminder.org. This package contains a subset of the data, including six variables (country, continent, year, life expectancy at birth, total population, and GDP per capita) for 142 countries, recorded every five years from 1952 to 2007.

In this dataset, each row represents a country, whereas each column represents a variable. To install and load the gapminder package, follow the same steps you would take to install and load any other R package, as discussed in previous chapters.

```
install.packages("gapminder")  
library(gapminder)
```

? And Data

There is a simple way to get more information about the gapminder package.

```
?gapminder
```

This command will open the documentation page for the gapminder dataset, providing details about its structure, variables, and usage. You can also find information on how to load the dataset into your R environment and explore its contents. Now that you have installed and loaded the gapminder package, load the gapminder dataset into the current R session:

```
data("gapminder")
```

By executing `data(gapminder)`, the gapminder dataset will be available for use in your R environment.

Subset Function

Let's say that we want to know the mean of the total population of 142 countries in 2007. But, the current gapminder dataset has all data for every 5 years from 1952 to 2007. We can subset the gapminder dataset for the year 2007 using the filter function of the tidyverse package.

```
library(tidyverse)
gapminder.2007 <- gapminder %>%
  filter(year == 2007)
```

The code above subsetting the gapminder dataset to include only observations for the year 2007.

Mean and Median

We will first use this subsetting dataset to compute the mean and median since the steps to get the mode are slightly more complicated.

```
mean(gapminder.2007$pop)
median(gapminder.2007$pop)
```

The above code produces the mean and median: The mean total population in 2007 for the countries in the gapminder dataset is approximately 44,021,220, and the median total population in 2007 for the countries in the gapminder dataset is approximately 10,517,531.

Mode

To compute the mode of the total population in 2007 for the countries in the gapminder dataset, you need to download the DescTools package that allows you to use the mode function.

```
install.packages("DescTools")
library(DescTools)
Mode(gapminder.2007$pop)
```

It is not an error that you do not see a single value for the mode; this happened because there was no most frequent score in the dataset. There was no same number of total population across countries in 2007.

Variance and Standard Deviation

Using the following syntaxes, you can compute the variance and standard deviation of the total population in 2007 for the countries in the gapminder dataset.

```
var(gapminder.2007$pop)
sd(gapminder.2007$pop)
```

Conclusion

In this chapter, we learned how to use R data packages and how to compute central tendency and variability measures. Central tendency and variability are critical statistical concepts that provide valuable insights into the characteristics of a dataset. In the next chapter, we will learn how to check the reliability of a scale.

References

Vogt, W. P., & Johnson, R. B. (2011). Dictionary of statistics & methodology: A nontechnical guide for the social sciences. Sage.

This page titled [1.4: Central Tendency and Variability](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Jaeyong Choi \(The Pennsylvania Alliance for Design of Open Textbooks \(PA-ADOPT\)\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.