Distributions

by Sophia

☰ WHAT'S COVERED

This tutorial will cover the topic of distributions. Our discussion breaks down as follows:
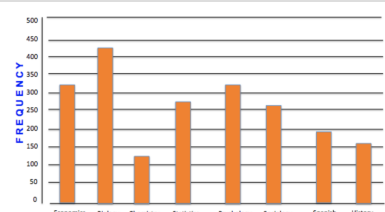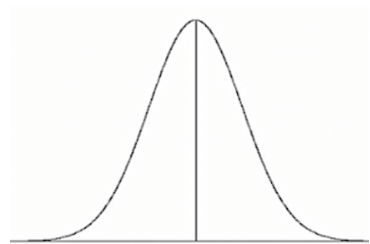
---

1. Distributions

A **data set** is not just a random list of numbers or values; there is some context associated with it, usually the units, or what type of measurement is used, or perhaps some kind of descriptor. Usually, multiple variables comprise the data set.

A **variable** is any single characteristic of the individual members of the population that can be measured. A variable of interest can take on different values for each member of the population.

> ⤷ EXAMPLE For example, suppose we are interested in the variable of height for a group of people. This could vary from person to person because people have different heights.

A **distribution** is a way to visually show how many times a variable takes a certain value; it is the values the variable takes and how often they show up. There are many kinds of distributions:

Types of Distributions	Description	Examples

Frequency tables	Can visually show how often a variable takes on a certain value	**Height / Frequency table:** 55: 11, 56: 21, 57: 33, 58: 37, 59: 55, 60: 51, 61: 44, 62: 32, 63: 30, 64: 12, 65: 7
Qualitative Data	The variables in these distributions are categories.	Bar Graphs Pie Charts Dot Plots
Quantitative Data	The variables in these distributions are measures of values or counts.	Stem-and-Leaf Plots Dot Plots Histograms Line Charts Time-Series Diagrams
Mathematical Rules	Can visually show variables through a certain pattern and are not strictly data-driven.	Normal Distribution Poisson Distribution

For the Frequency tables row:

Height	Frequency
55	11
56	21
57	33
58	37
59	55
60	51
61	44
62	32
63	30
64	12
65	7

For the Quantitative Data stem-and-leaf plot:

```
4 | 0
3 | 0 1 2 4 6 6 8
2 | 0 2 3 6 8 9      Key: 2 | 0 means GPA rounds to 2.0
1 | 9
```

📄 **TERMS TO KNOW**

Data Set

A collection of responses or observations associated with a particular context and collected from a sample or population.

Variable

A measurable factor, characteristic, or attribute of an individual or a system.

Distribution
> A way to visually display the values a variable takes and how often it takes each value.

3. Matching Distribution Types to Data Sets

Why are there so many different kinds of distributions? The point of a distribution is to make the data--possibly a large data set that is unwieldy--simpler to understand. You want to make it easy for yourself and your readers to understand. Therefore, different kinds of distributions will lend themselves better to different types of data sets.

⤷ **EXAMPLE** A dot plot is better for data that are close together and doesn't have a lot of values, whereas certain other distributions are better for larger data sets. A histogram is better than a dot plot when the data is very spread out.

You can determine which kind of distribution to use based on the kind of data you have.

> ☆ **BIG IDEA**

Each distribution has its own situation for which it is ideal. The data will determine which distribution is best to use.

> 🗒 **SUMMARY**
>
> There are many types of distributions. The point of all of them is to visually display your data so the reader can take a large data set and succinctly understand what is going on with it. Some distributions contain every observation or data point, and some only contain summaries; you can match your distribution types to the data set. Each type of distribution discussed here can be explored further in its own tutorial.
>
> Good luck!

Source: THIS TUTORIAL WAS AUTHORED BY JONATHAN OSTERS FOR SOPHIA LEARNING. PLEASE SEE OUR **TERMS OF USE**.

> 📄 **TERMS TO KNOW**
>
> **Data Set**
> > A collection of responses or observations associated with a particular context and collected from a sample or population.
>
> **Distribution**
> > A way to visually display the values a variable takes and how often it takes each value.

Variable

A measurable factor, characteristic, or attribute of an individual or a system.