

4.1: Introduction to Distributions

Refresher

Let's remind ourselves three important ideas we learned about last chapter: distribution, central tendency, and variance. These terms are similar to their everyday meanings (although I suspect most people don't say central tendency very often).

Distribution

When you order something from Amazon, where does it come from, and how does it get to your place? That stuff comes from one of Amazon's distribution centers. They distribute all sorts of things by spreading them around to your doorstep. "To Distribute" is to spread something. Notice, the data in the histogram is distributed, or spread across the bins. We can also talk about a distribution as a noun. The histogram is a distribution of the frequency counts across the bins. Distributions are **very, very, very, very, very** important. They can have many different shapes.

We will start this chapter on distributions talking about frequency distributions, which show the frequency of scores. When you hear "distribution," "distribution of data," or "frequency distribution," you should be imagining a histogram or line graph (frequency polygon).

- The X-axis has all possible scores.
- The Y-axis has frequencies of occurrences.

As we learn about other types of distributions, the Y-axis might be measured differently (like in percentages or probability), but the idea is the same: How often did that score occur?

Measures of Central Tendency

Central Tendency is all about sameness: What is common about some numbers? For example, if we had a distribution in which most of scores were near 0, we could say that there is a tendency for most of the numbers to be centered near 0. Notice we are being cautious about our generalization about the numbers. We are not saying they are all 0. We are saying there is a tendency for many of them to be near zero. There are lots of ways to talk about the central tendency of some numbers. There can even be more than one kind of tendency. For example, if lots of the numbers were around -1000, and a similar large amount of numbers were grouped around 1000, we could say there was two tendencies. The three common measures of central tendency that we discussed last chapter are:

- Mode
- Median
- Mean

? Exercise 4.1.1

Which measures of central tendency can be used with quantitative data?

Answer

Add texts here. Do not delete this text first. All three measures of central tendency can be used with quantitative data. The mode is the only measure of central tendency that can be used with qualitative variables, though.

Measures of Variance

Variance (or variability) is all about differentness: What is different about some numbers? For example, is there anything different about all of the numbers in the histogram? YES!!! The numbers are not all the same! When the numbers are not all the same, they must vary. So, the variance in the numbers refers to how the numbers are different. There are many ways to summarize the amount of variance in the numbers, (see [Measures of Variability](#)).

The most common measure of variance that we will use is standard deviation. Some researchers also like to know the range, or the maximum and minimum scores (highest score and lowest score).

Why?

As we begin talking about different kinds of special distributions, you might be asking yourself why these matter. We will go through some of these special distributions to show why statistical analyses can be used to predict information about a population from information about a sample. This will, ultimately, let us test research hypotheses. I could explain why, but it wouldn't make sense because you don't know about the special distributions yet! So, for now, let's learn the basics about some of these special distributions so that you can interpret your results later.

Let's begin!

Contributors and Attributions

- [Matthew J. C. Crump](#) ([Brooklyn College of CUNY](#))
- [Dr. MO](#) ([Taft College](#))

This page titled [4.1: Introduction to Distributions](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Matthew J. C. Crump](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- **2.3: Important Ideas - Distribution, Central Tendency, and Variance** by [Matthew J. C. Crump](#) is licensed [CC BY-SA 4.0](#). Original source: <https://www.crumplab.com/statistics/>.