

1.6: Outliers

Learning Objectives

At the end of this section you should be able to answer the following questions:

- Explain the differences between a spurious and non-spurious outlier.
- Identify the level of z score for a response that typically indicates an outlier.

One of the major concerns when analysing data is the effect that *outliers* – which are unusually high or low data points – can have on the overall results. For example, if you were asking everyday people how many cups of coffee they consume a day, and most of the responses were between zero to four, that would be a normal spread of responses.

However, if you had one participant who responded that they consumed 17 cups of coffee a day, we would consider this response to be an outlier when compared to the rest of the participants. We could assume this participant has either a caffeine problem or they have incorrectly entered their response. Either way, this response will increase the mean for this particular sample, without being representative of the average coffee drinker.

There are a number of ways to statistically identify outliers in your data set. Participant responses to any variable can be transformed to a “z score,” which is a basic transformation allowing you to compare responses across cases to a standardized response, which has a mean of 0 and standard deviation of 1. If a response has a z score of greater than ± 3.3 , it is to be considered to be an outlier. Another way is to graph the data using a box plot or bar graph, and visually identify the outliers. We will run through these options in greater detail later.

Normally, when you find outliers you can do two things: include them in the final analysis if you consider the outliers to be *non-spurious*, or you can remove them if the outliers have occurred for *spurious* reasons, meaning that they do not reflect accurate responses. If the outliers don't make sense in the context of the question, or are extreme without any potential justification, is a good idea to consider these as spurious responses and just remove them from the analysis. However, if you do find some responses that make sense or are only slightly outside the acceptable z score (± 3.3), it may be worth considering them to be non-spurious outliers and keeping them for analysis.

This page titled [1.6: Outliers](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Erich C Fein, John Gilmour, Tayna Machin, and Liam Hendry](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.