

### 3.1: The Goal of Summaries

When summaries are created, a statistician must balance two competing goals: gaining simplicity and retaining specificity. To simplify data means to reduce what is shown or stated to emphasize a key aspect or aspects of the data or the variable the data represent. To specify data means to retain and show details about the data or the variable the data represent. These two goals are in conflict with each other and some of the work a statistician must do is aimed at balancing the two so that simplicity is gained without losing too much specificity. Different summaries can be created, some of which emphasize simplicity over specificity and others that focus on adding some specificity back in. The measures of central tendency are focused primarily on simplicity rather than specificity. However, there are different things that one can focus on when summarizing data for simplicity. Each measure of central tendency focuses on one aspect of the data when summarizing. Therefore, we must understand the focus of each measure of central tendency to know which one will provide the best summary for a given situation.

Take a look at Data Set 3.1; there are 22 scores for age. It is hard to state anything succinct about age when we have 22 pieces of information about age in their raw form. The more data we have, the harder it gets to simply eyeball what we have in the dataset. Imagine how hard it would be to look at data for age from 10,000 cases and get a clear sense of what the data showed. To reduce the strain, data can be summarized with a measure of central tendency rather than simply viewed in their raw form. Take a moment to look at Data Set 3.1 and try to describe the data for age succinctly.

Data Set 3.1

Age
47
46
42
39
36
34
33
33
32
29
29
29
28
27
25
23
20
19
19
18
16
14

How did you describe or summarize the data for age in Data Set 3.1? You may have noticed that age varied and included teens through adults. You may have noticed that the lowest age was 14 and that the highest age was 47, that most people were in their teens, 20s or 30s, and that only 3 people were in their 40s. You may have even noticed that most ages were only reported once but that 29 was reported three times. Though these statements are true, they all differ and none of them provides a single number or description that can be used to summarize all of the data. In addition, different individuals might focus on a different datum when summarizing and may use different words to describe what they notice in the data. These differences can create confusion when several individuals are trying to understand the same variable, even when they are looking at the same data for that variable. These differences can make it especially difficult to understand the similarities or dissimilarities across different datasets. Therefore, it is helpful to have tools, such as measures of central tendency that allow us to make simple descriptive statements.

Measures of central tendency are strategies used to summarize raw data. Though there are many, we will focus on three common measures of central tendency: the mode, the median, and the mean. Each of these is calculated or identified in a systematic way. One of the benefits of using one of these techniques is that others will recognize the name of the measure of central tendency and know exactly which calculations were used to summarize the data. This helps improve clarity when summaries are shared with others, including large audiences.

---

This page titled [3.1: The Goal of Summaries](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .