

What is Winsorizing data and what are some examples of it?

Authored by
stats writer

April 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Winsorizing data and what are some examples of it?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138940>

Winsorizing data is a statistical technique used to minimize the effects of extreme or outlier values in a dataset. This is achieved by replacing these extreme values with the next highest or lowest value in the dataset, depending on the type of Winsorization being applied. This process helps to reduce the impact of outliers on statistical analyses and allows for more accurate and reliable results.

Some examples of Winsorizing data include:

1. Winsorizing in finance: In finance, Winsorizing is commonly used to adjust the returns of a stock or portfolio. This helps to reduce the effect of extreme fluctuations in stock prices, which can skew the overall performance.
2. Winsorizing in economics: In economic analysis, Winsorizing is used to adjust economic data that may contain outliers, such as income distribution or consumer spending. This helps to avoid biased results and provides a more accurate representation of the data.
3. Winsorizing in healthcare: In healthcare, Winsorizing is used to adjust medical data that may contain outliers, such as patient length of stay or medical costs. This helps to identify trends and patterns more accurately, which can aid in decision making and resource allocation.

Overall, Winsorizing is a useful technique for handling extreme values in a dataset and promoting more reliable and accurate statistical analysis.

Winsorize Data: Definition & Examples

To winsorize data means to set extreme outliers equal to a specified percentile of the data.

For example, a 90% winsorization sets all greater than the 95th percentile equal to the value at the 95th percentile and all observations less than the 5th percentile equal to the value at the 5th percentile.

In effect, to winsorize data means to change extreme

values in a dataset to less extreme values.

Example: How to Winsorize Data

Suppose we have the following dataset:

3, 14, 16, 16, 17, 29, 34, 36, 39, 47, 59, 64, 65, 66, 68, 79, 91, 98

To perform a 90% winsorization on this dataset, we would first find the 5th percentile and the 95th percentile, which turn out to be:

5th percentile: 12.35 95th percentile: 92.05

We would then set any values below 12.35 equal to 12.35 and any values above 92.05 equal to 92.05:

12.35, 14, 16, 16, 17, 29, 34, 36, 39, 47, 59, 64, 65, 66, 68, 79, 91, 92.05

In this case, the value 3 became changed to 12.35 and the value 98 became changed to 92.05.

Why Winsorize Data?

The mean and the standard deviation are two common

ways to measure the of a dataset and in a dataset, respectively.

However, these two metrics can both be influenced by extreme outliers. Thus, winsorizing data allows us to set extreme outliers equal to less extreme values.

This often allows us to get a more accurate view of the mean and the standard deviation of the dataset.

Trimming vs. Winsorizing

Another common way to deal with outliers is to trim them from the dataset, which means to remove them entirely.

3, 14, 16, 16, 17, 29, 34, 36, 39, 47, 59, 64, 65, 66, 68, 79, 91, 98

If we wanted to trim the values that fall below the 5th percentile or above the 95th percentile, we would simple remove the values 3 and 98.

Here are a couple rules of thumb for when to use trimming vs winsorizing:

Trimming: It makes sense to trim data values when some values seem completely unreasonable, i.e. they're a result of a data entry error.

Winsorizing: It makes sense to winsorize data when we want to retain the observations that are at the extremes but we don't want to take them too literally.

Cautions on Winsorizing Data

Here are a few things to keep in mind when deciding to winsorize data:

1. If there aren't extreme outliers, then winsorizing the data will only modify the smallest and largest values slightly. This is generally not a good idea since it means we're just modifying data values for the sake of modifications.
2. Outliers can represent interesting edge cases in the data. Thus, before modifying outliers it's a good idea to take a closer look at them to see what could have caused them.
3. You should decide whether or not to winsorize data *after* collecting the data, not before. You should see if

there actually are extreme outliers before you decide to perform winsorization. If no extreme outliers are present, winsorization may be unnecessary.

Tutorial: Winsorize Data in Excel

Refer to for a step-by-step example of how to winsorize a dataset in Excel.

ARABPSYCHOLOGY.COM