

What is the definition of truncated and censored data, and can you provide some examples?

Authored by
stats writer

April 27, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the definition of truncated and censored data, and can you provide some examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=139939>

Truncated and censored data are two types of incomplete data that can occur in statistical analysis. Truncated data refers to a dataset where certain values or observations are missing due to a selection process, usually based on a specific criteria. This can occur when only a certain group or range of values are included in the dataset, while others are excluded. For example, a survey may only include responses from individuals between the ages of 18-65, truncating the data for those outside of this age range.

On the other hand, censored data refers to a dataset where certain values or observations are missing due to the limitations of the data collection process. This can occur when the true value of a variable is unknown or unmeasurable, and is replaced with a known value such as a range or a symbol. For instance, in a medical study, a patient's exact time of death may not be known, and is instead recorded as "greater than 5 years" for censorship purposes.

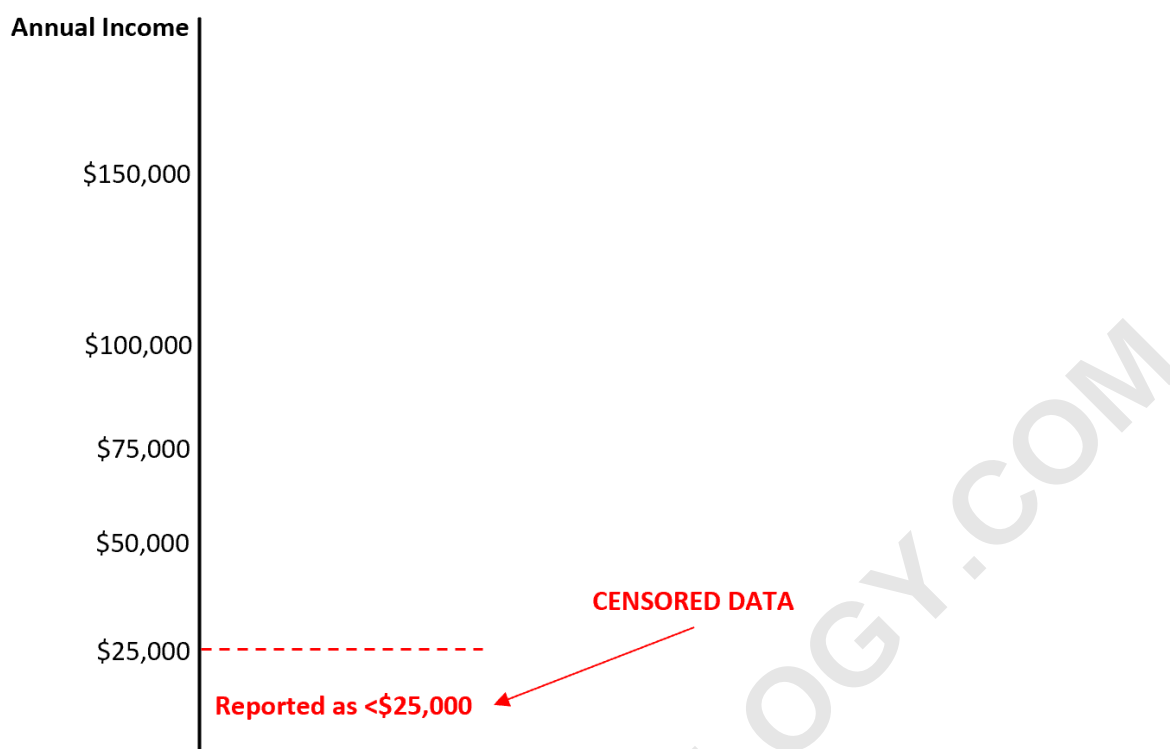
In both cases, truncated and censored data can affect the accuracy and validity of statistical analysis, as the missing values may not accurately represent the true population. It is important for researchers to be aware of these types of incomplete data and properly account for them in their analyses.

Truncated & Censored Data: Definition + Examples

Often when collecting data, researchers may decide to censor or truncate certain values.

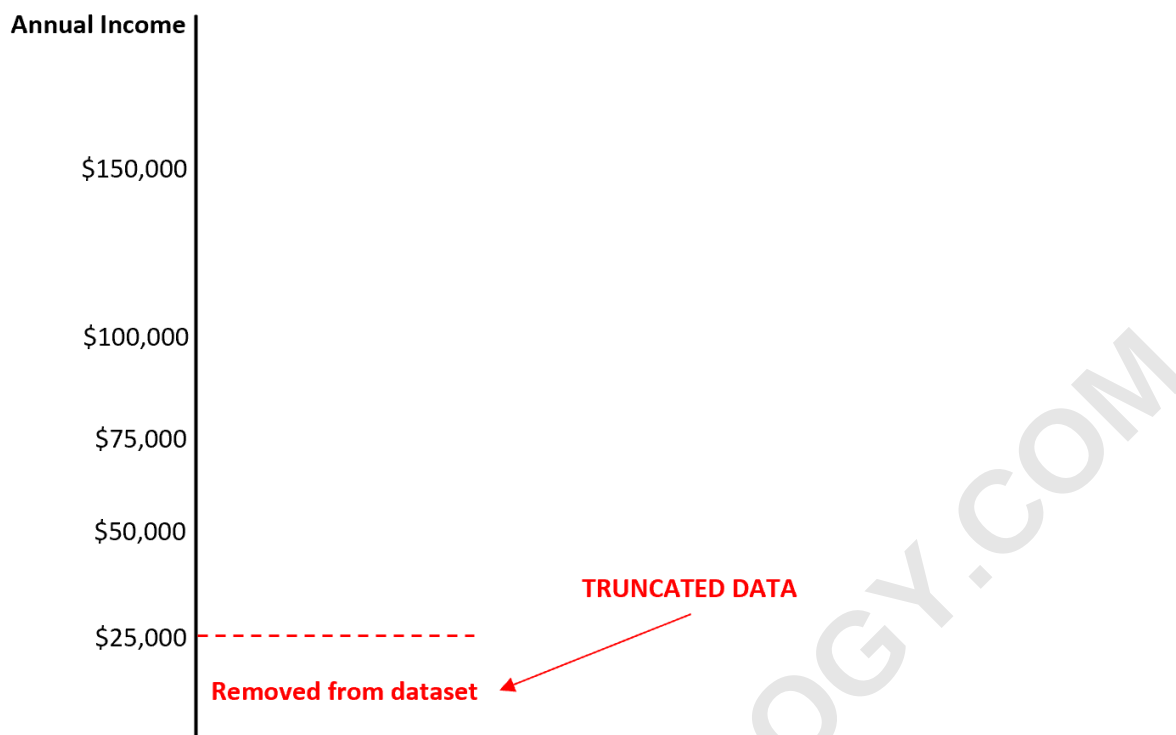
To censor data values means to only collect partial information about values that fall below or above a certain value.

For example, we may know that an individual earns less than \$25,000 per year but we may not know their exact annual income.



To truncate data values means to remove values from a dataset that fall below or above a certain value.

For example, a researcher may only be interested in studying individuals who earn more than \$25,000 per year. Thus, any individuals who earn less than \$25,000 are simply removed from the dataset.



This tutorial provides several examples of when data may be either censored or truncated.

Censoring Data

To censor data values means to only collect partial information about values that fall below or above a certain value.

The following examples illustrate scenarios where we may decide to censor data values.

Example 1: Annual Income

Suppose a researcher is collecting survey data about annual income. If an individual earns less than \$25,000 per year he decides to report this income as "<\$25,000" in a database rather than specifying their exact annual income.

This represents an example of censoring data because we know that an individual earns less than a certain amount but we don't know their exact annual income.

Example 2: Pollution Levels

Suppose a biologist uses a certain tool to measure the pollution levels in different bodies of water. Her tool is incapable of measuring pollution below .002 parts per million. Thus, any body of water that has pollution levels below this threshold will simply be reported as "<.002" rather than the exact amount.

This represents an example of censoring data because we know that certain bodies of water have pollution levels below .002 parts per million, but we don't know their exact pollution levels.

Truncating Data

To truncate data values means to remove values from a dataset that fall below or above a certain value.

The following examples illustrate scenarios where we may decide to truncate data values.

Example 1: Number of Crimes

Suppose a law enforcement officer is researching the types of crimes committed by individuals in a certain area. By default, any individual who has committed 0 crimes will not be included in the dataset because they haven't committed any type of crime.

This represents an example of truncating data because any individual who has committed 0 crimes is simply excluded from the dataset entirely.

Example 2: Education Level

Suppose a professor wants to study the relationship between a certain study program and student achievement.

Because of the intensity of the study program, the professor only wants to monitor students who currently

have a GPA greater than 3.5. Thus, any student who applies to the program but has a GPA less than 3.5 will simply not be included in the program

This represents an example of truncating data because any individual who has a GPA below a certain threshold is simply excluded from the dataset.

Summary

To censor data means to only collect partial information about data values and to truncate data means to remove data values from a dataset entirely.

Both censoring and truncating lead to loss of information in a dataset, but truncating results in greater information loss because it involves excluding certain data values entirely.