

What is an Open Ended Distribution?

Authored by
stats writer

April 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is an Open Ended Distribution?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=140965>

An open ended distribution refers to a type of statistical distribution that has no specific upper or lower limit. This means that the data can extend infinitely in either direction, making it possible for extreme values to occur. It is often used in situations where there is a wide range of possible outcomes, such as in financial markets or natural phenomena. Unlike a closed ended distribution, where the data is restricted to a specific range, an open ended distribution allows for more variability and the potential for outliers. This makes it a useful tool for analyzing complex and diverse data sets.

What is an Open Ended Distribution?

In statistics, an open ended distribution is a frequency distribution in which one or more classes (or "bins") are open-ended.

For example, the following frequency distribution represents an open ended distribution in which the smallest class is open ended:

Annual Income	Frequency
<\$20,000	6
\$20,000 < \$39,999	12
\$40,000 < \$59,999	17
\$60,000 < \$79,999	19
\$80,000 < \$99,999	14
\$100,000 < \$119,999	4

And the following frequency distribution shows an open ended distribution in which the largest class is open ended:

Annual Income	Frequency
\$10,000 < \$20,000	6
\$20,000 < \$39,999	12
\$40,000 < \$59,999	17
\$60,000 < \$79,999	19
\$80,000 < \$99,999	14
> \$100,000	4

Conversely, a closed ended distribution is one in which each class in the frequency distribution has an upper and lower boundary, such as the following:

Annual Income	Frequency
\$10,000 < \$20,000	6
\$20,000 < \$39,999	12
\$40,000 < \$59,999	17
\$60,000 < \$79,999	19
\$80,000 < \$99,999	14
\$100,000 < \$119,999	4

What Causes Open Ended Distributions?

Open ended distributions are often the result of researchers choosing to collect data in such a way that one of the classes ends up being open ended.

For example, suppose a researcher surveys residents in a certain city and asks them about their annual

household income.

The researcher may choose to make the largest possible response "> \$100,000" because he knows that high-income residents may not be comfortable sharing how much they make if it's significantly more than \$100,000.

Conversely, the researcher may choose to make the smallest possible response open ended because he knows that residents who earn very little will also not be comfortable sharing how little they make.

In a nutshell, researchers often include open ended classes in their surveys because they want to maximize the number of individuals who feel comfortable responding to the survey questions.

The Problem with Open Ended Distributions

The problem with open ended distributions is that true data gets . In other words, we might know the number of individuals who earn over \$100k in a certain city, but we don't actually know their exact annual incomes.

It's possible that some individuals may earn \$150k,

\$250k, \$500k, or even more but we have no idea since each of these individuals is only able to indicate that they make ">\$100,000" on the survey.

Because data is censored in open ended distributions, we're also unable to calculate the exact mean and standard deviation of the values in the dataset since we don't have access to all of the raw data values.

How to Analyze an Open Ended Distribution

Since we can't calculate the exact mean of an open ended distribution, we often use the median as a measure of the "center" of the dataset.

Recall that the median represents the middle value of the dataset.

When working with open ended distributions, we can use the following formula to find the best estimate of the median:

Best Estimate of Median: $L + ((n/2 - F) / f) * w$

where:

L: The lower limit of the median group n: The total

number of observations**F:** The cumulative frequency up to the median group
f: The frequency of the median group
w: The width of the median group

For example, suppose we have the following open ended distribution from earlier:

Annual Income	Frequency
\$10,000 < \$20,000	6
\$20,000 < \$39,999	12
\$40,000 < \$59,999	17
\$60,000 < \$79,999	19
\$80,000 < \$99,999	14
> \$100,000	4

There are a total of 72 values in the dataset. Thus, we know the median value will be located between the value of the 36th and 37th largest value in the dataset. Each of these values is located within the class "\$60,000 - \$79,999" so we know that the median income lies within this range.

Our best estimate of the median would be:

$$\text{Median: } 60,000 + ((72/2 - 25) / 19) * 19,999 = \$71,578$$

This value represents our best estimate of the median

annual income for individuals in this dataset.

ARABPSYCHOLOGY.COM