

What is Aggregation Bias and can you provide an explanation and example?

Authored by
stats writer

April 19, 2024

RECOMMENDED CITATION

stats writer (2024). *What is Aggregation Bias and can you provide an explanation and example?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137147>

Aggregation bias refers to a type of statistical bias that occurs when data is combined or aggregated in a way that skews the results. This can happen when individual data points are grouped together, leading to a distorted understanding of the overall data.

An example of aggregation bias can be seen in a study on average household income in a city. If the data is grouped by neighborhood, the results may show a higher average income, as wealthy neighborhoods would have a larger impact on the overall average. However, when the data is disaggregated by individual households, a more accurate picture may emerge, showing a wider range of incomes and a lower average.

Aggregation bias can also occur in surveys or polls, where the results may be skewed if certain demographic groups are underrepresented or not included in the sample. This can lead to inaccurate conclusions and decisions being made based on incomplete or biased data.

In order to avoid aggregation bias, it is important to carefully consider the grouping or aggregation of data and to ensure that all relevant data points are included in the analysis. Additionally, using a diverse and representative sample can help to mitigate this bias.

What is Aggregation Bias? (Explanation & Example)

Aggregation bias occurs when it is wrongly assumed that the trends seen in aggregated data also apply to individual data points.

The easiest way to understand this type of bias is with a simple example.

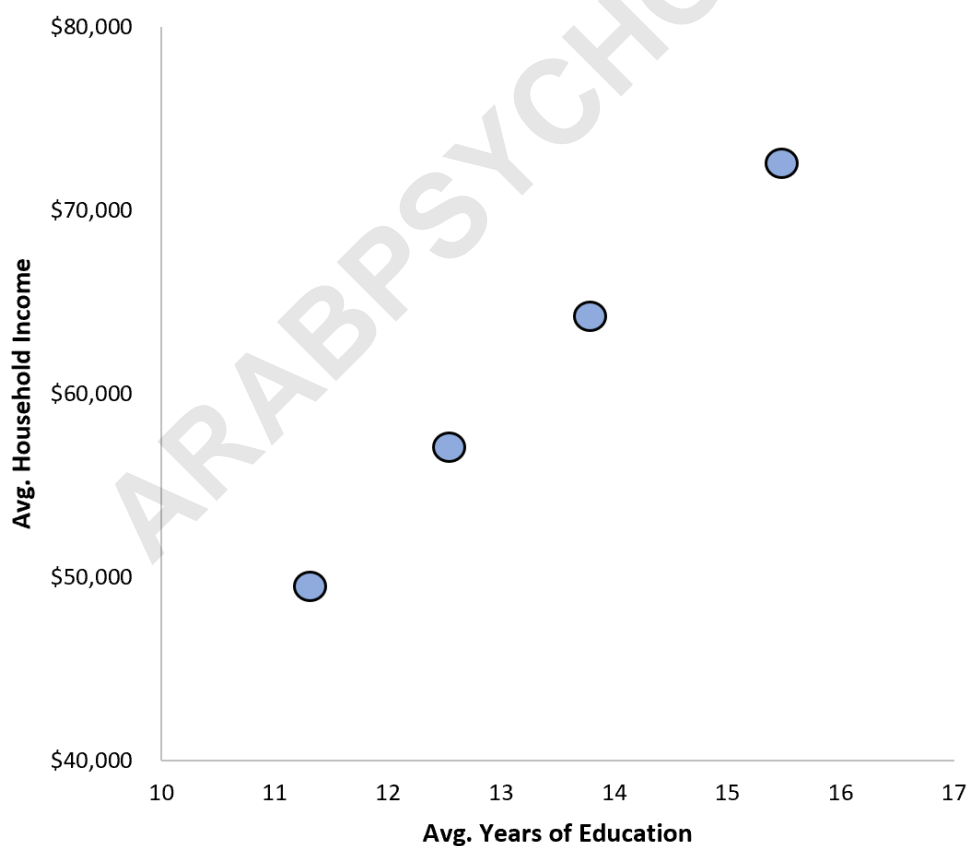
Example: Aggregation Bias

Suppose researchers want to understand the relationship between the average number of years of education and average household income in a certain state. They obtain aggregated data for 4 different cities

within the state and calculate the correlation between average education and average household income.

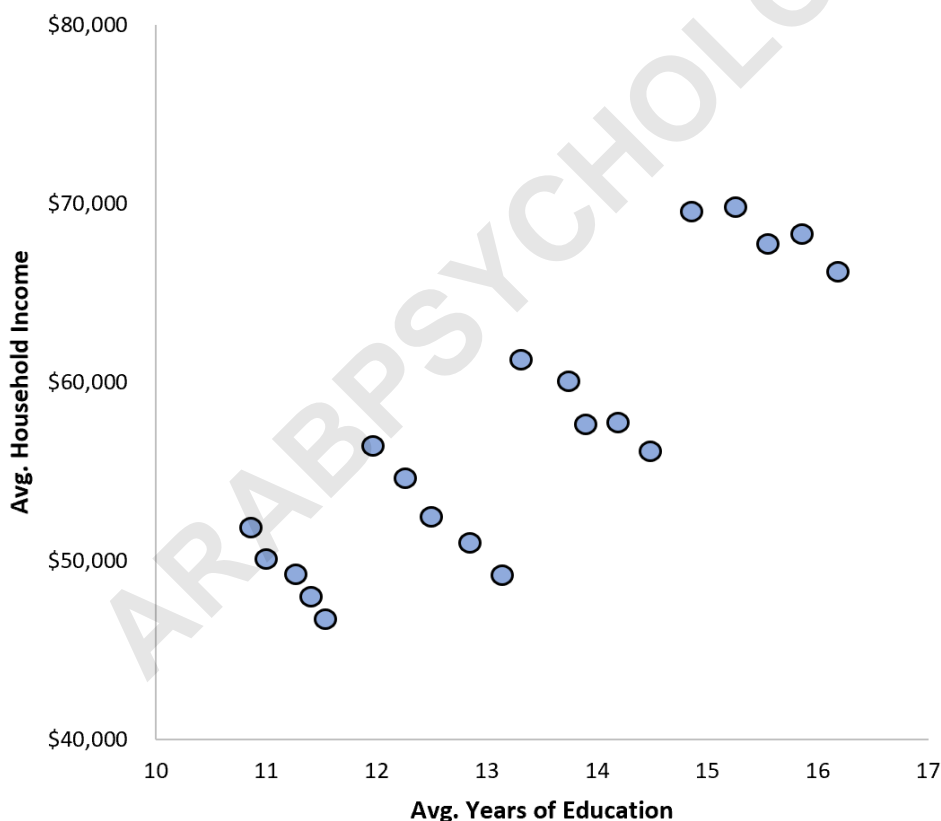
It turns out that the correlation between average number of years of education and average household income is 0.9632. This is a highly positive correlation coefficient.

The researchers even create a scatterplot to visualize the relationship between average number of years of education and average household income:



Without actually looking at the individual data, they may publish a report that claims that more years of education is strongly positively correlated with household income.

However, suppose a new researcher comes along a year later and obtains data for individual households across the same set of cities. Suppose she creates the following scatterplot of the data:

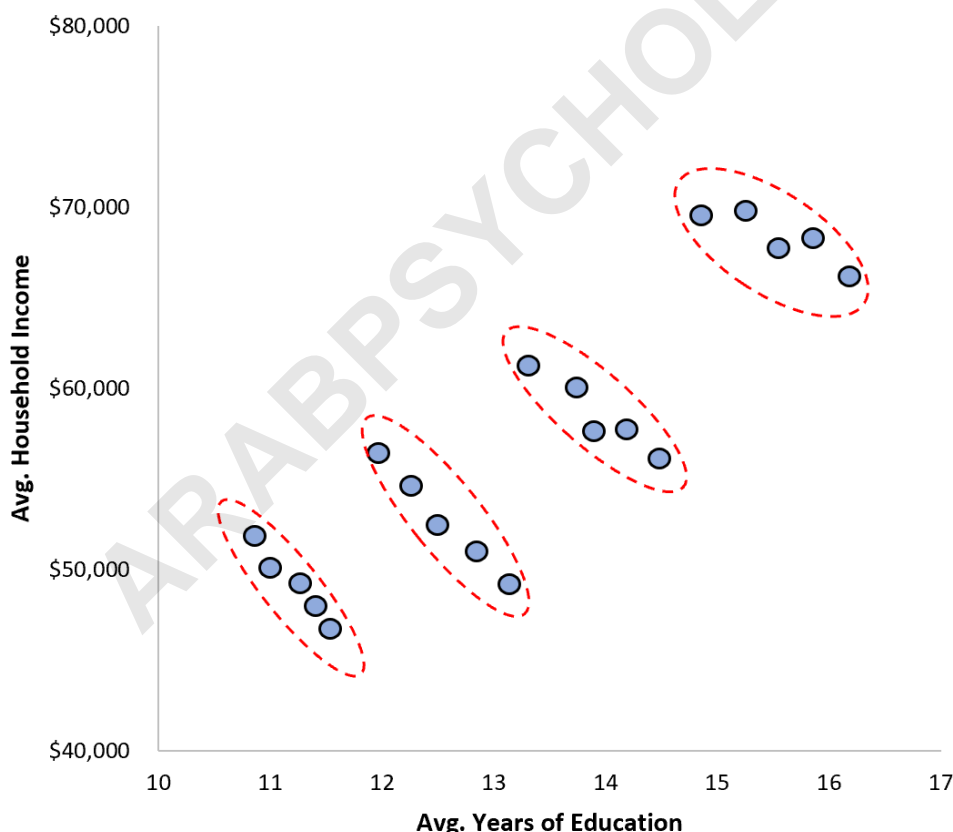


She calculates the correlation between the two variables and finds that it's actually only 0.1788 - still a

positive correlation but not nearly as strong as the correlation found by the previous researchers.

It turns out that when the data became aggregated, it covered the true trend between education and income that was taking place at the individual level.

In fact, when we look at a city-by-city basis in the scatterplot the relationship between education and income is actually negative!



Effects of Aggregation Bias

Aggregation bias occurs quite often in research simply because it's often wrongly assumed that the trends that appear at an aggregate level must also appear at an individual level. Unfortunately, this is not always the case as the previous example showed.

Aggregation bias can cause the results of a study to draw the wrong conclusions and can be misleading. This type of bias is particularly harmful when it relates to correlations between variables.

Negative correlation No correlation Positive correlation

The way to avoid this type of bias is to conduct studies using individual data points as opposed to aggregated data points so that the true relationship between two variables can be discovered.