5 Most Important Methods For Statistical Data Analysis

1. Mean

The arithmetic mean, more commonly known as "the average," is the sum of a list of numbers divided by the number of items on the list. The mean is useful in determining the overall trend of a data set or providing a rapid snapshot of your data. Another advantage of the mean is that it's very easy and quick to calculate.

Pitfall:

Taken alone, the mean is a dangerous tool. In some data sets, the mean is also closely related to the mode and the median (two other measurements near the average). However, in a data set with a high number of outliers or a skewed distribution, the mean simply doesn't provide the accuracy you need for a nuanced decision.

2. Standard Deviation

The standard deviation, often represented with the Greek letter sigma, is the measure of a spread of data around the mean. A high standard deviation signifies that data is spread more widely from the mean, where a low standard deviation signals that more data align with the mean. In a portfolio of data analysis methods, the standard deviation is useful for quickly determining dispersion of data points.

Pitfall:

Just like the mean, the standard deviation is deceptive if taken alone. For example, if the data have a very strange pattern such as a non-normal curve or a large amount of outliers, then the standard deviation won't give you all the information you need.

3. Regression

Regression models the relationships between dependent and explanatory variables, which are usually charted on a scatterplot. The regression line also designates whether those

relationships are strong or weak. Regression is commonly taught in high school or college statistics courses with applications for science or business in determining trends over time.

Pitfall:

Regression is not very nuanced. Sometimes, the outliers on a scatterplot (and the reasons for them) matter significantly. For example, an outlying data point may represent the input from your most critical supplier or your highest selling product. The nature of a regression line, however, tempts you to ignore these outliers. As an illustration, examine a picture of ANSCOMBE'S QUARTET, in which the data sets have the exact same regression line but include widely different data points.

4. Sample Size Determination

When measuring a large data set or population, like a workforce, you don't always need to collect information from every member of that population – a sample does the job just as well. The trick is to determine the right size for a sample to be accurate. Using proportion and standard deviation methods, you are able to accurately determine the right sample size you need to make your data collection statistically significant.

Pitfall:

When studying a new, untested variable in a population, your proportion equations might need to rely on certain assumptions. However, these assumptions might be completely inaccurate. This error is then passed along to your sample size determination and then onto the rest of your statistical data analysis

5. Hypothesis Testing

Also commonly called *t* testing, hypothesis testing assesses if a certain premise is actually true for your data set or population. In data analysis and statistics, you consider the result of a hypothesis test *statistically significant* if the results couldn't have happened by random chance. Hypothesis tests are used in everything from science and research to business and economic

Pitfall:

To be rigorous, hypothesis tests need to watch out for common errors. For example, the placebo effect occurs when participants falsely expect a certain result and then perceive (or actually attain) that result. Another common error is the Hawthorne effect (or observer effect), which happens when participants skew results because they know they are being studied.

Overall, these methods of DATA ANALYSIS add a lot of insight to your DECISION-MAKING PORTFOLIO, particularly if you've never analyzed a process or data set with statistics before. However, avoiding the common pitfalls associated with each method is just as important. Once you master these fundamental techniques for statistical data analysis, then you're ready to advance to more powerful data analysis tools.