What is the difference between univariate and multivariate analysis?

Authored by

stats writer

June 28, 2024

Univariate and multivariate analysis are two types of statistical techniques used to analyze data. Univariate analysis involves examining a single variable at a time, while multivariate analysis involves examining two or more variables simultaneously.

In univariate analysis, the focus is on understanding the characteristics and relationships of a single variable. This type of analysis is useful in identifying patterns, trends, and outliers within a dataset. It is commonly used in descriptive statistics and can provide a basic understanding of the data.

On the other hand, multivariate analysis involves examining the relationships between multiple variables in a dataset. This type of analysis is more complex and can provide a deeper understanding of the data by identifying the interdependencies and interactions between variables. It is commonly used in inferential statistics and can help in making predictions and drawing conclusions about a population.

In summary, the main difference between univariate and multivariate analysis is that univariate analysis focuses on a single variable, while multivariate analysis examines multiple variables to gain a more comprehensive understanding of a dataset. Both techniques have their own strengths and are commonly used in different types of research and data analysis.

Univariate vs. Multivariate Analysis: What's the Difference?

The term univariate analysis refers to the analysis of one variable. You can remember this because the prefix "uni" means "one."

The term multivariate analysis refers to the analysis of more than one variable. You can remember this because the prefix "multi" means "more than one."

There are three common ways to perform univariate analysis:

1. Summary Statistics

We can calculate like the mean or median for one variable.We can also calculate such as the standard deviation for one variable.

2. Frequency Distributions

We can create a , which describes how often each value occurs for one variable.

3. Charts

We can create charts like boxplots, histograms, density curves, etc. to visualize the distribution of values for one variable.

There are two common ways to perform multivariate analysis:

1. Scatterplot Matrix

We can create a scatterplot matrix, which allows us to visualize the relationship between each pairwise combination of variables in a dataset.

2. Machine Learning Algorithms

We can use a supervised learning algorithm to fit a model like that quantifies the relationship between multiple predictor variables and a response variable.We can also use an unsupervised learning algorithm like to find structure and relationships between multiple variables in a dataset at once.

The following examples show how to perform both univariate and multivariate analysis with the following dataset:

Household ID	Household Size	Annual Income	Number of Pets
1	2	$37,000	0
2	4	$49,000	0
3	4	$58,000	1
4	1	$68,000	3
5	3	$61,000	2
6	5	$64,000	2
7	6	$79,000	1
8	4	$89,000	1
9	7	$104,000	1
10	2	$95,000	0

Note: When you analyze exactly two variables, this is referred to as .

Example: How to Perform Univariate Analysis

We could choose to perform univariate analysis on any

of the individual variables in the dataset.

For example, we may choose to perform univariate analysis on the variable Household Size:

Household ID	Household Size	Annual Income	Number of Pets
1	2	$37,000	0
2	4	$49,000	0
3	4	$58,000	1
4	1	$68,000	3
5	3	$61,000	2
6	5	$64,000	2
7	6	$79,000	1
8	4	$89,000	1
9	7	$104,000	1
10	2	$95,000	0

We can calculate the following measures of central tendency for Household Size:

Mean (the average value): 3.8Median (the middle value): 4

These values give us an idea of where the "center" value is located.

We can also calculate the following measures of dispersion:

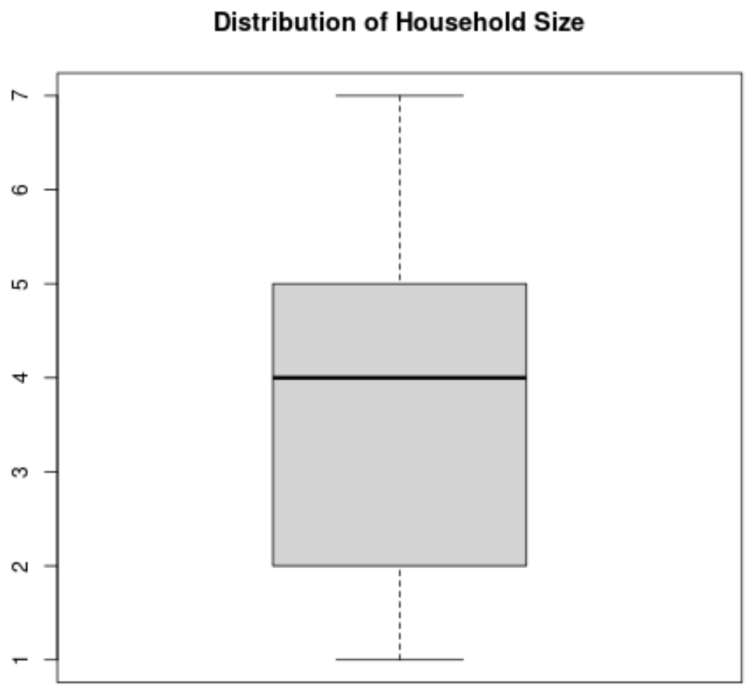
Range (the difference between the max and min): 6Interquartile Range (the spread of the middle 50% of values): 2.5Standard Deviation (an average measure of spread): 1.87

These values give us an idea of how spread out the values are for this variable.

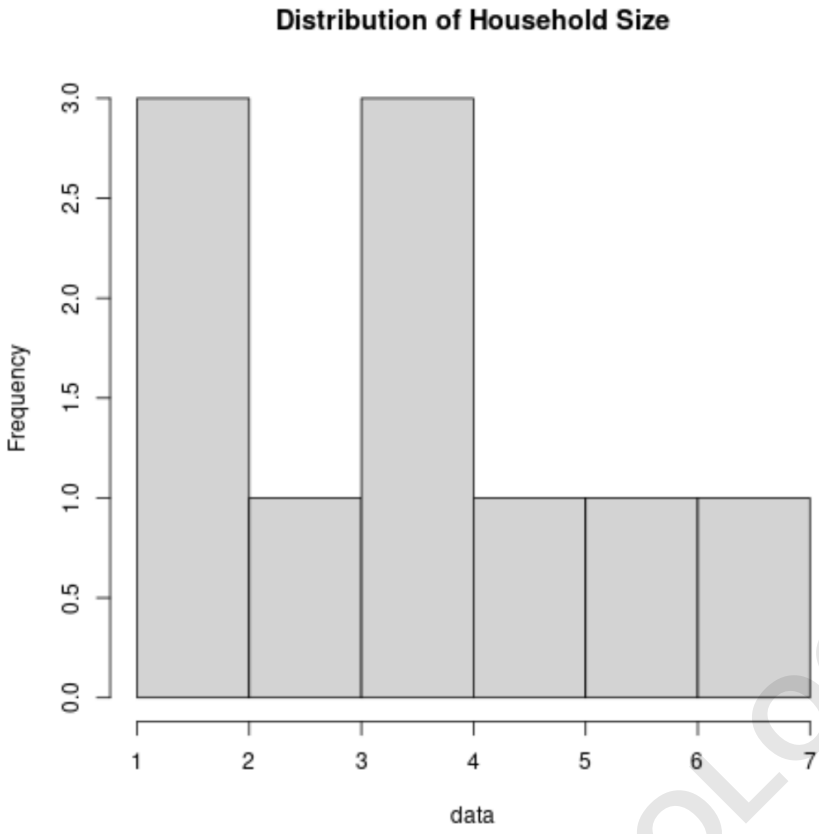
We can also create the following frequency distribution table to summarize how often different values occur:

Household Size	Frequency
1	1
2	2
3	1
4	3
5	1
6	1
7	1

We can also create a boxplot to visualize the distribution of values for household size:



Alternatively, we could create a histogram to visualize the distribution of values:



By calculating these metrics and creating these charts, we can gain a strong understanding of how the values are distributed for the variable Household Size.

Example: How to Perform Multivariate Analysis

Once again suppose we have the same dataset:

Household ID	Household Size	Annual Income	Number of Pets
1	2	$37,000	0
2	4	$49,000	0
3	4	$58,000	1
4	1	$68,000	3
5	3	$61,000	2
6	5	$64,000	2
7	6	$79,000	1
8	4	$89,000	1
9	7	$104,000	1
10	2	$95,000	0

One simple form of multivariate analysis we could perform on this dataset is to create a scatterplot matrix, which is a matrix that shows a scatterplot for each pairwise combination of numeric variables in the dataset.

We could create this type of matrix to visualize the relationship between household size, annual income, and number of pets all at once.

Resource: Check out to see how to create a scatterplot matrix in R.

Another way to perform multivariate analysis on this dataset would be to fit a multiple linear regression model. For example, we could create a regression

model that uses household size and number of pets to predict annual income.

Resource: Check out to see how to perform multiple linear regression in R.

Yet another way to perform multivariate analysis on this dataset would be to perform principal components analysis, which allows us to find an underlying structure in the dataset.

Resource: Check out to see how to perform principal components analysis in R.

Conclusion

Here's a quick summary of this article:

Univariate analysis is the analysis of one variable.Multivariate analysis is the analysis of more than one variable.There are various ways to perform each type of analysis depending on your end goal.In the real world, we often perform both types of analysis on a single dataset.Univariate analysis allows us to understand the distribution of values for one variable while multivariate analysis allows us to understand the

relationship between several variables.