With Python use the NumPy library mean() method to find the mean of the values 4,11,7,14:

import numpy

values = [4,11,7,14]

x = numpy.mean(values)

print(x)

Statistics gives us methods of gaining knowledge from data.

## What is Statistics Used for?

Statistics is used in all kinds of science and business applications.

Statistics gives us more accurate knowledge which helps us make better decisions.

Statistics can focus on making predictions about what will happen in the future. It can also focus on explaining how different things are connected.

## Typical Steps of Statistical Methods

The typical steps are:

* Gathering data
* Describing and visualizing data
* Making conclusions

It is important to keep all three steps in mind for any questions we want more knowledge about.

Knowing which types of data are available can tell you what kinds of questions you can answer with statistical methods.

Knowing which questions you want to answer can help guide what sort of data you need. A lot of data might be available, and knowing what to focus on is important.

## How is Statistics Used?

Statistics can be used to explain things in a precise way. You can use it to understand and make conclusions about the group that you want to know more about. This group is called the population.

A population could be many different kinds of groups. It could be:

* All of the people in a country
* All the businesses in an industry
* All the customers of a business
* All people that play football who are older than 45

and so on - it just depends on what you want to know about.

Gathering data about the population will give you a sample. This is a part of the whole population. Statistical methods are then used on that sample.

The results of the statistical methods from the sample is used to make conclusions about the population.

Note: The word 'statistic' can also refer to specific bits of knowledge; like the average value of something.

## Important Concepts in Statistics

* Predictions and Explanations
* Populations and Samples
* Parameters and Sample Statistics
* Sampling Methods
* Data Types
* Measurement Level
* Descriptive Statistics
* Random Variables
* Univariate and Multivariate Statistics
* Probability Calculation
* Probability Distributions
* Statistical Inference
* Parameter Estimation
* Hypothesis Testing
* Correlation
* Regression Analysis
* Causal Inference

We will cover these topics step by step in this tutorial.

## Statistics and Programming

Statistical analysis is typically done with computers. Small amounts of data can analyzed reasonably well without computers.

Historically, all data analysis was performed by manually. It was time-consuming and prone to errors.

Nowadays, programming and software is typically used for data analysis.

In this course, we will show examples of code to do statistics with the programming languages Python and R.

## Gathering Data

Gathering data is the first step in statistical analysis.

Say for example that you want to know something about all the people in France.

The population is then all of the people in France.

It is too much effort to gather information about all of the members of a population (e.g. all 67+ million people living in France). It is often much easier to collect a smaller group of that population and analyze that. This is called a sample.

### A representative sample

The sample needs to be similar to the whole population of France. It should have the same characteristics as the population. If you only include people named Jacques living in Paris who are 48 years old, the sample will not be similar to the whole population.

So for a good sample, you will need people from all over France, with different ages, professions, and so on.

If the members of the sample have similar characteristics (like age, profession, etc.) to the whole population of France, we say that the sample is representative of the population.

A good representative sample is crucial for statistical methods.

Note: Data from a proper sample is often just as good data from the whole population, as long as it is representative!

A good sample allows you to make accurate conclusions about the whole population.

Describing data is typically the second step of statistical analysis after gathering data.

## Descriptive Statistics

The information (data) from your sample or population can be visualized with graphs or summarized by numbers. This will show key information in a simpler way than just looking at raw data. It can help us understand how the data is distributed.

Graphs can visually show the data distribution.

Examples of graphs include:

* [Histograms](https://www.w3schools.com/statistics/statistics_histograms.php)
* [Pie charts](https://www.w3schools.com/statistics/statistics_pie_charts.php)
* [Bar graphs](https://www.w3schools.com/statistics/statistics_bar_graphs.php)
* [Box plots](https://www.w3schools.com/statistics/statistics_box_plots.php)

Some graphs have a close connection to numerical summary statistics. Calculating those gives us the basis of these graphs.

For example, a box plot visually shows the quartiles of a data distribution.

Quartiles are the data split into four equal size parts, or quarters. A quartile is one type of summary statistics.

### Summary statistics

Summary statistics take a large amount of information and sums it up in a few key values.

Numbers are calculated from the data which also describe the shape of the distributions. These are individual 'statistics'.

Some important examples are:

* [Mean, median and mode](https://www.w3schools.com/statistics/statistics_average.php)
* [Range](https://www.w3schools.com/statistics/statistics_range.php) and [interquartile range](https://www.w3schools.com/statistics/statistics_interquartile_range.php)
* [Quartiles and percentiles](https://www.w3schools.com/statistics/statistics_quartiles_and_percentiles.php)
* [Standard deviation and variance](https://www.w3schools.com/statistics/statistics_standard_deviation.php)

Note: Descriptive statistics is often presented as a part of statistical analysis.

Descriptive statistics is also useful for guiding further analysis, giving insight into the data, and finding what is worth investigating more closely.

Using statistics to make conclusions about a population is called statistical inference.

## Statistical Inference

Statistics from the data in the sample is used to make conclusions about the whole population. This is a type of statistical inference.

Probability theory is used to calculate the certainty that those statistics also apply to the population.

When using a sample, there will always be some uncertainty about what the data looks like for the population.

Uncertainty is often expressed as confidence intervals.

Confidence intervals are numerical ways of showing how likely it is that the true value of this statistic is within a certain range for the population.

Hypothesis testing is a another way of checking if a statement about a population is true. More precisely, it checks how likely it is that a hypothesis is true is based on the sample data.

Some examples of statements or questions that can be checked with hypothesis testing:

* People in the Netherlands taller than people in Denmark
* Do people prefer Pepsi or Coke?
* Does a new medicine cure a disease?

Note: Confidence intervals and hypothesis testing are closely related and describe the same things in different ways. Both are widely used in science.

## Causal Inference

Causal inference is used to investigate if something causes another thing.

For example: Does rain make plants grow?

If we think two things are related we can investigate to see if they correlate. Statistics can be used to find out how strong this relation is.

Even if things are correlated, finding out of something is caused by other things can be difficult. It can be done with good experimental design or other special statistical techniques.

Note: Good experimental design is often difficult to achieve because of ethical concerns or other practical reasons.

https://www.w3schools.com/statistics/statistics\_making\_conclusions.php