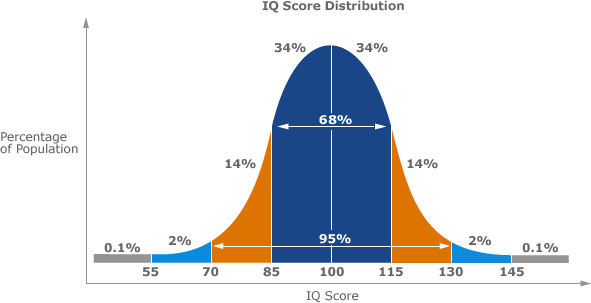
## **Statistics Overview**

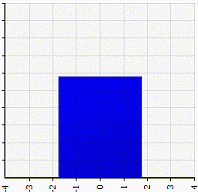
Statistics is the discipline that concerns the collection, organization, displaying, analysis, interpretation and presentation of data. In applying statistics to a scientific, industrial, or social problem, it is conventional to begin with a [statistical population](https://en.wikipedia.org/wiki/Statistical_population) or a [statistical model](https://en.wikipedia.org/wiki/Statistical_model) to be studied. Populations can be diverse groups of people or objects such as "all people living in a country" or "every atom composing a crystal". Statistics deals with every aspect of data, including the planning of data collection in terms of the design of [surveys](https://en.wikipedia.org/wiki/Statistical_survey) and [experiments](https://en.wikipedia.org/wiki/Experimental_design)

The most common basic statistics terms you’ll come across are the [mean, mode and median](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/). These are all what are known as “Measures of [Central Tendency](https://www.statisticshowto.datasciencecentral.com/central-tendency-2/).” Also important in this early chapter of statistics is the [shape of a distribution](https://www.statisticshowto.datasciencecentral.com/shapes-of-distributions/). This tells us something about how data is spread out around the [mean](https://www.statisticshowto.datasciencecentral.com/mean) or [median](https://www.statisticshowto.datasciencecentral.com/median). Perhaps the most common distribution you’ll see is the [**normal distribution**](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/normal-distributions/), sometimes called a bell curve. Heights, weights, and many other things found in nature tend to be shaped like this:



*IQ scores fit a bell curve shape.*

On the other end of the scale, you can also get a **flat distribution**. With this shape, the odds of anything happening are equal. For example, a [uniform distribution](https://www.statisticshowto.datasciencecentral.com/uniform-distribution/) can represent choosing a particular card from a standard deck; all the cards have a 1/52 chance of being chosen. Or tossing a coin, where you have a 50% chance of tossing a heads or a tails.

[](https://www.statisticshowto.datasciencecentral.com/wp-content/uploads/2014/02/shape_uniform.gif)

*A uniform distribution.*

**Type of Statistics:**

1. Descriptive Statistics
2. Inferential statistics

## [**Descriptive Statistics**](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/)

[Descriptive statistics](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/) are one of the fundamental “must know” with any set of data. It gives you a general idea of trends in your data including:

* The [mean, mode, median](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/) and [range](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/range-statistics/).
* [Variance](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/variance/)and [standard deviation](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/standard-deviation/).
* [Skewness](https://www.statisticshowto.datasciencecentral.com/skewness/).
* Count, maximum and minimum.

Descriptive statistics is useful because it allows you to take a large amount of data and summarize it. For example, let’s say you had data on the incomes of one million people. No one is going to want to read a million pieces of data; if they did, they wouldn’t be able to glean any useful information from it. On the other hand, if you summarize it, it becomes useful: an average wage, or a median income, is much easier to understand than reams of data.

## **1. Sub-Areas**

Descriptive statistics can be further broken down into several sub-areas, like:

* [Measures of central tendency.](#_Central_Tendency)
* [measures of dispersion](https://www.statisticshowto.datasciencecentral.com/dispersion/).
* Charts & graphs.
* [Shapes of Distributions.](https://www.statisticshowto.datasciencecentral.com/shapes-of-distributions/)

## 

## [**Central Tendency**](https://www.statisticshowto.datasciencecentral.com/central-tendency-2/)

Central tendency (sometimes called “measures of location,” “central location,” or just “centre”) is a way to describe what’s typical for a set of data. Central tendency doesn’t tell you specifics about the individual pieces of data, but it does give you an overall picture of what is going on in the entire data set. There are three major ways to show central tendency: [mean, mode and median](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/).

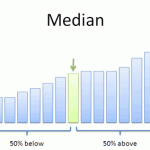
## **Central Tendency Measures**

**Mean**  
**The**[**mean**](https://www.statisticshowto.datasciencecentral.com/mean/)**is the**[**average**](https://www.statisticshowto.datasciencecentral.com/arithmetic-mean/)**of a set of numbers.** Add up all the numbers in a set of data and then divide by the number of items in the set. For example, the mean of 2 3 5 9 11 is:  
(2 + 3 + 5 + 9 + 11) / 5 = 30 / 5 = 6.

For more examples of finding the mean, see:  
[What is a mean?](https://www.statisticshowto.datasciencecentral.com/mean/)

**Median**  
**The**[**median**](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/#median)**is the middle of a set of numbers.** Think of it like the median in a road (that grassy area in the middle that separates traffic). Place your data in order, and the number in the exact center of a list is the median. For example:  
1 2 3 **4** 5 6 7  
The median is 4 because it’s in the center, with three numbers either side.

For more about the median, see:  
[What is a median?](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/#median)

[](https://www.statisticshowto.datasciencecentral.com/wp-content/uploads/2013/09/median.png)

**Mode**  
The mode is the most common number in a set of data. For example, the [mode](https://www.statisticshowto.datasciencecentral.com/mode/) of 1 2 2 3 5 6 is 2. Some data sets have no mode, like this one: 1 2 3 4 5 6. Others have multiple modes, like this one: 1 1 2 3 3.

For more on finding modes, see:  
[What is a Mode?](https://www.statisticshowto.datasciencecentral.com/mode/)

**Outliers**  
[Outliers](https://www.statisticshowto.datasciencecentral.com/find-outliers/)are extremely high or extremely low values. Outliers can affect central tendency, especially the mean. For example, if you got paid three weeks in a row but took vacation in the fourth week, your paychecks might be: $300 $300 $300 $0. Your four week mean would be ($300 + $300 + $300 + $0) / 4 = $900/4 = $225. That outlier of zero dollars brought your mean down very low.

## **2. Difference Between Descriptive and Inferential Statistics**

Statistics can be broken down into two areas:

* **Descriptive statistics:** describes and summarizes data. You are just describing what the data shows: a trend, a specific feature, or a certain [statistic](https://www.statisticshowto.datasciencecentral.com/statistic/)(like a [mean](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/#mean)or median).
* [**Inferential statistics**](https://www.statisticshowto.datasciencecentral.com/inferential-statistics/): uses statistics to make predictions.

Descriptive statistics just describes data. For example, descriptive statistics about a college could include: the [average](https://www.statisticshowto.datasciencecentral.com/arithmetic-mean/)SAT score for incoming freshmen; the [median](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/statistics-definitions/mean-median-mode/#median)income of parents; racial makeup of the student body. It says nothing about why the data might exist, or what trends you might be able to see from the data. When you take your data and start to make predictions about future behaviour or trends, that’s inferential statistics. Inferential statistics also allows you to take [sample](https://www.statisticshowto.datasciencecentral.com/sample/)data (e.g. from one university) and apply it to a larger [population](https://www.statisticshowto.datasciencecentral.com/what-is-a-population/) (e.g. all universities in the country).

## **3. Excel Descriptive Statistics**

Using the descriptive statistics feature in Excel means that you won’t have to type in individual functions like MEAN or MODE. One button click will return a dozen different stats for your data set. If you want to calculate Excel descriptive statistics, you must have the Data Analysis Toolpak loaded in Excel. Click the “Data” tab in Excel. If you don’t see “Data analysis” on the right of the toolbar, you need to load the Toolpak first. See: [Load the Excel Data Analysis Toolpak.](https://www.statisticshowto.datasciencecentral.com/excel-data-analysis-toolpak/)

### **How to Calculate Excel Descriptive Statistics: Steps**

[Watch the video](https://www.youtube.com/watch?v=ZdcoTVYJNF4) or read the steps below:

Step 1:**Type your data into Excel,** in a single column. For example, if you have ten items in your data set, type them into cells A1 through A10.

Step 2:**Click the “Data” tab**and then click “Data Analysis” in the Analysis group.

Step 3:**Highlight “Descriptive Statistics”**in the pop-up Data Analysis window.

Step 4:**Type an input range into the “Input Range” text box.**For this example, type “A1:A10” into the box.

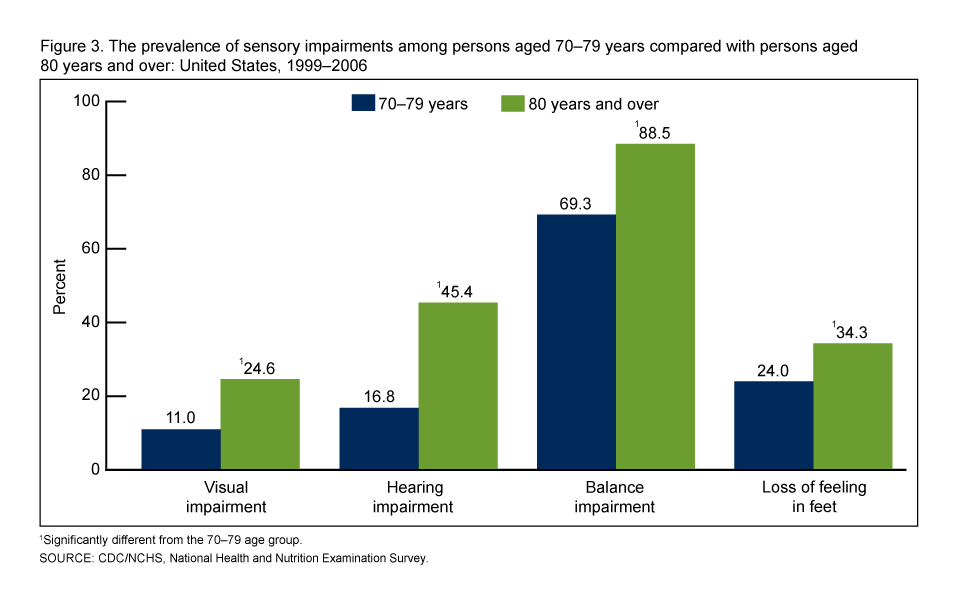
Step 5:**Check the “Labels in first row” check box**if you have titled the column in row 1, otherwise leave the box unchecked.

Step 6:**Type a cell location into the “Output Range” box.** For example, type “C1.” Make sure that two adjacent columns do not have data in them.

Step 7:**Click the “Summary Statistics” check box and then click “OK”**to display Excel descriptive statistics. A list of descriptive statistics will be returned in the column you selected as the Output Range.

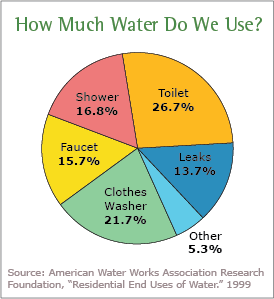
## **4. Descriptive Statistics: Charts, Graphs and Plots**

There are literally dozens of charts and graphs you can make from data. which one you choose depends upon what kind of data you have and what you want to display. For example, if you wanted to display relationships between data in categories, you could make a [bar graph.](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/bar-chart-bar-graph-examples/)



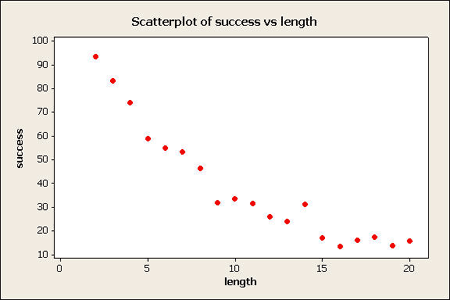
*Grouped bar graph. Image: CDC.*

A [pie chart](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/pie-chart/) would show you how categories in your data relate to the whole set.

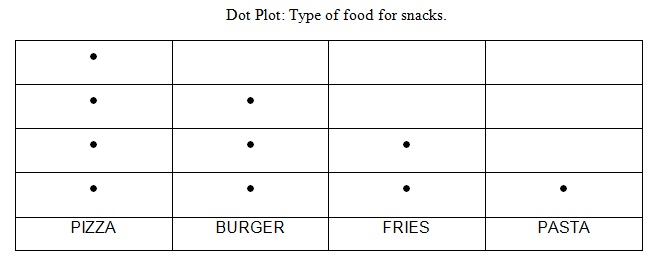


*Pie chart showing water consumption. Image courtesy of EPA.*

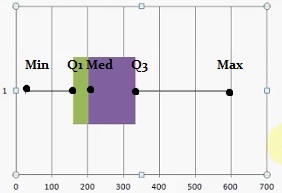
[Scatter plots](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/scatter-plot-chart/#definition) are a good way to display data points.

 *Image: Penn State*

Less common, but useful in some cases, include [dot plots](https://www.statisticshowto.datasciencecentral.com/what-is-a-dot-plot/) and [box and whisker charts](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/box-plot/#definition):

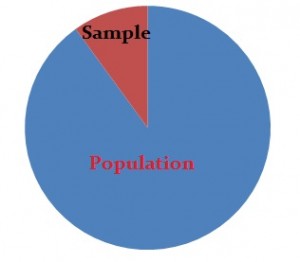
[](https://www.statisticshowto.datasciencecentral.com/wp-content/uploads/2013/10/dot-plot-2.jpg)

*Simple dot plot showing the types of foods a group of friends eats.*

[](https://www.statisticshowto.datasciencecentral.com/wp-content/uploads/2013/11/box-and-whiskers-graph-.jpg)

*Box and whiskers graph*

## **Inferential Statistics**



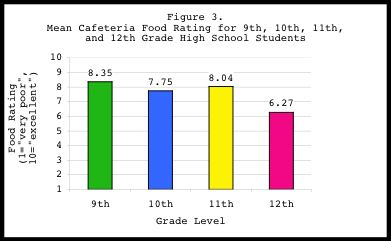
[Descriptive statistics](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/) describes data (for example, a chart or graph) and **inferential statistics** allows you to make predictions (“inferences”) from that data. With inferential statistics, you take data from [samples](https://www.statisticshowto.datasciencecentral.com/sample/)and make generalizations about a [population](https://www.statisticshowto.datasciencecentral.com/what-is-a-population/). For example, you might stand in a mall and ask a sample of 100 people if they like shopping at [Sears](http://www.sears.com/). You could make a [bar chart](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/bar-chart-bar-graph-examples/) of yes or no answers (that would be [descriptive statistics](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/)) or you could use your research (and inferential statistics) to reason that around 75-80% of the population (**all**shoppers in **all malls**) like shopping at Sears.

There are two main areas of inferential statistics:

1. **Estimating parameters**. This means taking a [statistic](https://www.statisticshowto.datasciencecentral.com/statistic/)from your sample data (for example the [sample mean](https://www.statisticshowto.datasciencecentral.com/sample-mean/)) and using it to say something about a population parameter (i.e. the population mean).
2. [**Hypothesis tests**](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/). This is where you can use sample data to answer research questions. For example, you might be interested in knowing if a new cancer drug is effective. Or if breakfast helps children perform better in schools.

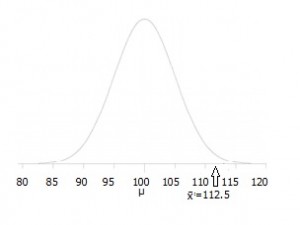
Let’s say you have some sample data about a potential new cancer drug. You could use descriptive statistics to describe your sample, including:

* Sample [mean](https://www.statisticshowto.datasciencecentral.com/mean/)
* Sample [standard deviation](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/standard-deviation/)
* Making a [bar chart](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/bar-chart-bar-graph-examples/) or [boxplot](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/descriptive-statistics/box-plot/)
* Describing the shape of the sample [probability distribution](https://www.statisticshowto.datasciencecentral.com/probability-distribution/)



*A bar graph is one way to summarize data in descriptive statistics. Source: NIH.GOV.*

With inferential statistics you take that sample data from a small number of people and try to determine if the data can predict whether the drug will work for everyone (i.e. the population). There are various ways you can do this, from calculating a [z-score](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/z-score/) (z-scores are a way to show where your data would lie in a [normal distribution](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/normal-distributions/) to [post-hoc](https://www.statisticshowto.datasciencecentral.com/post-hoc/) (advanced) testing.

[](https://www.statisticshowto.datasciencecentral.com/wp-content/uploads/2014/10/hypothesis-testing-example.jpg)

*A hypothesis test can show where your data is placed on a distribution like this one.*

Inferential statistics use statistical models to help you compare your sample data to other samples or to previous research. Most research uses statistical models called the Generalized Linear model and include [Student’s t-tests](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/t-test/), [ANOVA (Analysis of Variance](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/anova/)), [regression](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/regression-analysis/)analysis and various other models that result in straight-line (“linear”) probabilities and results.

References:

1. <https://www.statisticshowto.datasciencecentral.com/statistics-basics/>
2. <https://stattrek.com/tutorials/ap-statistics-tutorial.aspx>