



## CHAPTER 1

---

# Basic Statistics in Spreadsheets

When you completed the problems in the last section, you probably noticed how long it took to compute statistics like the mean, the median, and variance by hand. Luckily, computers were designed to free us from these sorts of tedious tasks. The most basic tool for automating calculations is the spreadsheet program.

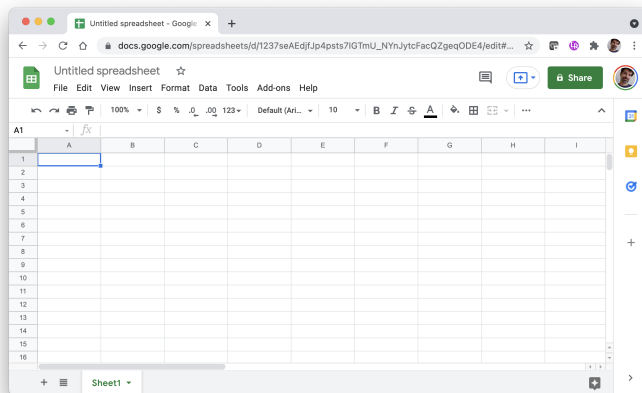
There are lots of spreadsheet programs including Microsoft's Excel and Apple's Numbers. Any spreadsheet program will work; they are all very similar. The instructions and screenshots here will be from Google Sheets – a free spreadsheet program you use through your web browser.

### 1.1 Your First Spreadsheet

In whatever spreadsheet program you are using, create a new spreadsheet document.

A spreadsheet is essentially a grid of cells. In each cell you can put data (like numbers or text) and formulas.

## 2 Chapter 1. BASIC STATISTICS IN SPREADSHEETS



Let's put some labels in the column:

- Select the first cell (A1) and type "A number".
- Select the cell below it (A2) and type "Another number".
- Select the cell below that one (A3) and type "Their product".
- In the next column, type the number 5 in B1 and 7 in B2.

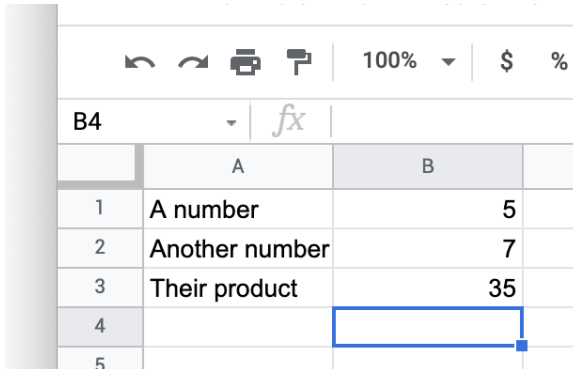
It should look like this:

	A	B
1	A number	5
2	Another number	7
3	Their product	
4		

Now put a formula in cell B3. Select B3, and type " $= B1 + B2$ ". The spreadsheet knows this is a formula because it starts with '='. It will look like this as you type:

	A	B
1	A number	5
2	Another number	7
3	Their product	$= B1 * B2$
4		

When you press Return or Tab, the spreadsheet will remember the formula, but display its value:



	A	B
1	A number	5
2	Another number	7
3	Their product	35
4		
5		

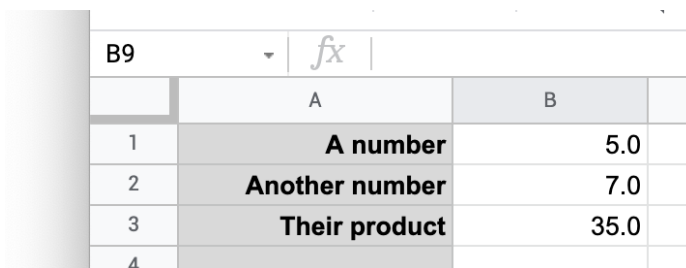
If you change the values of cell B1 or B2, the cell B3 will automatically be recalculated. Try it.

## 1.2 Formatting

Every spreadsheet lets you change the formatting of your columns and cells. They are all a little different, so play with your spreadsheet a little now. Try to do the following:

- Set the background of the first column to light gray.
- Right-justify the text in the first column.
- Make the text in the first column bold.
- Make the numbers in the second column have one digit after the decimal point.

It should look something like this:



	A	B
1	<b>A number</b>	5.0
2	<b>Another number</b>	7.0
3	<b>Their product</b>	35.0
4		

That's a spreadsheet. You have a grid of cells. Each cell can hold a value or a formula that uses values from other cells. The cells with formulas automatically update as you edit the values in the other cells.

### 1.3 Comma-Separated Values

A lot of data is exchanged in a file format called *Comma-Separated Values* or just CSV. Each CSV file holds one table of data. It is a text file, and each line of text corresponds to one row of data in the table. The data in each column is separated by a comma. The first line of a CSV is usually the names of the columns. A CSV might look like this:

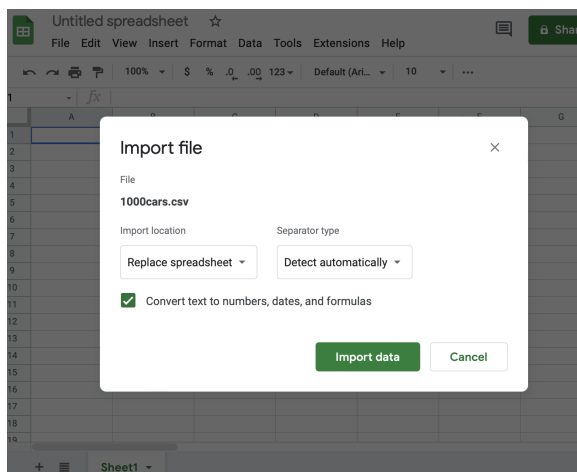
```
studentID,firstName,lastName,height,weight
1,Marvin,Sumner,260,45.3
2,Lucy,Harris,242,42.2
3,James,Boyd,261,44.2
```

In your digital resources for this module, you should have a file called `1000cars.csv`. It is a CSV with only one column called “speed”. The first few lines look like this:

```
speed
33.8000
29.9920
34.8699
27.9936
```

There is a title line and 1000 data lines.

Import this CSV into your spreadsheet program. In Google Sheets, it looks like this:



You should see a long, long column of data appear. (Mine goes from cell A2 through A1001.)

Untitled spreadsheet ☆ 📄 ☁

File Edit View Insert Format Data Tools Extensions Help

100% \$ % .0 .00 123 Default (Ari... 1

	A	B	C
1	speed		
2	33.8		
3	29.992		
4	34.8699		
5	27.9936		
6	26.2875		
7	31.6701		
8	27.3347		

## 1.4 Statistics in Spreadsheets

Let's take the mean of all 1000 numbers. In cell B2, type in a label: "Mean". (Feel free to format your labels as you wish. Bolding is recommended.)

In cell C2, enter the formula `"=AVERAGE(A2:A1001)"`. When you press return, the cell will show the mean: 31.70441, if done correctly.

	A	B	C
1	speed		
2	33.8	<b>Mean</b>	31.7044106
3	29.992		
4	34.8699		
5	27.9936		

Notice that by specifying that the function AVERAGE was to be performed on a range of cells: cells A2 through A1001.

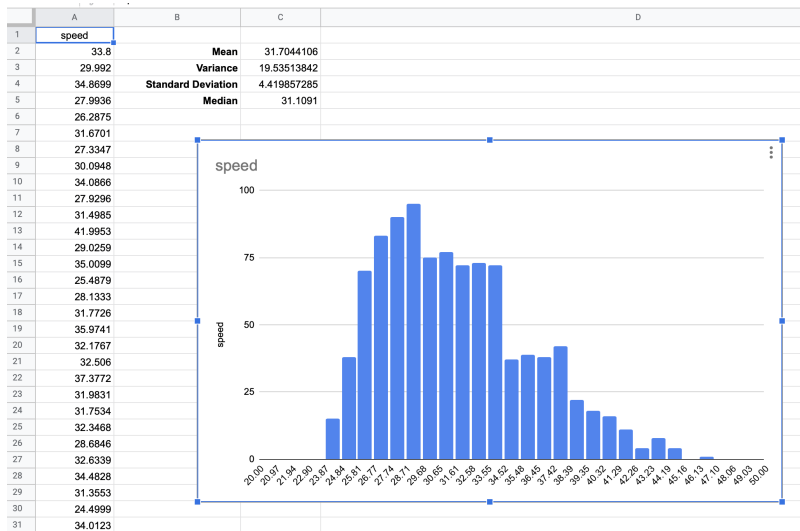
Do the calculations for variance, standard deviation, and median.

- The function for variance is VAR.
- The function for standard deviation is STDEV.
- The function for median is MEDIAN.

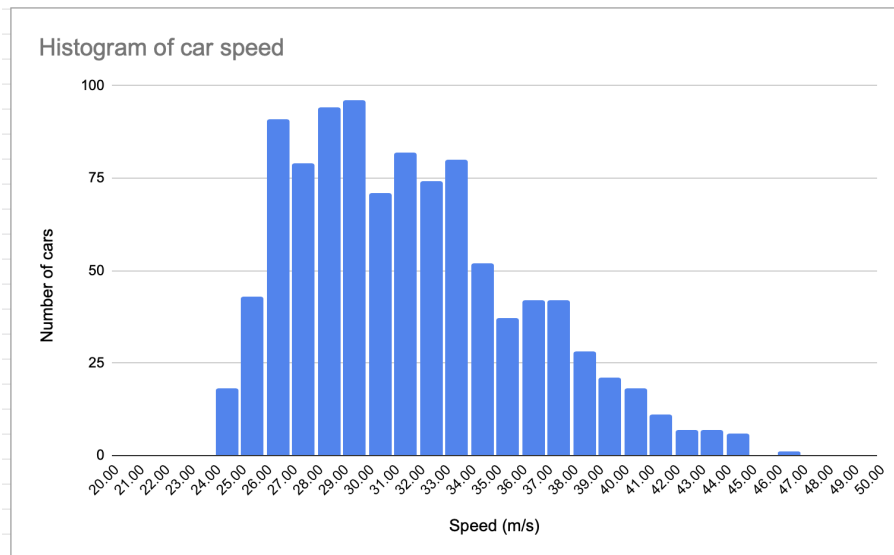
	A	B	C
1	speed		
2	33.8	<b>Mean</b>	31.7044106
3	29.992	<b>Variance</b>	19.53513842
4	34.8699	<b>Standard Deviation</b>	4.419857285
5	27.9936	<b>Median</b>	31.1091
6	26.2875		
7	31.6701		

## 1.5 Histogram

Most spreadsheets have the ability to create a histogram. In Google Sheets, you select the entire range A2:A1001 by selecting the first cell and then shift-clicking the last. Then you choose Insert→Chart. In the inspector, change the type of the chart to a histogram. This will get you a basic histogram.



Play with the formatting to see how unique you can make data. Here is an example:



**Exercise 1      RMS**

In your spreadsheet, calculate the quadratic mean (the root-mean-squared) of the speeds. You will need the following three functions:

- SUMSQ returns the sum of the squares of a range of cells.
- COUNT returns the number of cells in a range that contains numbers.
- SQRT returns the square root of a number.

*Working Space*

*Answer on Page ??*







## APPENDIX A

---

# Answers to Exercises

### **Answer to Exercise ?? (on page ??)**

The formula for the RMS is “=SQRT(SUMSQ(A2:A1001)/COUNT(A2:A1001))”.





---

# INDEX

Spreadsheet  
    Entering formula, [2](#)  
spreadsheet, [1](#)