**Lesson Three**

# Basic Statistical Analysis with Excel

Excel is an extremely versatile tool for statistical analysis. With a set of data, the user can readily calculate such descriptive statistics as the **mean**, the **standard deviation**, the **median**, and the **range**. We show here how Excel can be used to find these basic statistical quantities and to obtain confidence intervals for the means of data sets.

## Calculating a Mean with Excel

In this example, we learn to calculate the mean of a set of data. First, we define formulas to calculate the mean, and then we use the built-in functions of Excel to accomplish the task. The data used are from six replicate determinations of iron(III) concentration in a given sample. The results ( values) were 19.4, 19.5, 19.6, 19.8, 20.1, and 20.3 ppm iron(III).

### Entering the Data

Let’s begin by starting Excel with a clean spreadsheet. In cell B1 enter the heading **Data[↵]** and make it boldface. Now enter in column B under the heading the results given above. Click on cell A11, and type

**Total[↵]**

**N[↵]**

**Mean[↵]**

Make these labels boldface. Your worksheet should now look like that shown in Figure 3-1.

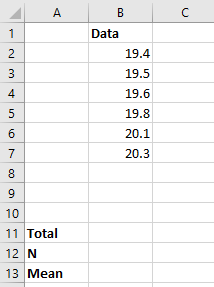


Figure 3-1 Data entry for statistical calculations.

### Finding the Mean

The mean is calculated according to the equation



Click on cell B11, and type

=SUM(B2:B7) **[↵]**

This formula calculates the sum of the values in cells B2 through B7 and displays the result in cell B11. Now, we will obtain the number of data points N in cell B12 by typing

=COUNT(B2:B7)**[↵]**

The COUNT function counts the number of nonzero cells in the range B2:B7 and displays the result in cell B12. Since we have found the sum of the values and the number N of data points, we can find the mean of the data set by typing the following formula in cell B13.

=B11/B12**[↵]**

At this point in the exercise, your worksheet should appear as shown in Figure 3-2.

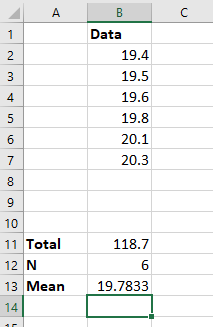


Figure 3-2 Spreadsheet after computing the sum, count and mean values.

Note that the mean contains many more figures after the decimal point than any of the data values. It is appropriate to round the mean to 19.8 either by invoking the **Format Cells** window or by clicking on the Decrease Decimal button the Number group on the Home ribbon.

### Using Excel's Built-in Functions

Excel has built-in functions to compute many important and useful quantities. Now we shall see how to use them to calculate the mean, or in Excel's syntax, the average. Click on cell C13 and type

=AVERAGE(B2:B7)[↵]

Notice that the mean determined using the built-in **AVERAGE** function is identical to the value in cell B13 that you determined by typing a formula. Again format the result to contain only 1 figure after the decimal point.

### Finding the Deviations from the Mean

We can now use Excel to determine the deviations of each of the data values from the mean using the definition



Click on cell C1 and type the boldface label

**Deviation[↵]**

With cell C2 active, type

=ABS(B2-$B$13)[↵]

This formula computes the absolute value **ABS()** of the difference between our first value in B2 and the mean value in B13. Note that this formula is a bit different than those that we have used previously. We have typed a dollar sign, $, before the B and before the 13 in the second cell reference. This type of cell reference is called an absolute reference. This syntax means that no matter where we might copy the contents of the cell C2, the reference will always be to cell B13. The other type of cell reference that we consider here is the relative reference, which is exemplified by B2. The reason that we use a relative reference for B2 and an absolute reference for B13 is that we want to copy the formula in C2 into cells C3-C7, and we want the mean referred to by **$B$13** to be subtracted from each of the successive values in column B. Let’s now copy the formula by clicking on cell C2, clicking on the fill handle, and dragging the rectangle through C7. When you release the mouse button, your worksheet should look like that shown in Figure 3-3.

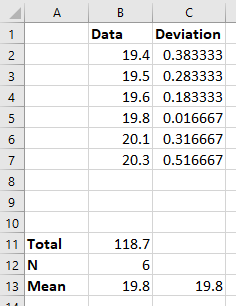


Figure 3-3 Worksheet with deviations computed.

Now click on cell C3, and notice that it contains the formula **=ABS(B3-$B$13)**. Compare this formula with the one in cell C2 and C4-C7. The absolute cell reference $B$13appears in all of the cells. As you can see, we have accomplished our task of calculating the deviation from the mean for all of the data. Now we will edit the formula in cell C13 to find the mean deviation of the data.

### Editing Formulas

To edit the formula to calculate the mean deviation of the data, click on C13, and then click on the formula in the formula bar. Use the arrow keys, **[←]** and **[→]**, and either the **[Backspace]** or the **[Delete]** key to replace both Bs in the formula with Cs so that it reads: **=AVERAGE(C2:C7)**. Finally, type **[↵]**, and the mean deviation will appear in cell C13. Your worksheet should appear as in Figure 3-4.

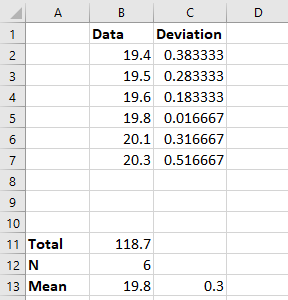
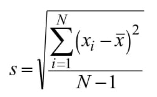


Figure 3-4 Worksheet showing deviations and mean deviation.

## Computing the Standard Deviation

In this exercise, we will calculate the sample standard deviation, the sample variance, and the relative standard deviation of two sets of data. We begin with the spreadsheet and data from the previous example. The sample standard deviation is given by the equation



and the variance is ***s*2**.

### Finding the Variance

Begin with the worksheet shown in Figure 3-4 on your computer screen. Make cell D1 the active cell, and type the label

**Deviation^2[↵]**

Make the label boldface. With cell D2 the active cell, type

=C2^2[↵]

and the square of the deviation in cell C2 appears in D2. Copy this formula into the other cells in column D by once again clicking on cell D2, clicking on the fill handle, and dragging the fill handle through cell D7. You have now calculated the squares of the deviations of each of the data from the mean value in cell B13.

### A Shortcut for Performing a Summation

To find the variance, we must first find the sum of the squares of the deviations, so now click on cell D11, and then click on the AutoSum icon in the Editing group on the Home ribbon. The worksheet should now appear as shown in Figure 3-5

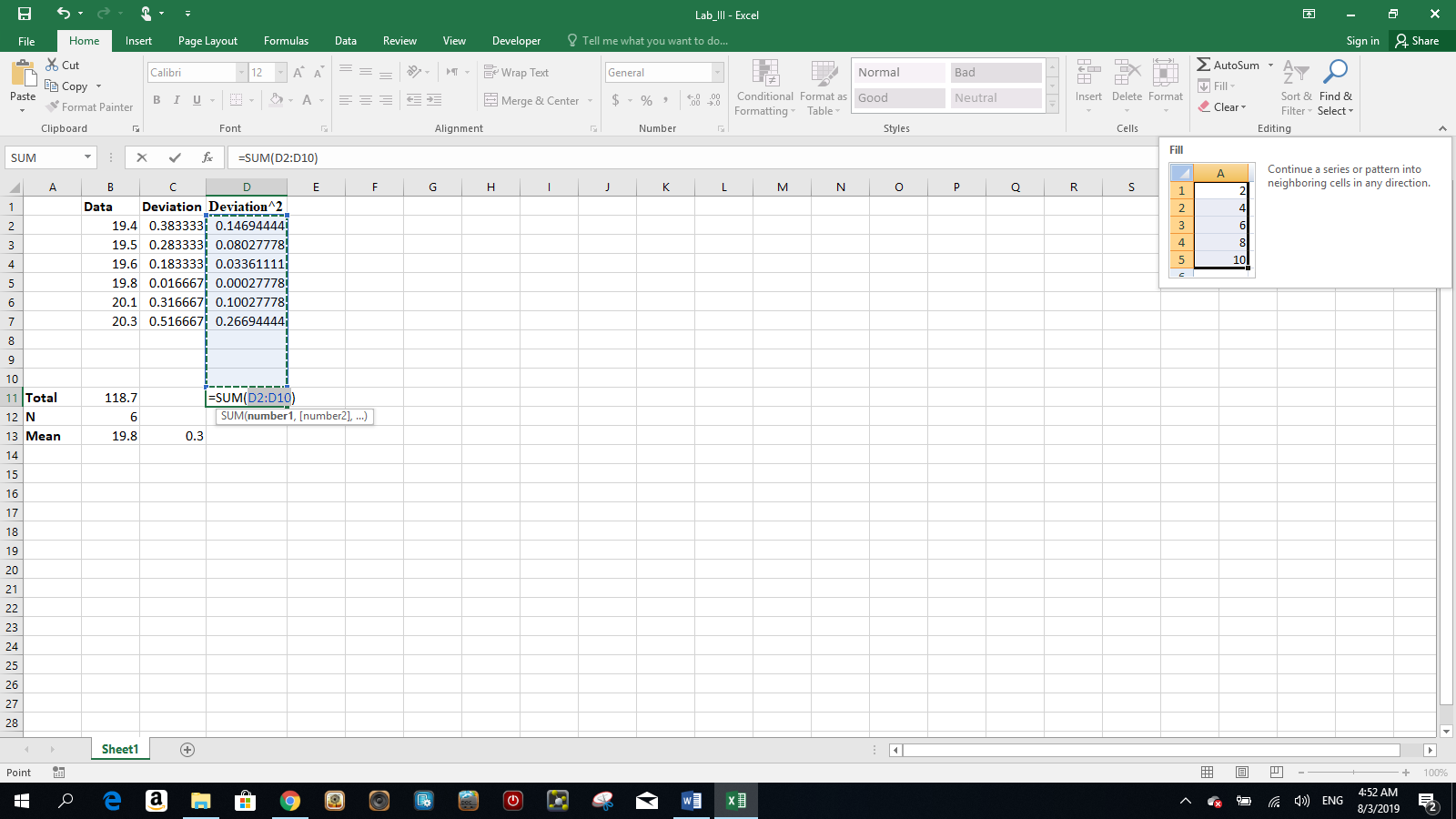


Figure 3-5 Worksheet with AutoSum function selected.

The dashed box shown above now surrounds the column of data in cells D2-D10, which appear as arguments of the SUM function in cell D11 and in the formula bar. Note that Excel assumes that you want to add all of the numerical data above the active cell and automatically completes the formula. When you type **[↵]**, the sum of the squares of the deviations appears in cell D11. Since cells D8 through D10 are blank, they contribute nothing to the sum, and so there is no harm in leaving the references to blank cells in the formula. Be aware, however, that references to blank cells could pose difficulty under certain circumstances. The final step in calculating the variance is to divide the sum of the squares of the deviations by the number of degrees of freedom, which is *N – 1*. We’ll type the formula for carrying out this last calculation in cell D12. Before proceeding, type the label **Variance** in F12 in boldface font. Now click on D12, and type

=D11/(B12-1)[↵]

Excel calculates the variance and it appears in the cell. Notice that you must enclose the difference B12 – 1 in parentheses so that Excel computes the number of degrees of freedom before the division is carried out. If we had not enclosed the number of degrees of freedom, B12 – 1, in parentheses, Excel would have divided D11 by B12 and then subtracted one, which is incorrect. To illustrate this point, suppose D11 = 12 and B12 = 3. If we leave off the parentheses, D11/B12 –1 = 3, but if we put them in, D11/(B12 – 1) = 6. The order of mathematical operations in Excel is extremely important. Remember that just as in algebra, Excel performs exponentiation before multiplication and division, and it performs multiplication and division before addition and subtraction. As in the present example, we can change the order of operations by properly placing parentheses.

### Finding the Standard Deviation

Our next step is to calculate the standard deviation by extracting the square root of the variance. Click on D13, and type

=SQRT(D12)[↵]

Then click on F13, and type in bold

**Standard Devotion[↵]**

Your worksheet should then appear similar to Figure 3-6.

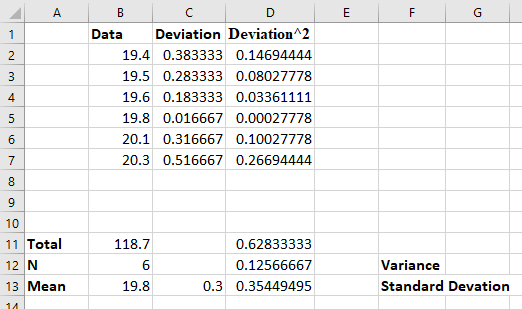


Figure 3-6 Worksheet after computing variance and standard deviation.

Notice that we have deliberately left cells E12 and E13 blank. We will now use the built-in variance and standard deviation functions of Excel to check our formulas.

### The Built-in Statistical Functions of Excel

Click on cell E12, and then type

=VAR(

Now click in cell B2, and drag the mouse into cell B7 so that the worksheet appears as shown in Figure 3-7.

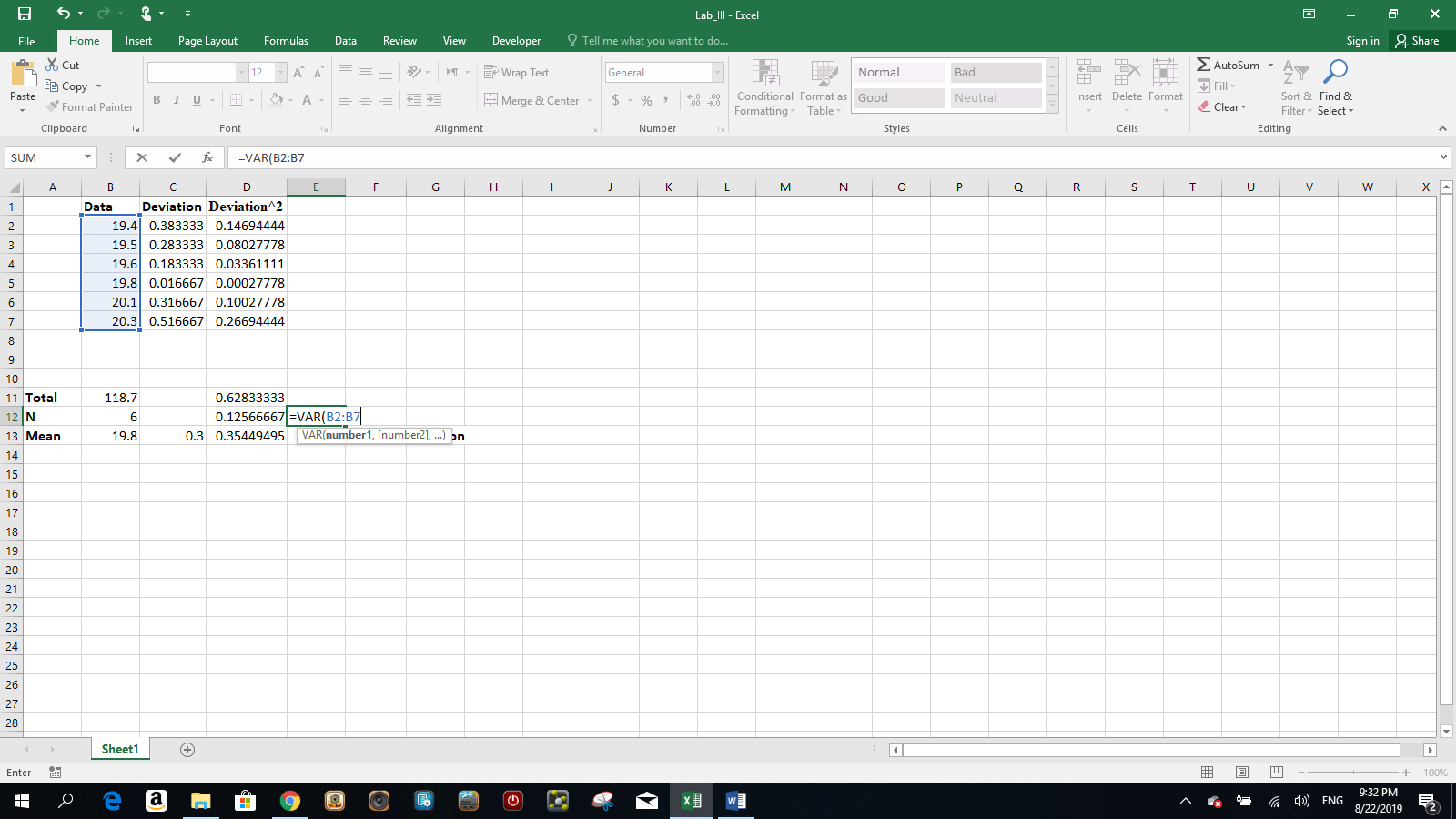


Figure 3-7 Defining the range for the variance (VAR) function.

Notice that the cell references B2:B7 appear in cell E12 and in the formula bar. Now, let up on the mouse button, and type [↵], and the variance appears in cell E12. If you have performed these operations correctly, the values displayed in cells D12 and E12 are identical. The active cell should now be E13. If it is not, click on it, and type

**=**STDEV(

and click and drag to highlight cells B2:B7 as you did previously. Let up on the mouse button, type [↵], and the standard deviation appears in cell E13. The computed values in cells D13 and E13 should be equal although they may show different numbers of decimal places. It is important to note that the Excel **STDEV** and **VAR** functions calculate the sample standard deviation and the sample variance, not the corresponding population statistics. These built-in functions are quite convenient since your sample will generally be sufficiently small that you will want to calculate sample statistics rather than population statistics. Excel 2016 has the functions VAR,P, VAR,S, STDEV,P and STDEV,S for the population and sample variances and standard deviations, respectively. Now decrease the number of decimal places in D11, D12, D13, E12, E13 cells until only one significant figure is displayed.

### Percent Relative Standard Deviation (%RSD)

Click in cell E14, and type

=E13\*100/B13[↵]

Then click on cell F14 and type **%RSD** in bold.

Note that we have multiplied the ratio of E13 to B13 by 100 so that the relative standard deviation is expressed as a percentage. Move the decimal point to indicate only two significant figures in the %RSD.

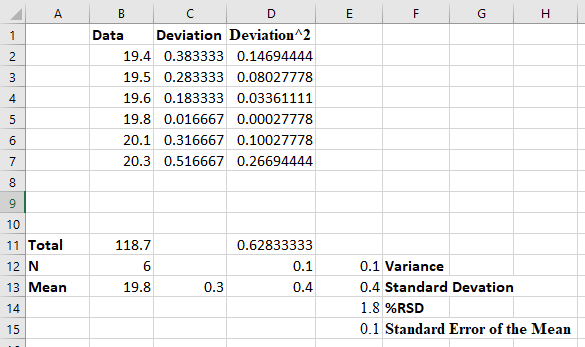
### Standard Error of the Mean



Click in cell F15 and type the label **Standard Error of the Mean[↵]**in bold. In cell E15, type

=E13/SQRT(B12)[↵]

The standard error of the mean should now appear in cell E15. Format this result to contain 1 digit after the decimal place. Your worksheet should now look like that shown in Figure 3-8.

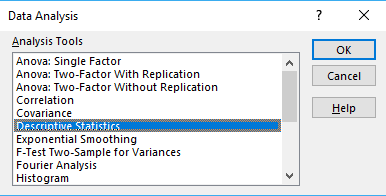


**Figure 3-8** Final worksheet for computing basic statistics.

We have now constructed a general-purpose spreadsheet that you may use to make basic statistical calculations. You can use the spreadsheet for other data sets by deleting or overtyping the data in cells B2:B7. To clear the data from your worksheet, just click and drag to highlight cells B2:B7, and strike **[Delete]** or choose Clear All from the Editing group on the Home ribbon. Alternatively, you may simply click on B2, and begin typing the data. Note that room is provided for 9 data points in the worksheet. If you need more rows, select cell A10 and then click on the Insert command in the Cells group on the Home ribbon. From the pull-down menu, select Insert Sheet Rows. You will have to extend any formulas from row 7 to include any added values.

## Computing Descriptive Statistics with the Analysis ToolPak

There is yet another way to compute the mean, the standard deviation and other statistical quantities. This method takes advantage of the Descriptive Statistics package in Excel’s Analysis ToolPak. The advantage of this method is that the formulas and calculations are done automatically. The Analysis ToolPak On the Data ribbon, select Data Analysis in the Analysis group or else you can add it form File 🡪 Options 🡪 add-Ins 🡪 Manage Excel add-Ins 🡪 Go 🡪 Analysis ToolPak 🡪 OK. The window shown below should appear if the ToolPak has been installed and enabled. Select Descriptive Statistics and click OK.



With the Descriptive Statistics window open, click and drag to highlight cells B2:B7 as shown in Figure 3-9. Note that these cells appear in the Input Range: box.

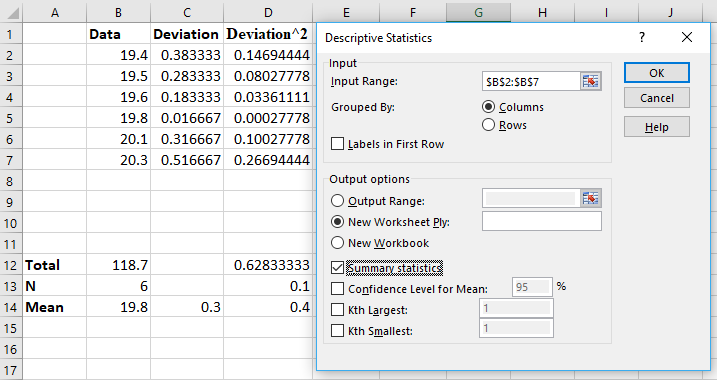


Figure 3-9 Worksheet with Descriptive Statistics Window

Click on New Worksheet Ply, select Summary Statistics and click on OK. This action inserts the statistical results into a new worksheet. You will have to increase the width of column A to read all the labels. The statistical results should be as shown in Figure 3-10. Note that the results are identical to those calculated previously. If desired, you could format the cells with the appropriate number of significant figures.

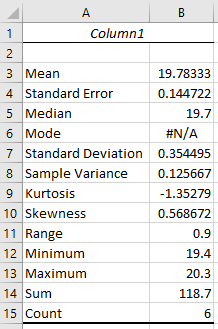


Figure 3-10 Descriptive Statistics.

Note that Excel produces some additional statistics such as the median, the range, and the minimum and maximum values. Excel also produces some measures of the symmetry of the data distribution called the kurtosis and the skewness. The skewness is negative for a distribution that tails to the left and positive for one that tails to the right. Kurtosis is a measure of the amount of peakedness in the distribution. The detailed descriptions of kurtosis and skewness are beyond the scope of this discussion.