Introduction to Statistics: Excel Lab Manual

Bianca Sosnovski

05/18/2022

Contents

Preface

Statistics is present in many ways in our lives. Statistical methodology can be found in surveys, sampling, clinical trials, studies of biomedical treatments, digital marketing, finance, etc. In recent years, Statistics has undergone changes in its techniques and approaches because of the need to analyze exceptionally large and complex data sets that arise all around us but cannot be done by hand. Since we have more powerful computers available to us at the present time, we can employ them to analyze these large data sets. Thus, it is even more important to learn how to use computer applications for statistical analysis, enhancing technology skills effectively for academic research and lifelong learning.

About this Manual

This manual has been written as a companion text for an introductory statistics course using Excel software. Topics include descriptive statistics; probability; binomial and normal distributions; sampling; confidence intervals and tests of hypotheses.

This text is to be used as a tutorial on how to use Excel to learn some fundamentals concepts covered in a typical Introduction to Statistics course and as a tool to advance technology skills effectively for academic research and lifelong learning. It is to be used alongside lectures and/or textbooks in such a course. This text is not a comprehensive introduction to Excel or Statistics.

The mathematical equations required for statistical concepts such as the mean and the standard deviation have not been provided as these are easily performed with functions in Excel.

About the Author

Bianca Sosnovski (**https://www.qcc.cuny.edu/mathcs/people/BSosnovski.html**) is an Assistant Professor for the Department of Mathematics and Computer Science at Queensborough Community College, CUNY. (**https://**

www.qcc.cuny.edu). She earned her Ph.D. and M.Phil. in Mathematics from the Department of Mathematics at the Graduate Center, CUNY in New York, and her M.S. in Applied Mathematics from the Department of Mathematics at the State University of Campinas (UNICAMP) in Brazil. In addition to her interest in mathematics education, she is also interested in abstract algebra and cryptography.

She also shares some other resources on GitHub (**https://github.com/bsosnovski**). In her free time, she enjoys photography and traveling.

License

Figure 1: Creative Commons License

Disclaimer

Introduction

Microsoft Excel® is a spreadsheet software for business applications available in the Microsoft Office® package. Excel enables users to format, organize, and calculate data with formulas and functions.

This basic Excel lab manual provides students and the novice with the skills needed to use Excel's built-in statistical functions and how to produce some useful graphs. It also prepares them for more advanced skills using the Excel software.

Although Excel is a powerful tool for executing many personal and professional activities, it is not a statistical data analysis package. It was never designed to be one, but it has the *Data Analysis ToolPak*, an extra feature extending Excel's standard capabilities. Data Analysis ToolPak enables users to perform various statistical analyses in Excel. In many cases, macros (automated actions that can be used repeatedly to perform a task or tasks) can be created to produce special graphs, such as box-and-whisker plots. However, this manual only shows how to use some of the existing built-in features and the Data Analysis ToolPak.

The Analysis ToolPak is part of Excel and can be added to the basic Excel program as an option on the program's list of Add-ins. The details for getting started with the Analysis ToolPak are presented in Chapter **1** of this manual. No additional software is required to use the Excel functions described.

Excel runs slightly differently on MACs and PCs. Whenever there is a difference between Excel for Mac and PC on performing a task, each system's instructions in this lab will be provided separately. Otherwise, the same set of instructions is applied to both systems.

Chapter 1

Getting Started with Excel

City University of New York (CUNY) Students can access Excel via Microsoft 365. On most campuses, students can also remotely access a computer lab at their specific campus.

If obtaining Excel via Microsoft 365, download the software to your machine. Do not run Excel from the cloud. Running it from the cloud does not give students access to the *Analysis ToolPak* (see *Section **1.4***).

1.1 The Excel Window

Note that in the Excel Windows shown in Figures **1.1** and **1.2**, the title bar shows `Book1` as a temporary name for the file. When opening a new Excel file (also called a Workbook, see *Section **1.6***), it has a suggested temporary name, `Book1`, `Book2`, etc, depending on how many new files are open before saving them. When saying it for the first, Excel will prompt the user to rename it.

1.1.1 The Excel Window on a PC

1.1.2 The Excel Window on a MAC

1.2 Saving Excel Files

On the main menu, click `File > Save As` when saving for the first time. A dialog will appear, similar to the one in figure **1.3** on a PC or figure **1.4** on a MAC. Fill in the file name, select the directory of your choice where the file will be saved on your machine, and the file type, for example, file with extensions `.xlsx` or `.xls`.

After using the `Save As` feature, you can use the `Save` command to overwrite your previously saved file with the file in its current state. After saving the Excel

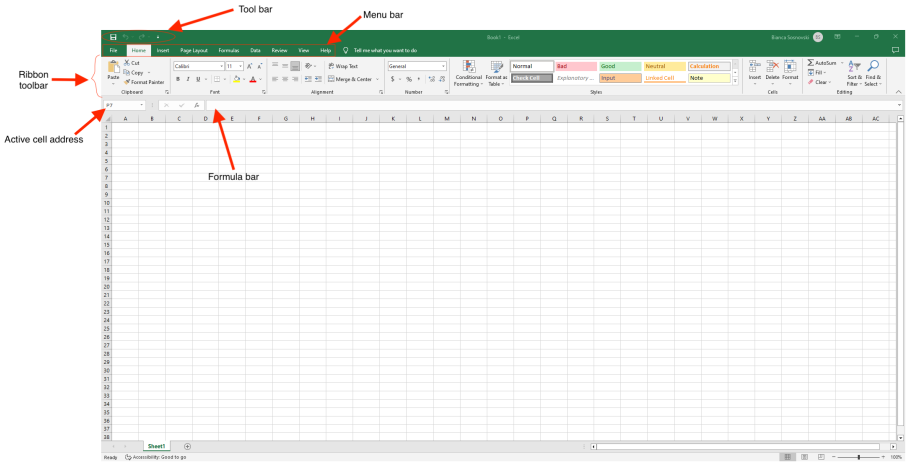


Figure 1.1: Excel Window on a PC.

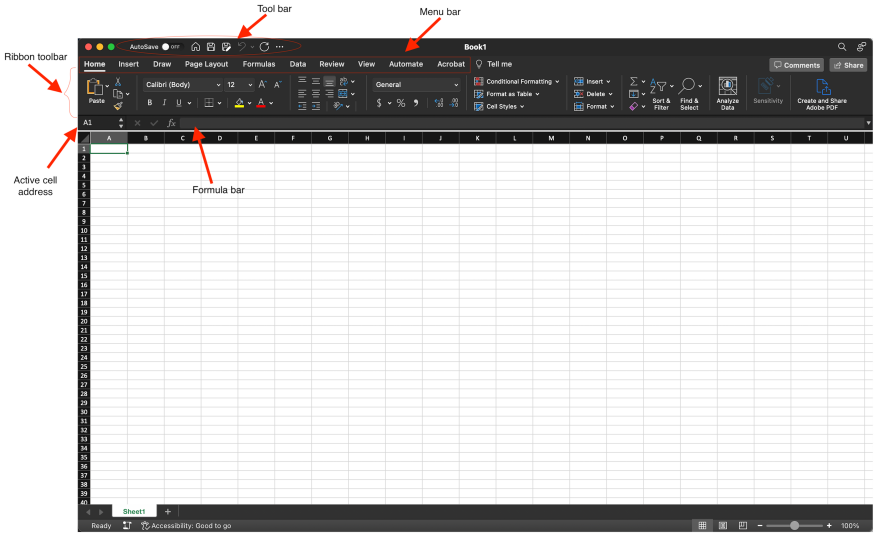


Figure 1.2: Excel Window on a MAC (in Dark Mode).

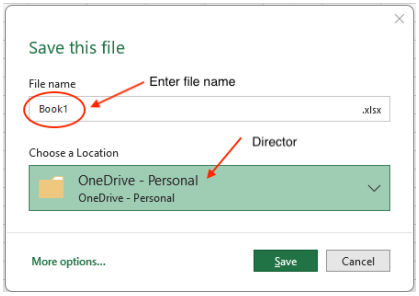


Figure 1.3: Dialog window when saving a workbook on a PC.

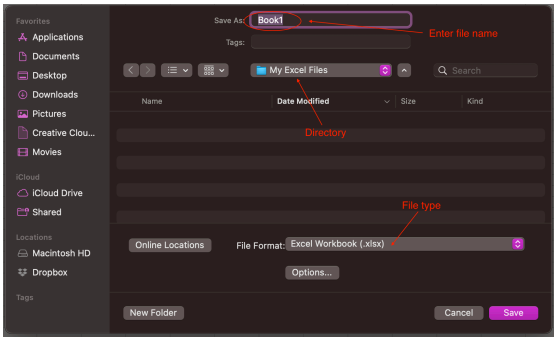


Figure 1.4: Dialog window when saving a workbook on a MAC.

file for the first time, you can save the updates during your working session using the save button on the title toolbar.

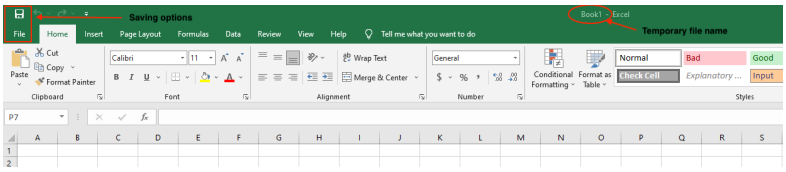


Figure 1.5: Saving options.

It is essential to save your Excel files whenever you start a new project or periodically when you make changes to an existing file. Saving early and often can prevent your work from being lost. You'll also need to pay close attention to where you save the workbook so it will be easy to find later.

You can also turn on the *AutoSave* feature to save your data automatically while working on Excel. The following are the steps to turn on Auto-save.

1.2.1 AutoSave on a PC

1. Click the `File` tab.
2. Click on `Options`.
3. In the Options dialog window, click on the `Save` option on the left.
4. Check the option `Save AutoRecover information every`. By default, the value is 10 minutes, but you can choose a lower or higher value.
5. Then click `Ok`.

1.2.2 AutoSave on a Mac

1. On the Excel title bar, click `Preferences`.
2. Under the category `Sharing and Privacy`, click `Save`.
3. Select the `AutoSave` or `AutoRecover` info in every check window, and then enter how frequently you want to save the Excel files.

A shortcut is available on Mac to activate the AutoSave. You can just toggle `ON` the `AutoSave` option in the Toolbar of the program.

Note: AutoSave in Excel is a special feature for Microsoft 365 subscription accounts. When you save your Excel spreadsheet online on OneDrive or SharePoint, Excel activates the AutoSave feature. AutoSave in Excel saves your progress every few seconds so that all changes appear online if someone else is working on the same file.

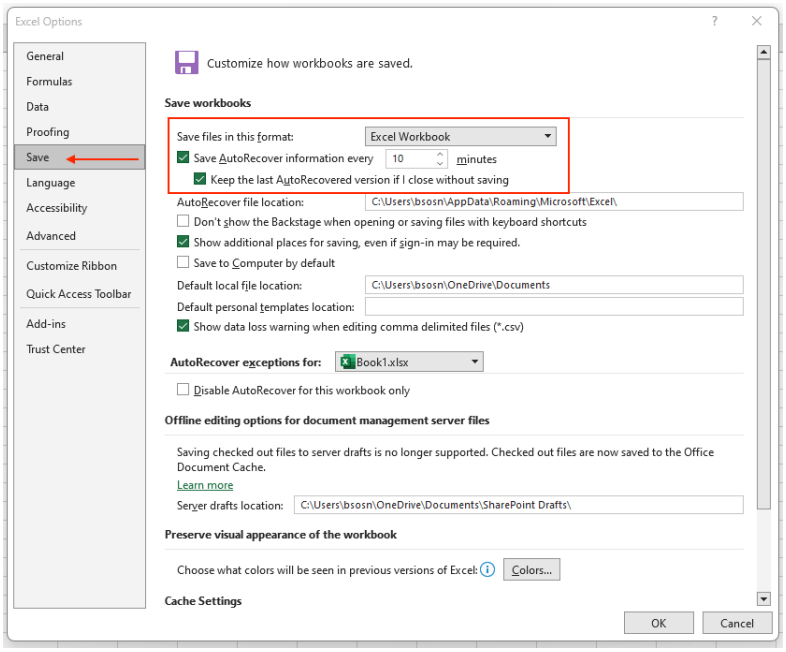


Figure 1.6: AutoSave option on PC.

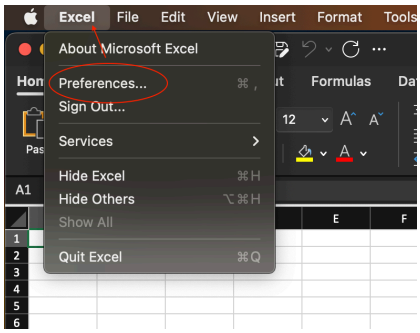


Figure 1.7: Dropdown menu in the Excel title bar.

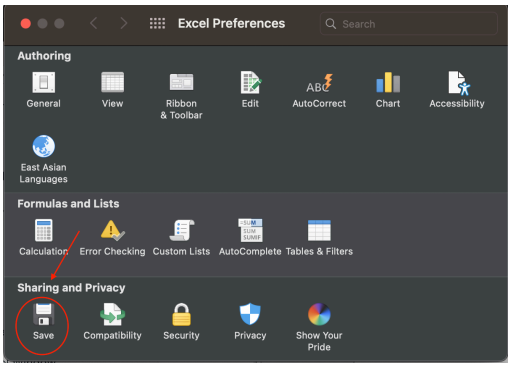


Figure 1.8: Save option in Excel Preferences

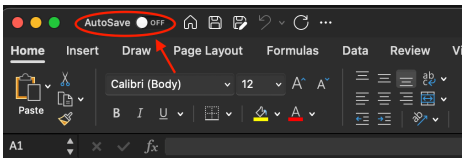


Figure 1.9: AutoSave option in the Toolbar.

1.3 Opening an Excel File

To open an Excel file, go to the main menu and click `File > Open`. Select the directory containing the file, then select the file you want to open.

Note: On PCs, you will get the Enable Editing warning when you open an Excel document downloaded from the Internet for the first time. You must click the Enable Editing button to modify and work on it.

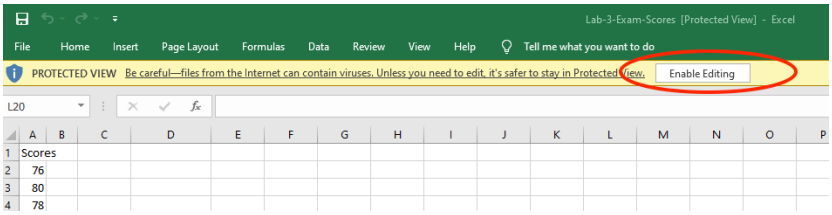


Figure 1.10: Enable Editing warning.

If you are working on your own computer, you can change Excel so that editing is enabled by default. Click File > Options >Trust Center > Trust Center Setting > Protected View. You can then uncheck the options you want related to Enable Protected View.

1.4 The Analysis ToolPak

You may need to install the *Analysis ToolPak* from the original Excel software if your computer does not have it.

1.4.1 Adding the Analysis ToolPak on a PC

With an Excel worksheet open:

1. Click on `File` in the main menu.
2. Click on `Options`.
3. Click `Add-ins` (left side menu) to see if the ToolPak is listed in the Active Application Add-ins. If it is, the ToolPak is installed. If you do not see a listing for Analysis ToolPak in Active Application Add-ins, it may be under the list of inactive applications. IYou must install it from the original Excel installation source if you do not see it.
4. To install or activate it, click `Go`.
5. Check the box next to `Analysis ToolPak`.
6. Click `OK`.

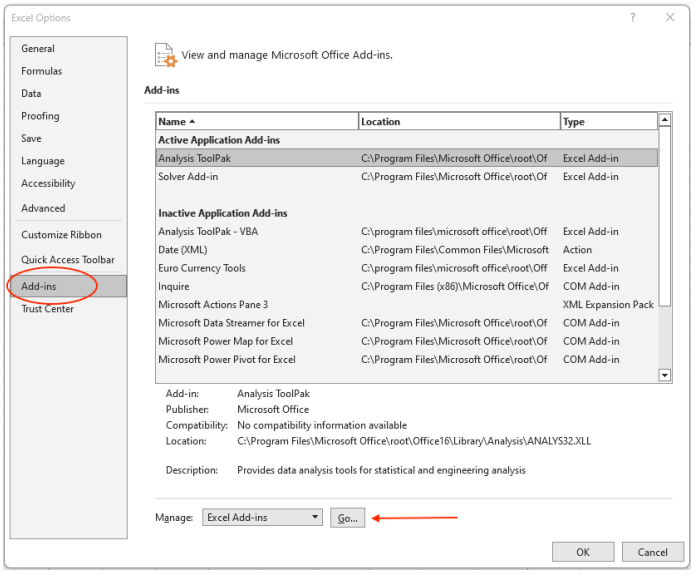


Figure 1.11: Excel Options on a PC.

1.4.2 Adding the Analysis ToolPak on a Mac

With an Excel worksheet open:

1. Click on `Tools` in the main menu.

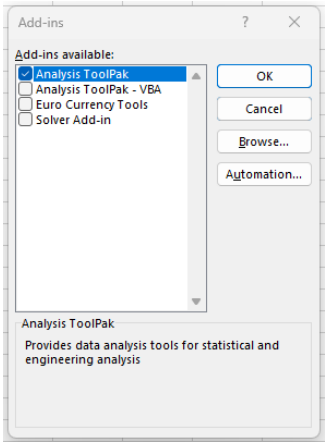


Figure 1.12: Add-ins Options on a PC.

2. Click on `Add-ins` and see if the `Analysis ToolPak` is listed as an active application. If it is not, check the box next to it.
3. Click `OK`.

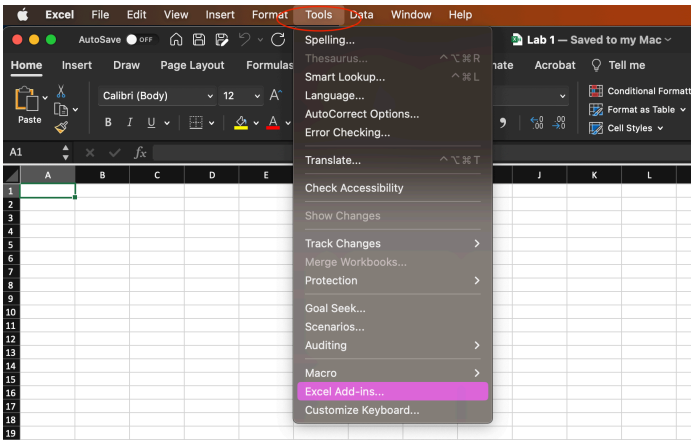


Figure 1.13: Excel Tools on a MAC.

1.5 Checking if the Analysis ToolPak is Active

1. Click on the tab `Data`.
2. You should be able to see it as a ribbon on the top right of your Excel Window.

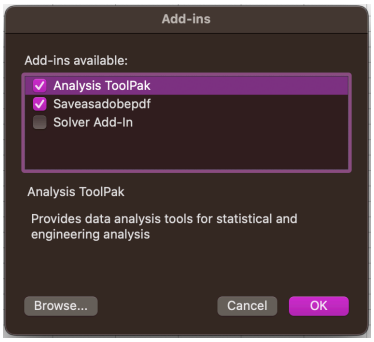


Figure 1.14: Add-ins Options on a MAC.

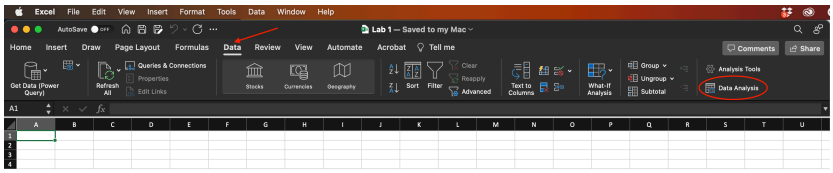


Figure 1.15: Data tab and Analysis ToolPak ribbon.

1.6 The Excel Workbook

The workbook is an Excel file that consists of one or more worksheets. The tabs near the bottom of the screen show that we are working with Sheet 1 (check Figures **1.1** and **1.2**) . To change worksheets, click on the corresponding worksheet tab. Alternatively, you can `RIGHT-click` (or `Control + click`) the arrows just to the left of the worksheet tabs to get a list of all the worksheets in the projects and select a worksheet.

1.6.1 The Cells in the Worksheet

When you look at a worksheet, you should see horizontal and vertical grid lines. Grid lines are light grey lines that appear around cells on Microsoft Excel worksheets. If they are missing, you will need to activate that feature. To do so, see lines:

1. Select the `View` tab.
2. Click the `Show` group on the tab.
3. Be sure that the `Gridlines` option is checked or selected.

1.6.2 Cell Addresses

Intersecting rows and columns form the cells. A cell's address consists of a column letter followed by a row number. For example, address B3 designates

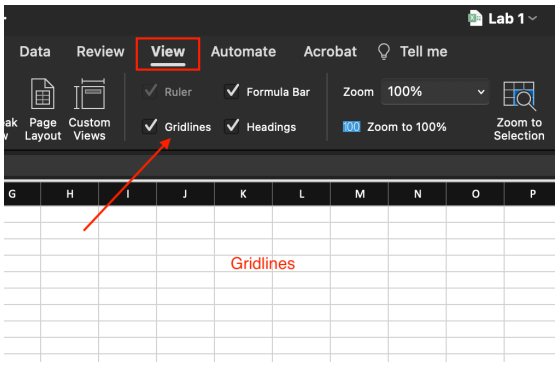


Figure 1.16: Gridlines option in the View tab.

the cell in Column B, Row 3. When Cell B3 is highlighted, it is the active cell, which means we can enter text or numbers into Cell B3.

1.6.3 Selecting Cells

To select a cell, position the cursor in the cell and click the left mouse button (or click on your computer's trackpad). Sometimes you will want to select several cells simultaneously to format them (as described next).

To select a rectangular block of cells, position the cursor in a corner cell, hold down the left mouse button, drag the cursor to the cell in the opposite corner of the block, and release the button.

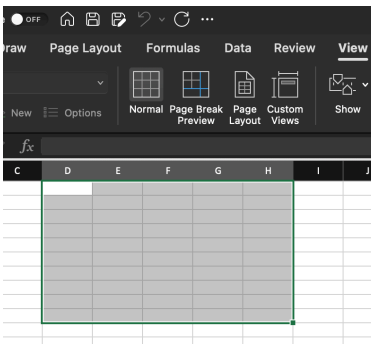


Figure 1.17: A block of cells selected in the worksheet.

To select an entire column, click on the letter above the column to be selected; to select an entire row, click on the number to its left.

To select every cell in the worksheet, click on the blank gray rectangle in the upper left corner of the worksheet (above row header 1 and left of column header A).

You can also select a block of cells by typing the two corner cells into the active cell address window. For example, typing B3:E4 and pressing `Enter` would select the rectangular block of cells from B3 to the opposite corner E4.

1.6.4 Formatting Cell Contents

In Excel, you may place text or numbers in a cell. As in other Windows applications, you can format the text or numbers by using the formatting toolbar buttons for bold (B), italics (I), underline (U), etc. Other options include left, right, and centered alignment within a cell.

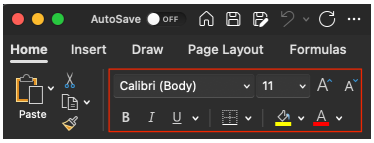


Figure 1.18: Text Formatting ribbon in the Home tab.

Numbers can be formatted to represent dollar amounts ($) or percent form (%) and can be shown with commas in large numbers (,). The number of decimal places to which numbers are carried is also adjustable. All these options appear on the formatting menu bar. Other options are accessible by `RIGHT-clicking` on a cell and selecting Format Cells.

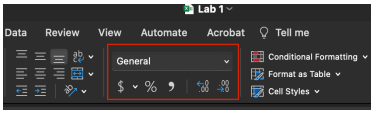


Figure 1.19: Number Formatting ribbon in the Home tab.

1.6.5 Changing Cell Width

Column widths and row heights can be adjusted by placing the cursor between two columns' letters or row numbers. Hold the left mouse button down when the cursor changes appearance, move the column or row boundary, and release. All these instructions may seem a little mysterious, but once you try them, you will find that they are fairly easy to remember.

1.7 Entering Data

In Excel, we enter data and labels in the cells. It is common to select a column for the data and place a label as the first entry of the column.

1.7.1 Entering and Correcting Data

To select a cell for content entry, move the mouse pointer to the cell and click. Then type the label or data and press Enter. Excel automatically moves to select the next cell in the same column. If you want to enter information in a different cell, click on it.

Errors are easily fixed. If you notice a mistake before you press Enter, back-arrow to the mistake and correct it. If you see the error after you have pressed Enter, select the affected cell, and then click on the formula bar to add a typing cursor to the cell contents displayed. Use standard keyboard editing techniques to make corrections, then press `Enter`.

If you want to erase the contents of a cell or range of cells but keep the formatting:

1. Select the cells and click `Home`
2. Click the arrow next to `Clear`
3. Select `Clear Content` (or press `Delete`)

The `Clear > Clear Formats` option keeps the content but clears the format. The `Clear > Clear All` option removes both content and format.

Note: `Clear > Formats` is handy for changing percent data back into decimal format.

1.7.2 Practice 1

The number of ads and the time devoted to ads were recorded for every twenty hours of prime-time viewing. Enter the data as shown in Figure **1.20** in your Excel workbook.

1.8 Arithmetic Options on the Standard Toolbar

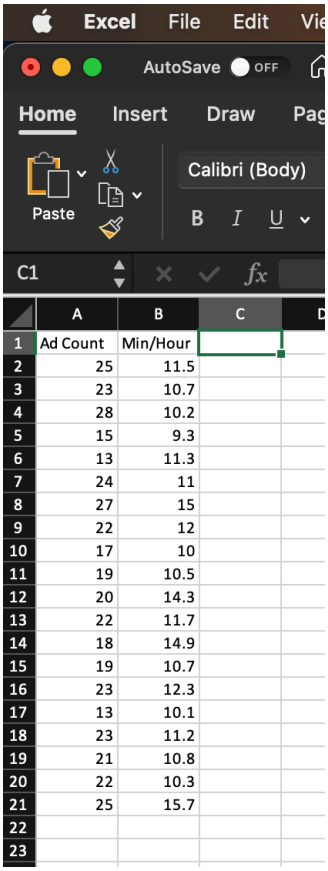
1.8.1 Sum Function

On the `Formula` Menu, the button you see in Figure **1.21** automatically sums the values in the selected cells.

When we sum the contents of an entire column, Excel places the sum under the selected cells. It is a good idea to type the label *Total* next to the cell where the total appears.

1.8.2 Practice 2

For the data you entered in your worksheet (see Practice 1), select cells in Column A containing numerical values (a2:a21). Press the button `AutoSum`,



	A	B	C
1	Ad Count	Min/Hour	
2	25	11.5	
3	23	10.7	
4	28	10.2	
5	15	9.3	
6	13	11.3	
7	24	11	
8	27	15	
9	22	12	
10	17	10	
11	19	10.5	
12	20	14.3	
13	22	11.7	
14	18	14.9	
15	19	10.7	
16	23	12.3	
17	13	10.1	
18	23	11.2	
19	21	10.8	
20	22	10.3	
21	25	15.7	
22			
23			

Figure 1.20: Excel worksheet with data.

Figure 1.21: Autosum button.

then type the word *Total* in the corresponding row of Column C. We see that the sum of Column A is 419.

1.8.3 Sorting Data

The `Sort` and `Filter` ribbon on the `Home` tab shown in Figure **1.22** can be used to sort the data in ascending or descending depending on the selection from the sub-menu.

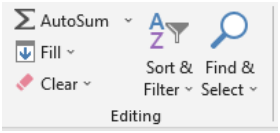


Figure 1.22: Sort and Filter buttons.

To sort just one column, highlight that column and press the button and select the sort order. To sort two or more columns by ascending or descending order of the data in the first column, highlight all the columns and click the appropriate button. In general, we will sort one column of data.

1.8.4 Practice 3

1. Sort the data in the first column in ascending order.
2. Now undo the ordering you just did to your data. – If you have not made any other changes since you did the sort, you can use `Undo` function of Excel from the `Title Bar` from the main menu. The data will appear in their original order. – Another option to undo actions in Excel, use the keys `CRTL-Z` on a PC keyboard or `Command-Z` on a MAC.

After Step 2 above, your data set should be in the original form.

1.8.5 Copying Cells

To copy one cell or a block of cells to another location on the worksheet:

1. Select the cells you wish to copy.
2. From the `Home` tab, select `Copy` (the shortcut with the keyboard for this process is `CRTL-C` on PC or `Command-C` on MAC). Notice that the range of cells being copied now has a blinking border around it.
3. Select the upper-left cell of the block that will receive the copy.
4. Press Enter. When you press Enter, the copy process is complete, and the blinking border around the original cells disappears.

Note: Even if you use the command `Paste` or the shortcut `CRTL-V` on PC or `Command-V` on MAC to paste, you must still press Enter to remove the blinking border around the original cells. To copy one cell or a range of cells to another worksheet or workbook, follow Steps 1 and 2 above. For Step 3, be sure you are

in the destination worksheet or workbook and that the worksheet or workbook is activated. Then proceed to Step 4.

1.8.6 Practice 4

1. Copy the contents of columns A and B and paste them into columns F and G. Then sort the data in Column F in descending order
2. Also, copy the contents of columns A and B and paste them into columns J and K. Then custom sort the data in Columns J and K by the values in Column J in ascending order

After Step 2 above, your worksheet should have 3 sets of data where the two (2) rightmost ones have modifications in the ordering of the columns according to the instructions in the practice above.

1.9 Using Formulas

A formula is an expression that generates a numerical value in a cell (like a calculator), usually based on values in other cells. Formulas usually involve standard arithmetic operations. Excel uses + for addition, `-` (hyphen) for subtraction, `*` for multiplication, `/` for division, and `^` (carat) for exponentiation (raising to a power). For instance, if we want to divide the contents of Cell A2 by the contents of Cell B2 and place the results in Cell C2, we do the following:

1. Make Cell C2 the active cell.
2. Click in the formula bar and type `=A2/B2`.
3. Press `Enter`.

Cell C2 will be the quotient of the values in Cells A2 and B2. If for a whole series of rows, we wanted to divide the entry in Column A by the entry in Column B and put the results in Column C, we could repeat the above process over and over. However, the typing would be tedious. We can accomplish the same thing more easily by replicating the formula in one cell in other cells.

1.9.1 Practice 5

1. Enter `=A2/B2` in Cell C2 as described above.
2. Move the cursor to the lower right corner of Cell C2. The cursor should change shape to a small black cross (+). Hold down the left mouse button and drag the + down until all the cells in Column C where you want the calculation done are highlighted.
3. Release the mouse button (you may need to press `Enter` right after, depending on how your Excel program is configured). The cell entries in Column C should equal the quotients of the same-row entries in Columns A and B.

1.10 Adjustment to Formulas

Now, if you click on one of the lower cells in Column C, you will see that the row number in the cell addresses is not 2, but rather the number of the new cell's row. Generally, when a formula is copied from one cell to another, the cell addresses in the formula are automatically adjusted. For example, if the formula `=D3+C7` is copied to a new cell, three columns right and two rows up from the old one, the pasted formula comes out as `=G1+F5`. (Three columns right from D is G, two rows up from 3 is 1, and so on.)

Sometimes you will want to prevent the automatic address adjustment. To do this, put a dollar sign before any row or column number you want to keep from changing. When the formula `=D$3+$C7` is copied to a new cell, three columns right and two rows up from the old one, the pasted formula comes out as `=G$3+$C5`. We will call an address with two `$` signs in an absolute address because it always refers to the same cell (`$` fix the addresses in a formula), no matter where the formula is copied/pasted. For a cell with only one `$` sign in it or none, we will call a relative address because the cell referred to can change as the formula is pasted from one location to another.

Chapter 2

Random Number Generators and Tables

There may be cases when you need to generate random numbers in Excel. Excel has a few random number generators: `RANDBETWEEN`, `RAND`, and `RANDARRAY`. In this manual, you will get acquainted with the `RANDBETWEEN` function.

Tables are one of the most powerful Excel features. This lab shows you what you can do with a table. You will learn how to insert tables in Excel, table formatting and gain an understanding of Excel table functions and formulas.

2.1 Random Samples

The one you will find most convenient for our labs is the function RANDBETWEENN(bottom,top), which Returns a random integer number between the numbers specified in the function's arguments. The other random functions are `RAND()`, which returns an evenly distributed random real number greater than or equal to 0 and less than 1, and `RANDARRAY(rows,columns,min,max,integer)`, which Returns an array of random numbers. For RANDARRAY, you can specify the number of rows and columns to fill, minimum and maximum values, and whether to return whole numbers or decimal values.

We focus now on the `RANDBETWEEN` function.

2.1.1 Insert the function RANDBETWEEN on a PC

To use `RANDBETWEEN`, select a cell in the active worksheet. Click in the formula tab and then click the `Insert Function` ribbon.

Select the category `All` in the dialog window. Then select `RANDBETWEEN` and then fill in the bottom and top numbers. Returns a random integer number

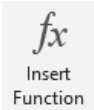


Figure 2.1: Insert Function ribbon.

between the numbers you specify.

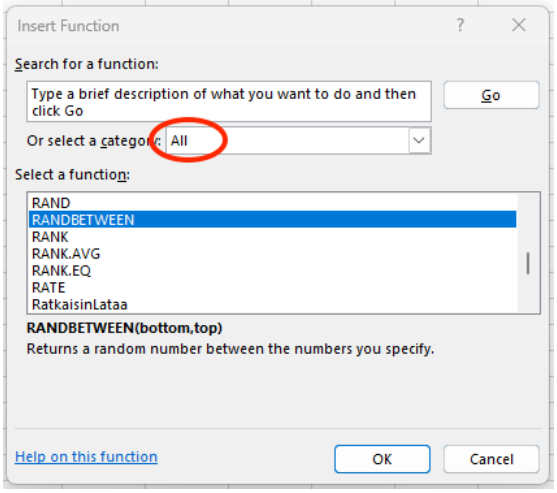


Figure 2.2: The dialog window of Insert Function on PC.

2.1.2 Insert the function RANDBETWEEN on a MAC

1. Click on the `Formula` tab.
2. Click the `Insert Function` ribbon.
3. You will see on the right side of the Excel window that the `Formula Builder` appears.
4. In the search box, start to type a keyword for the function you are looking for. Excel will list functions that match the criteria of the search you type.
5. Then select the function you wish to insert in the worksheet. In this case, the `RANDBETWEEN` function.
6. Adjust the arguments of the formula (top and bottom numbers) as needed.

Note: Alternatively, you can simply type `=RANDBETWEEN(bottom,top)` in the formula bar, with numbers in place of *bottom* and *top* for the arguments of the formula.

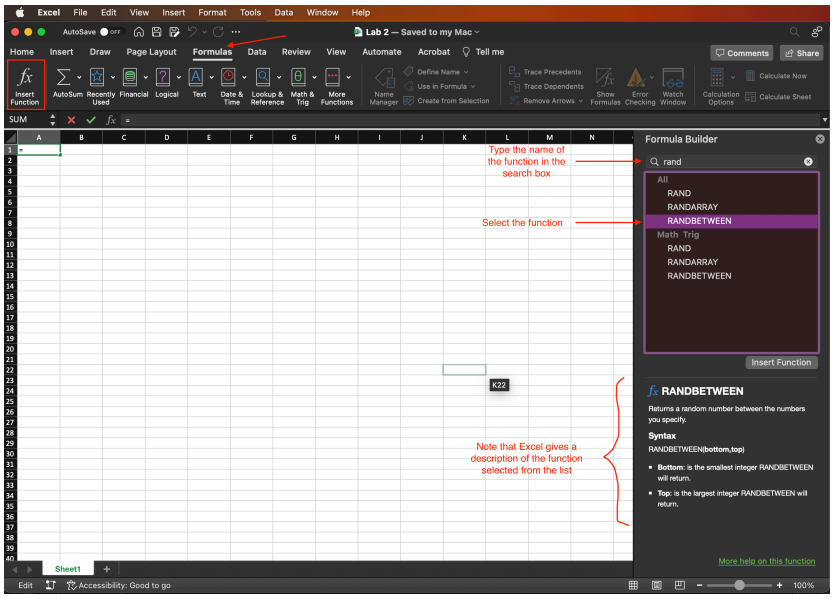


Figure 2.3: Dialog window of Insert Function on MAC.

2.2 Setting the Workbook to Manual Calculation

The random number generators of Excel have the characteristic that whenever a command is entered anywhere in the active workbook, the random numbers change because they are recalculated. To prevent this from happening, change the recalculate mode from automatic to manual.

2.2.1 Manual Calculation on a PC

1. Select `File` then `Options`.
2. Click on the tab labeled `Formulas`.
3. Select `Manual Calculation` for `Workbook Calculation` and uncheck the option `Recalculate workbook before saving`.
4. Then press `OK`.

2.2.2 Manual Calculation on a MAC

1. Go to `Excel Menu` then select `Preferences`
2. Under `Formulas and Lists`, click on `Calculation`
3. For `Workbook Calculation`, select `Manual calculation` and uncheck the option `Recalculate workbook before saving`.

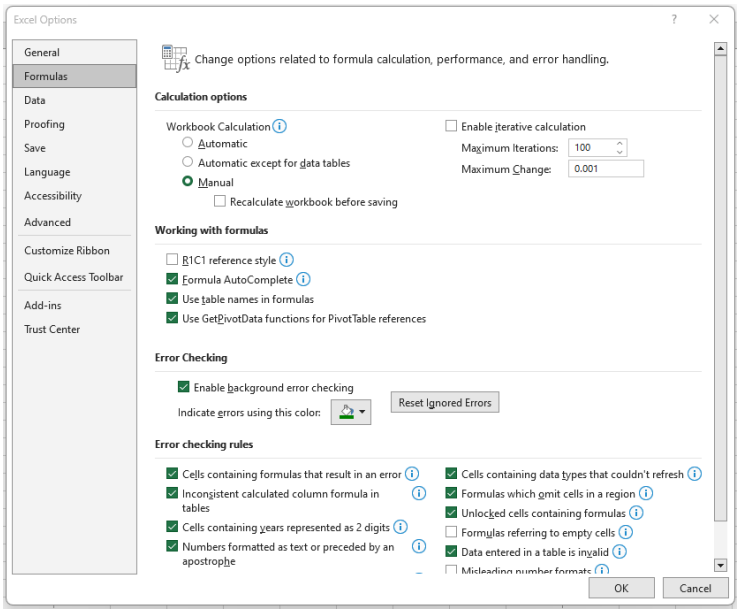


Figure 2.4: Calculation options under Formulas in a PC.

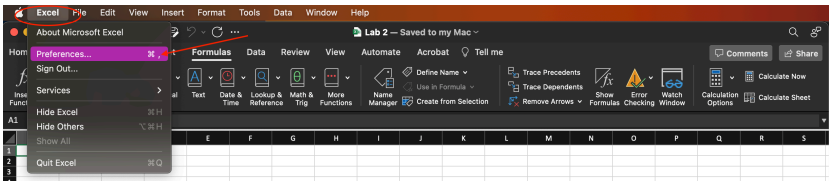


Figure 2.5: Preferences in the Excel Menu on a MAC.

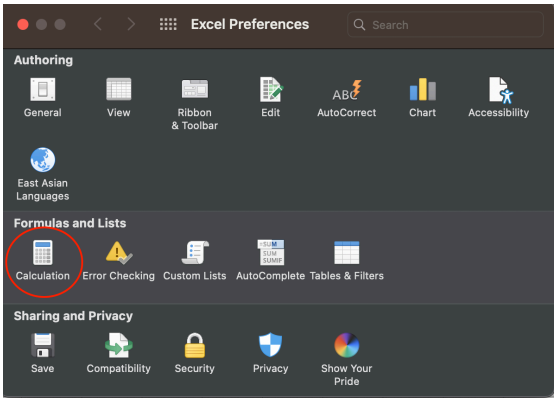


Figure 2.6: Formulas and Lists in Preferences on a MAC.

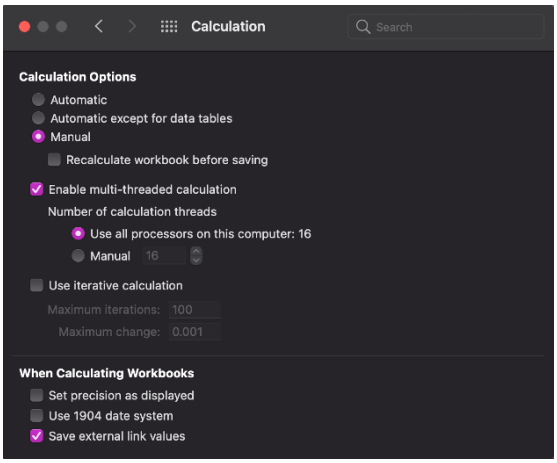


Figure 2.7: Calculation Options on a MAC.

2.3 Using the Recalculation Mode

Even if the `Calculation` option is set for `Manual`, you can use a ribbon command or keyboard shortcut to force a recalculation when needed.

You can still recalculate by pressing the keys `Shift` + `F9` on the keyboard on a PC (For MAC, use keys `Command` + `=`) or pressing the button `Calculate Now` in the `Formulas` Menu.

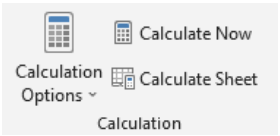


Figure 2.8: Calculation ribbon.

2.3.1 Practice 1

Suppose that there are 95 students enrolled in a section of the course Introductory Statistics. Draw a random sample of fifteen (15) of the students.

To draw the sample, assign each of the students a distinct number between 1 and 95. To find the numbers of the 15 students to be included in the sample, do the following steps.

1. Change the calculation mode to `Manual` (as described above).
2. Type the label *Sample* in Cell A1 ,
3. Select Cell A2.

4. Type `=RANDBETWEEN(1,95)` in the formula bar and press Enter.
5. Position the mouse pointer in the lower right corner of Cell A2 until it becomes a `+` sign, and click-drag downward until you reach Cell A16. Release. Then press the key `F9` (or `Calculate Now` button).
6. 6) Use one of the `Sort` buttons to sort the data to easily check for repetitions. If there are repetitions, press `Shift` + `F9` (or `Calculate Now`) again and re-sort. Below, with the data sorted, we can verify that there are no repetitions

2.4 Creating Tables

To make working with data easier, you can organize data in a table format on a worksheet.

You are going to use Excel's random number generator and tables to create a fictional grade book.

2.4.1 Practice 2

1. Select the block of cells D1:H11.
2. On the `Home` tab, in the `Styles` ribbon, click `Format as Table`, and then click the table style of your choice.

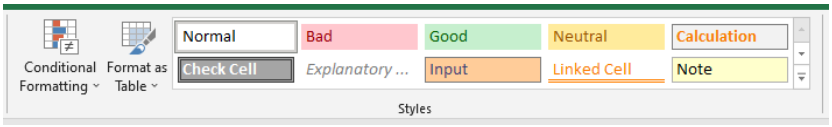


Figure 2.9: Style ribbon.

3. Select the `My table has headers` check box in the `Format as Table` dialog window.

Note: When you select a table by clicking the mouse pointer, the `Table Tools` menu becomes available, and a `Design` tab is displayed. To get a good idea of what you can add to or change in your table, click the `Design` tab, then explore the groups and options provided on this tab.

4. Replace the table headers. Instead of `Columns1` through `Column5`, replace headers by typing in them the following: *Name*, *Exam 1*, *Exam 2*, *Exam 3* and *Final Exam*.
5. Under header *Name*, in cells 2-11, create 10 students' names, writing last name, and first name.
6. Now select the block of cells D2:D11. Click `Home > Sort & Filter > Select A to Z`. The fictional names should then appear in alphabetical order.

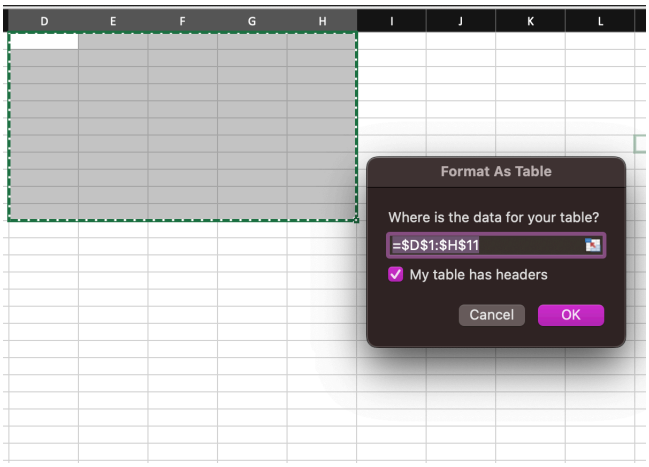


Figure 2.10: 32: Format As Table window.

7. Select cell E2. Click on the `Formula` bar, type in `=RANDBETWEEN(0,100)`.
8. Position the mouse pointer in the lower right corner of cell E2 until it becomes a **+** sign and click-drag downward until you reach cell E11. Release (*Note:* Sometimes this feature is not needed within a formatted table).
9. Repeat the process above for Column F using `=RANDBETWEEN(40,98)`.
10. Repeat the process above for Column G using `=RANDBETWEEN(64,105)`.
11. Repeat the process above for Column H using `=RANDBETWEEN(37,100)`.
12. If the list of numbers is not randomized, click the button `Calculate Now`.

Chapter 3

Frequency Tables and Histograms

You will construct a histogram with 5 classes (bars) using Excel. Graphing a histogram manually requires determining minimum and maximum values in a data set, class width, the lower and upper limits of classes of values, class boundaries, midpoints, etc. This is similar in Excel. You will need the upper-class limits and the midpoints for adjusting the graph created with Excel.

Throughout this manual, *Select* means to left-click with the mouse/pad.

3.1 Preparation

In this lab, you will use a data set that contains scores received on an MA-336 Exam. Download the data set. called *Lab 3 Exam Scores.xlsx* from here: **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-3-Exam-Scores.xlsx**

Open the Excel file you downloaded and follow the instructions below.

3.2 Initial Computations

1. Select cell C1 and in the formula bar type in `=COUNT(A2:A33)`.

Note: You can fill in the input range, A2:A33, in the formula count above by manually typing the address of the cells containing the data values. Alternatively, you can select cells that form the Input Range by clicking on A2, holding, and dragging downward until you reach cell A33. Then release.

2. Type in cell D1 the label *Total number of data entries.*

3. Select cell C2 and in the formula bar type in `=MIN(A2:A33)`.
4. Type in cell D2 the label *Min*.
5. Select cell C3 and in the formula bar type in `=MAX(A2:A33)`.
6. Type in cell D3 the label *Max*.
7. Select cell C4 and in the formula bar type in `=(C3-C2)/5`.
8. Type in cell D4 the label *Width Computation*.
9. 9) Round up the value resulting in cell C4 to the next whole number by using the formula `=ROUNDUP(C4,0)` in cell C5 the adjusted value for the width.

Note: The formula above uses the cell address (C4) to identify the number to be rounded up with 0 decimal places.

10. Type in cell D5 the label *Width*.

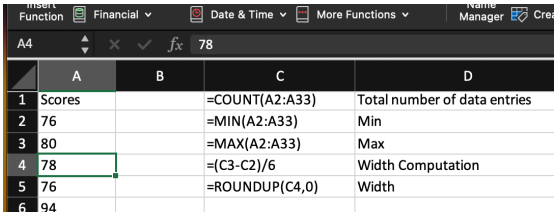


Figure 3.1: Initial computations for constructing a histogram.

3.3 Class Limit Calculations

1. Type in cell C7 the label *Lower Limits*.
2. In cell C8, type the formula `= C2`. This will copy the value in cell C2 as the histogram's lower limit for the first class.
3. 3) Select cell C9, and in the formula bar, type `=SUM(C8,$C$5)`. Remember the `dollar sign` because the value in C5, the width, is a fixed value for all formulas you use here.

Note: You may also do the computation in Step 3 above using the keyboard key `+`.

4. Position the mouse pointer in the lower right corner of cell C9 until it becomes a `+` sign and click, hold, and drag downward until you reach cell C12. Then release.

Note: You should have a total of 5 class limits since your histogram has 5 classes.

Note: If the formulas do not update automatically, do Step 5 below. Otherwise, skip to Step 6.

5. Select `Formulas>Calculate Now`.
6. Type in cell D7 the label *Upper Limits*.

7.Type the values of the upper limits of each class in cells D8 through D12.

Note: You can also be creative and use a formula to do the work in this step. There are different ways of entering the upper limits using formulas, but that is not being requested from you for this lab.

8. Type in cell E7 the label *Midpoints*.
9. Compute the values of the midpoints (lower limit + upper limit)/2 of each class in cells E8 through E12 by using Excel formulas.

Note: To do this step, use the techniques and Excel functions presented above (create the Excel formula!)

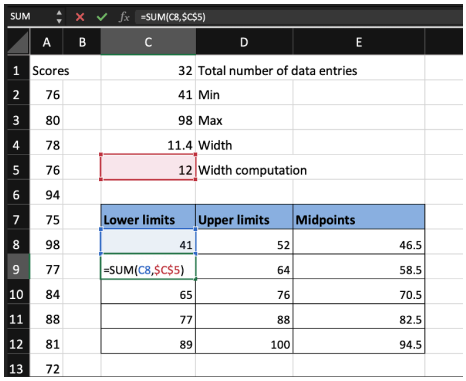


Figure 3.2: Class limits computations.

3.4 Creating the Histogram

1. Select `Data>Data Analysis>Histogram` and click `OK`.

Note: If `Data Analysis` does not appear in the Data menu, refer to the instructions in *Lab **1** Getting Started with Excel* how to make it available.

2. In the dialog window, first click in the `Input Range` field then type in `$A$2:$A$33`.

Note: Instead of typing, you can *select cells* that form the Input Range by clicking on A2, holding, and dragging downward until you reach cell A33. Then release.

3. 3) Click the `Bin Range` field in the dialog window first, then type in `$D$8:$D$12`.

Note: You can also select cells containing the upper limits instead of typing !

4. You have the option to have the graph placed in a specific cell or range of cells (for this option check `Output Range` as in the picture above, you may select a different cell address) or in a new worksheet (for this check option `New Worksheet Ply`).
5. In the pop-up window, Check the boxes for `Cumulative Percentage` and `Chart Output`.
6. Then click `OK`.

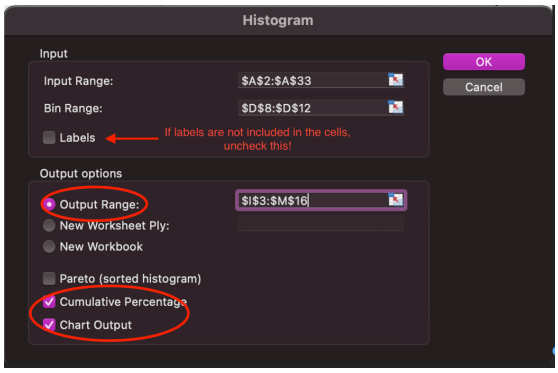
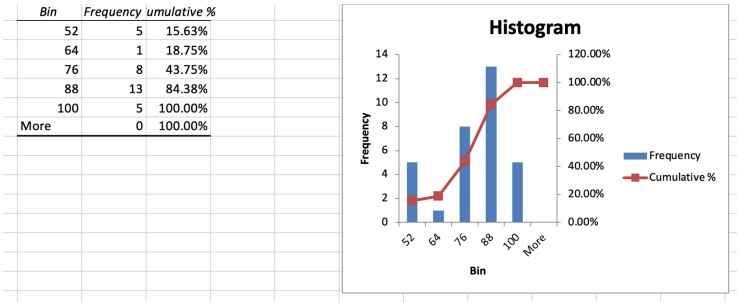


Figure 3.3: Dialog window with the histogram options.

After the steps above, Excel presents a frequency table and the corresponding graph. Excel counts the frequency of each class based on the *Bins* values you entered in the dialog window for the histogram. Thus, it counts how many values are in the date set (Input Range) up to the upper limits.

The graph is a *raw histogram* that needs to be adjusted to resemble an actual histogram.



Bin	Frequency	umulative %
52	5	15.63%
64	1	18.75%
76	8	43.75%
88	13	84.38%
100	5	100.00%
More	0	100.00%

Figure 3.4: Resulting histogram.

3.5 Removing the *More* bar of the Histogram

1. If you chose `New Worksheet Ply` in Step 4 above, select the Worksheet that contains the graph. Otherwise, go to next Step 2 below.
2. You may resize the graph by clicking anywhere within the figure. Handles appear along the perimeter of the graph. You can change the shape of the figure by clicking on a handle and dragging it. Make your graph to be displayed in an appropriate size.
3. A frequency table was created with the graph in the process. That table controls some aspects of the graph. Select the cells in the last row of the frequency table, the one that contains the word *More*.
4. Select `>Home>Delete` (you can also do this by `RIGHT-clicking` on the selected cells and then the option `Delete`).

Note: If you do not delete the row in the table, the graph will contain an extra bar on the right of the graph with a frequency equal to zero. **Be sure to delete the cells and not just delete the contents in those cells.**

5. If the graph doesn't update automatically, select `Formulas>Calculate Now`.

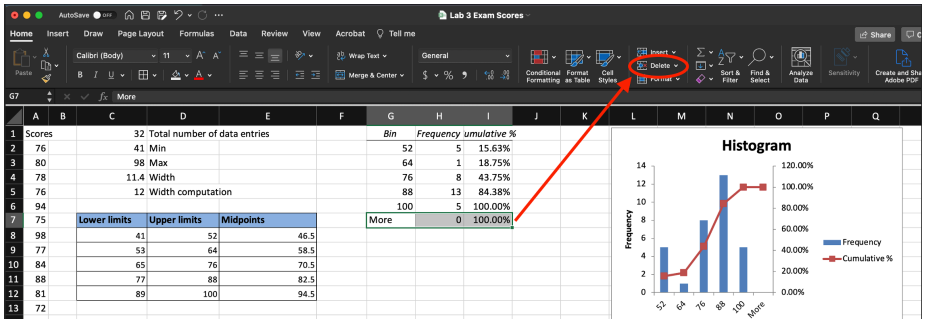


Figure 3.5: Deleting the *More* row from the frequency table of the histogram.

After Steps 4-5, the last column in the graph, labeled *More* will no longer be displayed in the graph.

6. `RIGHT-click` on any one of the vertical bars of the graph and then select `Format Data Point` (or `Format Data Series`). Alternatively, you can also double-click with your mouse on a vertical for that.
7. Move `Gap Width` to position `0%` (this makes the bars of the graph touch).

3.6 Adjusting the Horizontal Axis

3.6.1 MAC

1. `RIGHT-click` on any vertical bar of the graph. From the menu, click the option `Select Data`.
2. Click `Edit` under *Horizontal (Category) Axis Labels*. Click and drag over the range `E8:E12` (midpoints).
3. Click `OK`. This is to re-label the x-axis with the midpoints replacing the upper limits.
4. Click OK again.

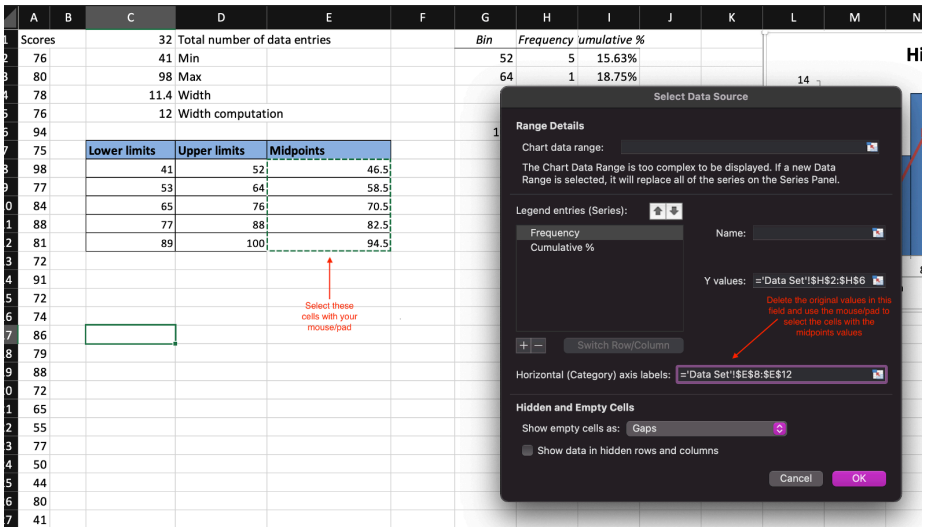


Figure 3.6: Adjusting the horizontal axis on a MAC.

3.6.2 PC

1. `RIGHT-click` on any vertical bar of the graph. From the menu, click the option `Select Data`.
2. Click `Edit` under *Horizontal (Category) Axis Labels*.
3. Click in the filed *Axis label range*, delete whatever values are in there, and drag over the range `E8:E12` to select midpoints cells.
4. Click `OK`. This is to re-label the x-axis with the midpoints instead of the upper limits.
5. Click `OK` again.

3.7 Adjusting the Title of the Histogram

1. To change the histogram title, click on the word *Histogram*, which appears in a window. Change the title to *MA336 Exam Scores*.

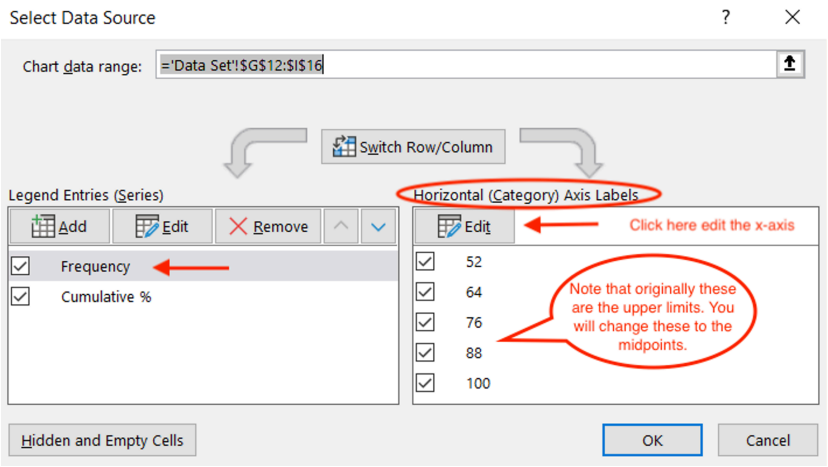


Figure 3.7: Adjusting the horizontal axis on a MAC.

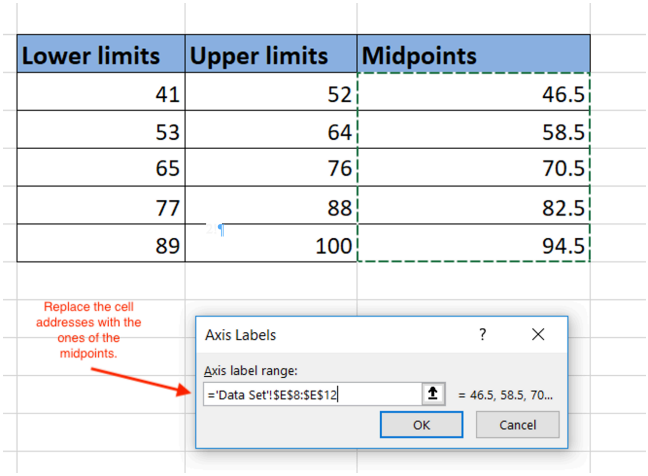


Figure 3.8: Relabeling the horizontal axis on a PC.

2. Click on the word *Bin* and change the x-axis label to *Scores*.

After all the adjustments are made, the graph will look like the one below.

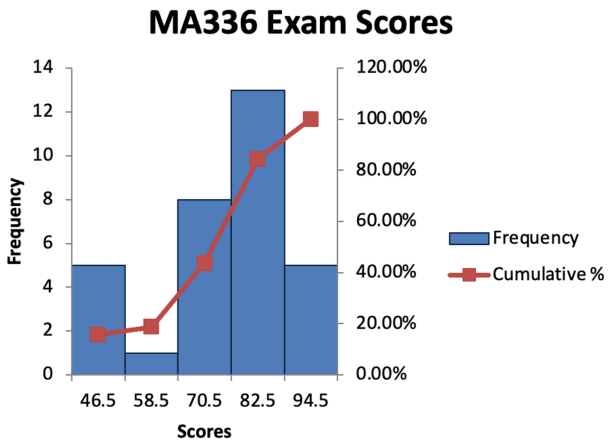


Figure 3.9: Adjusted histogram.

3.8 Practice

Practice: Dan gathered data about the number of minutes per day that a sample of teenagers spends in doing a physical activity.

15, 30, 45, 25, 30, 15, 20, 10, 30, 40, 25, 15, 20, 25, 25, 29, 40, 10, 20, 27, 10, 15, 20, 20, 30, 25, 20, 60, 60, 75, 50, 45

Use the data and the techniques learned in these lab instructions to build a histogram with five classes and descriptive statistics.

Chapter 4

Measures of Central Tendency and Variation

In this lab, you will use Excel to determine a data set's mean, median, mode, and standard deviation. A measure of central tendency shows where the center or middle of the data set is located. A measure of variation shows the spread among data values.

Throughout these instructions, *Select* means to left-click with the mouse/pad.

4.1 Preparation

In this lab, you will use a data set that contains three samples. You will compute descriptive statistics in different ways. Download the data set called *Lab 4 Samples.xlsx* from here: **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-4-Samples.xlsx**

- Sample 1 contains 100 data values.
- Sample 2 contains 120 data values.
- Sample 3 contains 95 data values.

Open the Excel file you downloaded and follow the instructions below.

4.2 Using the Descriptive Statistics in Excel to Obtain the Measures

Sample 1

1. Go to `Data > Data Analysis`.
2. Select `Descriptive Statistics` and click `OK`.

Note: If `Data Analysis` does not appear in the Data menu, refer to the instructions in *Lab **1** Getting Started with Excel* how to make it available.

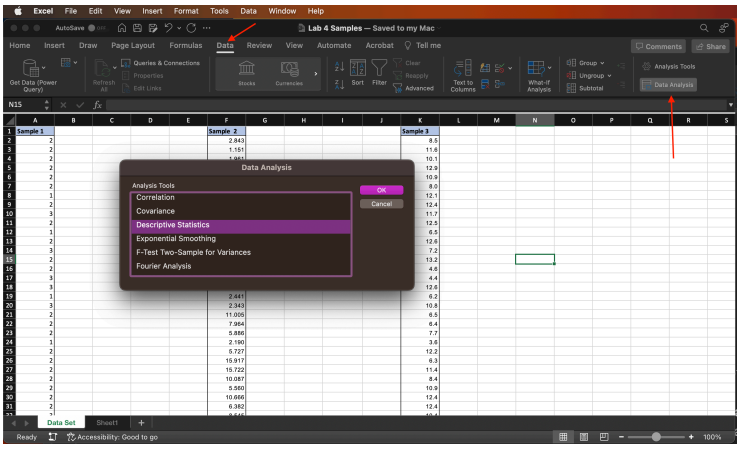


Figure 4.1: Data Analysis dialog window.

3. The *Descriptive Statistics* window should appear. Complete the Descriptive Statistics window with the information needed from the data in Sample 1.
4. The `Input Range` for Sample 1 has cell address `$A$2:$A$101` (not including the label). If you include the label cell in the address, the address is `$A$1:$A$101` (the first cell address is different). You may want to fill in the input range by manually typing the address or by selecting the cells.
5. If the data values are stacked in a column, then check the options `Grouped by Columns`. If they were entered into *Rows* instead, then you would select the rows option.
6. After you filled in the input range field, check or uncheck the box `Labels in the first row` according to what is in the input range.
7. Check the option `Output Range` and select a cell in a blank area of your worksheet. 8.Check the option `Summary statistics`.
8. Click `OK`.

These options give you the statistics of Sample 1 such as mean, median, standard deviation, and others.

4.2.1 Practice 1

Now you will do the same work above (Steps 1-8) for *Samples 2* and *3*.

Note: In the instructions above, you must adjust the cell addresses for Samples 2 and 3.

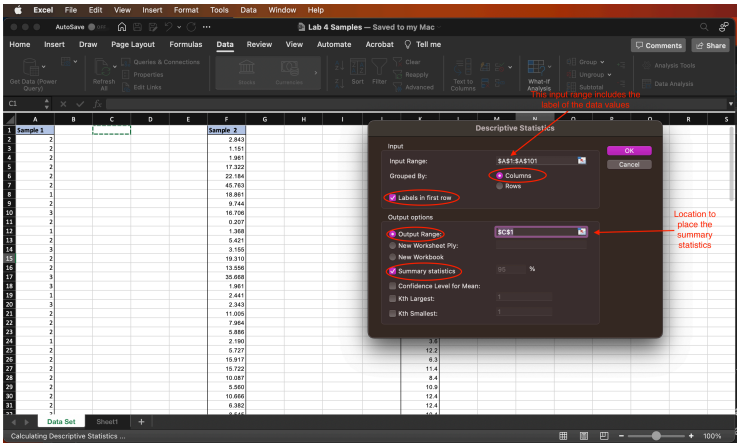


Figure 4.2: Descriptive Statistics dialog window.

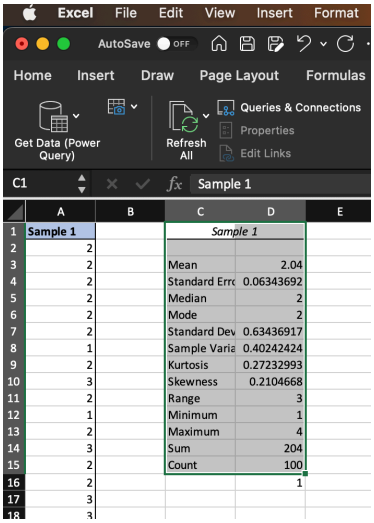


Figure 4.3: The resulting summary statistics for Sample 1.

4.3 Using Excel's Functions to Obtain the Measures

Sample 1

1. Next to the column containing Sample 1 (between Sample 1 and Sample 2), click on an empty cell of the worksheet and type *Mean*.
2. Click on the cell right next to it and type `=AVERAGE(A2:A101)`.

Note: You can also use the Excel function Average by clicking: `Formulas > Insert Function`.

3. Below the cell *Mean*, type the word *Median*.
4. Click on the cell right next to it and type `=MEDIAN(A2:A101)`.
5. Below the cell *Median*, type *Stand. Dev.*
6. Click on the cell right next to it and type `=STDEV.S(A2:A101)` (sample standard deviation).

Note: The `STDEV.P` calculates the standard deviation based on the entire population given as arguments and the `STDEV.S` estimates the standard deviation base on a sample.

7. Excel does not have a *Range* function, so below the standard deviation type the word *Range* and next to it compute the range by using the formula `Range = Max - Min`.
8. Below the cell *Range*, type *Mode(s).*

You will investigate the many modes in the data set and what they are. For that, you will use the function `MODE.MULT()`. This function returns a vertical array of the most frequently occurring, that is, the mode(s) of the data set.

9. Starting with the cell next to the word *Mode(s)*, select a vertical range of cells (a block of any number of vertical cells).
10. While the block of cells is selected, type the function `MODE.MULT(A2:A101)` into the Formula bar.
11. Press the keys `Control` + `Shift` + `Enter` (For MAC, use the keys `Command` + `Shift` + `Enter`).

Note: Be careful! If you only press Enter or use the function MODE, Excel will only display one value for the mode even if there are multiple modes.

Note: If there is no mode, #N/A will appear as a result. If there is only one mode, MODE.MULT returns the same value repeatedly.

4.3.1 Practice 2

Now you will do the same work above (Steps 1-10) for *Samples 2* and *3*.

Note: In the instructions above, you must adjust the cell addresses for Samples 2 and 3.

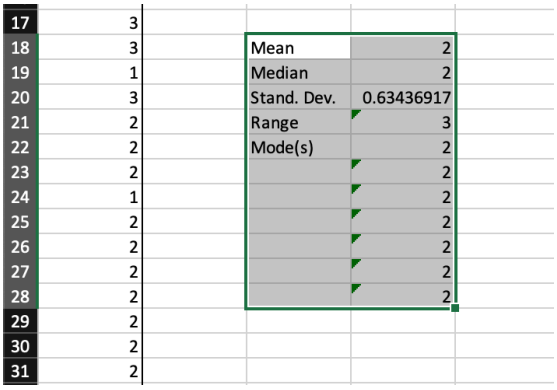


Figure 4.4: The results from the Excel functions used above..

Compare the results obtained in *Section* ***4.2*** with the results obtained in *Section* ***4.3***. The results should be the same.

Chapter 5

Time Series

A *time series* is a series of data points indexed (or listed or graphed) at successive equally spaced points in time.

In this lab, you will build a time series using Excel. In addition, as practice, you will use Excel to calculate the measures of tendency and variation.

5.1 Preparation

Below are some sources you can use for this lab. The websites below are financial sites for stock or index (a financial assest) prices and business information. Choose one asset and collect *30 related values or more data values* to complete this lab.

- **www.nasdaq.com**

Select a stock. On the left side menu in the webpage, there is the option for its historical quote, then download data.

- **www.finance.yahoo.com**

Go to the tab *Markets*, select the type of asset (e.g., *Trending Tickets*), then go to the tab *Historical Data* to download the data.

- **https://www.investing.com**

Select an asset available on the website. Go to *Historical Data* in the tab *General* to obtain the data set.

5.2 Download the Data

1. After choosing one of the financial sites to obtain your data, select an asset to download its prices.

2. Once you have selected an asset , choose a range of 30 or more consecutive days of prices.
3. Download the data into an Excel file.

5.3 Preparing the Data

You will label each day in the data set with sequential numbers from 1 to 30 (or more than 30, depending on your data). The asset's *Close*, *Open*, *Low*, and *High* values are contained in the file.

5.3.1 Renaming the File

1. Open the file you downloaded from the website.
2. Go to `File > Save As`. Include *Lab 5* in the file's name and save it with the extension *.xlsx*.

5.3.2 Inserting a Column on MAC

Next, you will insert a new column between the columns with *Dates* and *Prices*. The following steps will prompt Excel to insert a new column with all the columns with prices to the right of it.

1. Hold down `CONTROL` and simultaneously click the first column with Prices (e.g., Open).
2. Then on the pop-up menu, click `Insert`.
3. Type the label *Day* at the top of the new column.

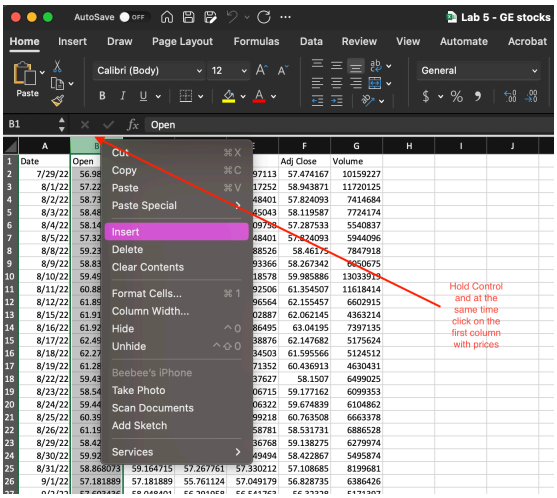


Figure 5.1: Insert column menu on MAC.

5.3.3 Inserting a Column on PC

1. `RIGHT-click` the top of the first column with Prices.
2. Select `Insert`.
3. Type the label *Day* at the top of the new column.

5.3.4 Creating a List of Sequential Numbers

Next, you will create a list in the new column with numbers 1 to 30 (or more, depending on your data).

1. Enter 1 (starting value of the series) in the cell below the label Day in the new column.

Note: If the number 1 changes to a date format, select the column and change it to the `General` format in the `Number` ribbon.

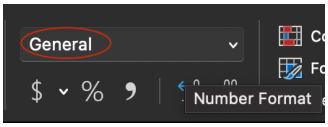


Figure 5.2: Numbers ribbon in the Home tab.

2. Type the number *2* in the next cell below to establish a pattern.
3. Select the cells that contain the two starting values.
4. Position the mouse pointer in the lower right corner of the selected cells until it becomes a `+` sign and click-drag downward across the range you want to fill.

5.4 Time Series

The following steps will guide you to build a time series for displaying your data. Use the asset's *Close* prices to build the time series.

1. Select all cells in both Days and Close columns.

Note: If the columns with *Days* and *Close* prices are not adjacent (separate), first select the cells in one column, and then while holding the `Control` key on MAC (`Control` or `CTRL` on PC), select the cells in the other column.

2. Click `Insert > Scatter > Scatter with Straight Lines and Markers`.

The resulting graph needed to be adjusted.

3. Select the chart created.
4. Replace the *Chart Title* with an appropriate title for your time series. This depends on your data set.

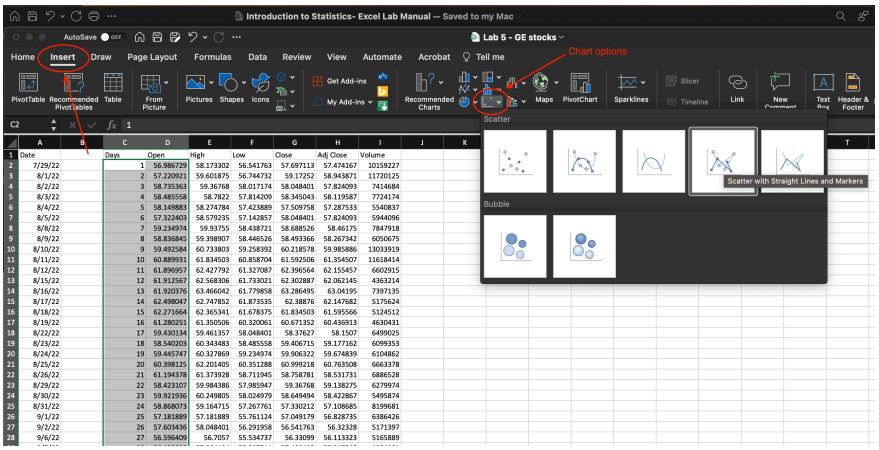


Figure 5.3: Scatter options in the Insert tab.

5. With the graph selected, go to the `Chart Design` tab.
6. To add axis titles, click `Add Chart Element > Axis Titles > Primary Horizontal` and then `Add Chart Element > Axis Titles > Primary Vertical`.
7. Replace the Y-axis Title with an appropriate name for the vertical axis.
8. Replace the X-axis Title with the word *Day*.

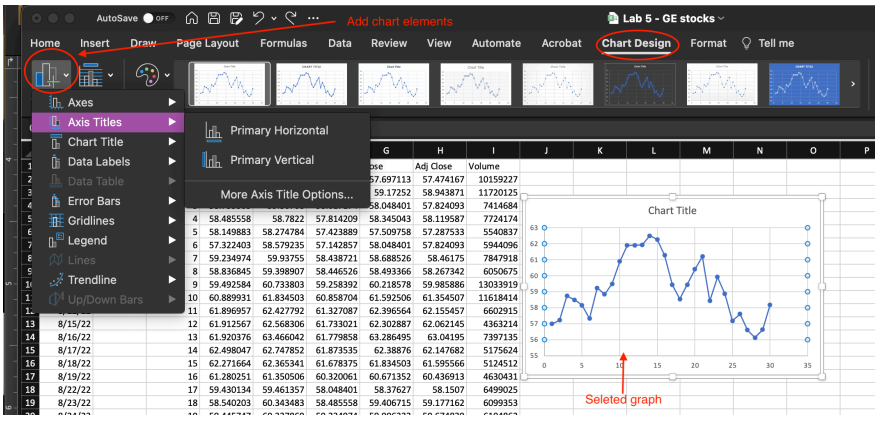


Figure 5.4: Chart adjustment options Chart Design.

9. Make any adjustments necessary to the chart's vertical and horizontal axes by `RIGHT-clicking` on them and selecting `Format Axis`.

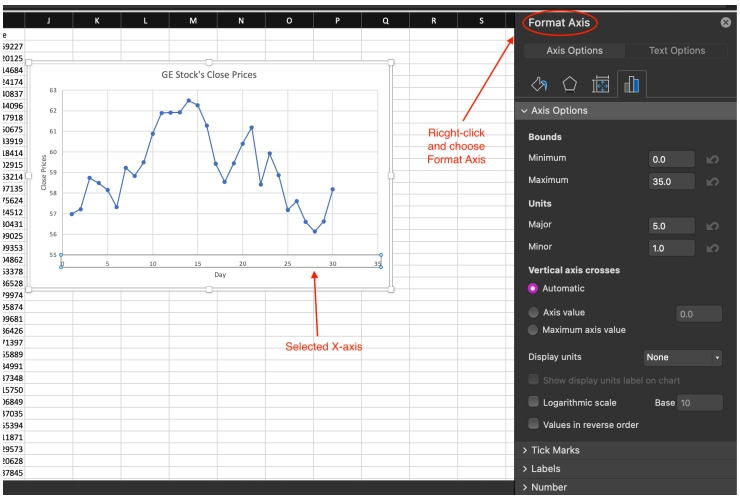


Figure 5.5: Chart adjustment of the horizontal axis.

5.5 Practice 1

Use Excel to build a time series for the *High* prices in your data set.

5.6 Practice 2

Use Excel to calculate your data set's mean, mode, median and sample standard deviation of the *Close* prices. Use *Lab 4 Measures of Central Tendency and Variation*'s instructions on how to calculate these measures.

Chapter 6

Pie and Pareto Charts

In this lab, you will generate pie and Pareto charts using Excel.

A *pie chart* is a tool to display basic statistical information for qualitative data, and is one of the easier charts to make in Excel.

A *Pareto chart* is a bar graph that visually depicts which situations are more significant.

6.1 Preparation

In this lab, you will use a data set that provides a snapshot of a sample Netflix user base. The data set includes information such as the user's subscription type, device type, and the country in which they are located.

Download the data set called *Lab 6 Netflix Sample.xlsx* from here: **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-6-Netflix-Sample.xlsx**.

6.2 Pie Charts

A *pie chart* (or a circle chart) is a circular statistical graphic that is divided into slices (or wedges) to illustrate numerical proportions.

1. Open a new Excel worksheet and enter *Earned Degree* data, as shown below.

2. Select the cells A1:B6.
3. Go to the `Insert` tab. Select the `leftmost 2D Pie diagram` in the top row in the `Charts` ribbon.

	A	B
1	Type of degree	Number (in thousands)
2	Associate's	525
3	Bachelor's	854
4	Master's	390
5	First professional	65
6	Doctoral	74
7		

Figure 6.1: Earned degree data.

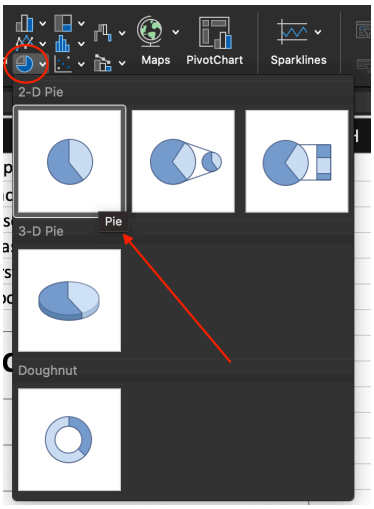


Figure 6.2: Pie chart options in the Insert tab.

4. If your graph does noy appear with a title and legend, do Steps 5 and 6 below.
5. When you select the graph, the tab `Chart Design` appears in the menu bar (you must select the graph for this tool to appear).
6. Choose Style 3 for your pie chart.

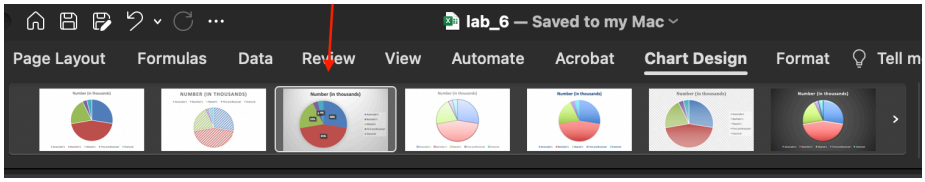


Figure 6.3: Options in the Chart Design tab for pie charts.

Note: Style 3 includes titles, legends, and labels in percent form. Check other styles to see the differences among them.

7. Replace *Number (in thousands)* with the title *Earned Degrees Conferred in 2000*.

Note: You can `RIGHT-click` on one of the data labels (e.g., on the percent of 45%) and then select `Format Data Labels` to change the format of the labels in your graph.

6.2.1 Practice 1

Use Excel to build a pie chart for the data containing *Device Type* in the *Lab 6 Netflix Sample.xlsx* file.

6.3 Pareto Charts

A *Pareto chart* is a bar graph with the bars arranged from the tallest bar to the shortest bar. Because of that, we need first to put the data in descending order.

6.3.1 Sorting Data

1. Highlight the cells A1:B6 again.
2. `Copy` (`CTRL` + `C`) the data, click on cell E1 (or any empty cells of the worksheet), and `Paste` (`CTRL` + `V`) the data.
3. Highlight the cells E1:F6 (or the cells that contain the copy of the data). Select `Custom Sort` in the *Sort & Filter* ribbon in the `Home` tab in the menu bar.
4. In the Sort dialog window that appears. Select the following options:
   - *Sort by* -> `Number (in thousands)`
   - *Sort on* -> `Values`
   - *Order* -> `Largest to Smallest`

This will sort both highlighted columns concerning the highest frequencies.

5. Check the box to the left on `My list has headers`. Click `OK`.

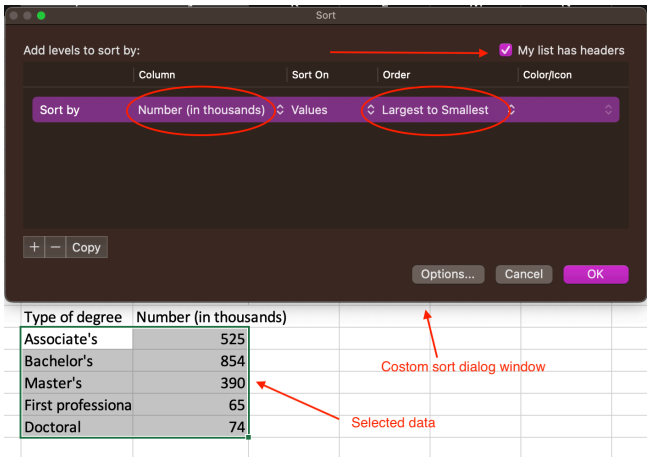


Figure 6.4: Custom sort options..

6.3.2 Building the Pareto Chart

1. Select the cells E1:F6 (or the cells containing the copy of the data) that have been sorted.
2. Go to `Insert > Column chart` and choose the leftmost top type.
3. Change the title of the graph to *Earned Degrees conferred in 2000*.
4. With the graph selected, go to the `Chart Design` tab.
5. To add axis titles, click `Add Chart Element > Axis Titles > Primary Horizontal` and then `Add Chart Element > Axis Titles > Primary Vertical`.
6. Replace the horizontal axis title with the *Type of Degree*.
7. Replace the vertical axis title with *In Thousands*.

6.3.3 Practice 2

Use Excel to build a Pareto chart for the data containing *Country* in the *Lab 6 Netflix Sample.xlsx* file.

Chapter 7

Measures of Location

Measures of location are numbers that indicate the position of data values in a data set. In this lab, you will use Excel to determine measures of location in a distribution: percentiles and standard scores (z-scores).

A *percentile* represents the location of a value as a cumulative percentage. The k-th (where 1 k 99) percentile is the data value that indicates the k% percentage of all values in the data set that falls at or below it, and (100-k)% of all data falls above it.

Another common measure of location is the standard value score, which relates to the distribution's spread. A *standard score* (or *z-score*) indicates the position of a value in terms of its distance to the mean in units of standard deviation.

7.1 Preparation

You will use a data set that provides measures of global atmospheric concentrations of carbon dioxide from 2002 to 2019.

Download the data set called Lab 7 Carbon Dioxide Concentrations.xlsx from here **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-7-carbon-dioxide-concentrations.xlsx**.

Open the Excel file *Lab 7 Carbon Dioxide Concentrations.xlsx* and follow the instructions below.

7.2 Percentiles

The function `PERCENTILE.INC(array,k)` returns the k-th percentile of values in a data set, where k is in the range 0..1, inclusive, representing the decimal form

of the percentage. The function `PERCENTILE.EXC(array,k)` works similarly, but k is exclusive in the range 0..1.

Example: Suppose a data set contains the test scores for a class of 20 students, and you want to calculate the 80th percentile score. Assuming the test scores are in cells A1:A20, you would use the formula `=PERCENTILE.INC(A1:A20, 0.80)`. This formula will return the value at the 80th percentile in the data set, representing the score separating the top 20% of students from the rest.

The following are instructions for you to find all percentiles of the carbon dioxide measured in *Mauna Loa, Hawaii*, contained in the file you downloaded.

1. Go to the top of an empty column and type *Percentiles- Mauna Loa, Hawaii* as the column's header.
2. Below *Percentiles- Mauna Loa, Hawaii*, create a list of numbers from 1 to 99 (use the instructions in *Section **5.3.4*** from *Lab **5** Time Series and Measures of Central Tendency*).
3. In the column next to it, type the header *Hawaii Measure*.
4. Below *Hawaii Measure*, enter the formula `=PERCENTILE.INC($C$4:$C$21,L2/100)`.
5. Select the cell containing the formula in Step 4.
6. Position the mouse pointer in the lower right corner of the selected cell until it becomes a `+` sign and click-drag downward across the range that covers all the values of the *Hawaii measure*.

Note: In the formula above, the address of the cells containing the data set has dollar signs (`$`) before the row and the column number to keep the calculations from changing the location of the data set.

Note: `PERCENTRANK.INC` uses interpolation to find percentiles for values that are not originally in the data set.

7.3 Percentile Ranks

The function `PERCENTRANK.INC(array, x, [significance])` calculates a specific value's relative position within a data set, expressed as a percentile rank between 0 and 1, inclusive. *Significance* is an optional argument. that identifies the number of significant digits for the returned percentage value. If omitted, `PERCENTRANK.INC` uses three digits (0.xxx).

Example: Suppose that a data set contains the test scores for a class of 20 students, and you want to find the percentile rank of the value in cell A1. You would use the following formula `=PERCENTRANK.INC(A1)`. This formula will return the corresponding percentile of the value with three (3) decimal places.

1. Go to the top of another empty column and type *Hawaii Measure* as the header of the column.

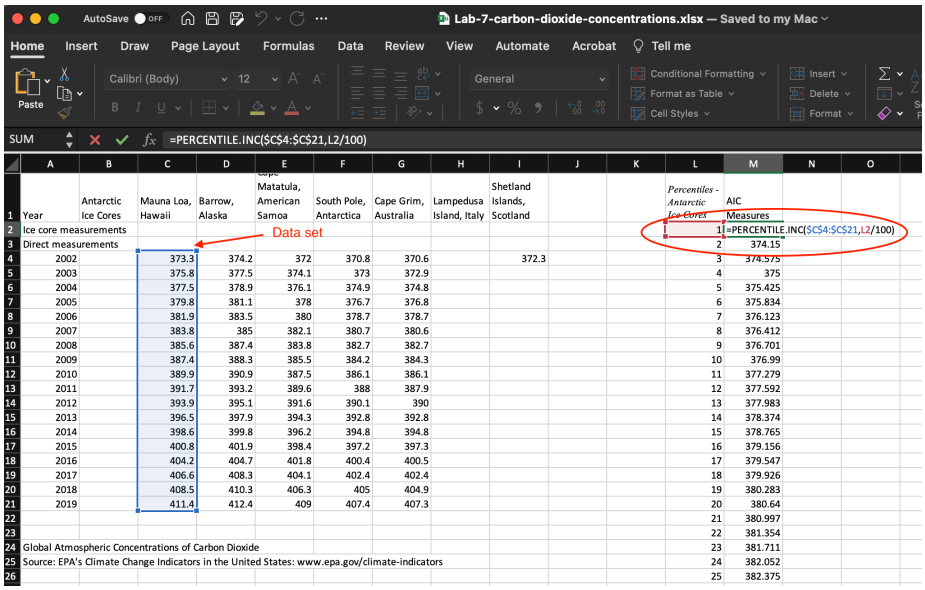


Figure 7.1: Percentiles calculations for Hawaii measures.

2. Copy the values in cells C4:C21 and paste them below the header *Hawaii Measure*.
3. Below *Percentile Rank*, enter the formula `=PERCENTRANK.INC ($O$2:$O$19,O2)`.
4. Select the cell containing the formula above.
5. Position the mouse pointer in the lower right corner of the selected cell until it becomes a + sign and click-drag downward across the range that covers all the values in Step 3. The results are in the decimal form.

Note: To convert the results to percentages, select the values in the column *Percentile Rank*, go to the `Home` tab, and then change the format to `Percentages` in the `Number` ribbon.

7.3.1 Practice 1

Use Excel to calculate the percentiles and percentile ranks of the values in the column *South Pole, Antarctica*, in *Lab 7 Carbon Dioxide Concentrations.xlsx*.

7.4 Standard Scores

The *standard score* (z) indicates the location of a value in the distribution as a function of how far the value is from the mean of the distribution in terms of the standard deviation.

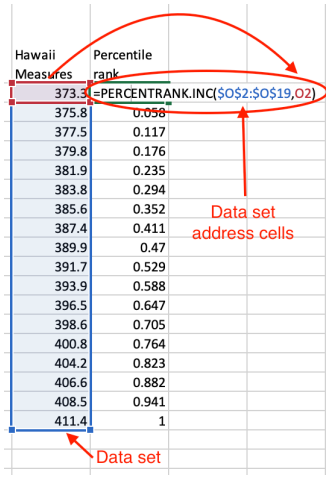


Figure 7.2: Percentile rank calculations for the Hawaii measures.

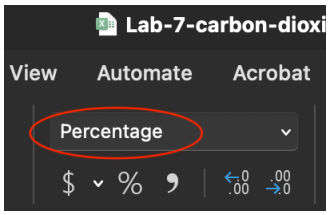


Figure 7.3: Numbers ribbon in the Home tab displaying the Percentage format.

7.4.1 Calculating the Mean and Standard Deviation

In this part of the lab, you will compute the mean and the standard deviation of the distribution of *Hawaii Measures*.

1. In an empty cell, type *Mean* and next to the cell, compute the mean using the function `=AVERAGE(C4:C21)`.
2. In another empty cell type *Stand Dev* and next to the cell, compute the standard deviation using the function `=STDEV.S(C4:C21)`.
3. In the `Number` ribbon, adjust the standard deviation value to have two (2) decimal places.

The result for the mean is 391.51, and the standard deviation is 11.78.

7.4.2 Converting to Z-Scores

The function `STANDARDIZE(x, mean, stand_dev)` returns a normalized value x from a distribution characterized by mean and standard deviation. The normalized value is the same as the z-score of the value.

1. Go to the top of an empty column and type *Z-Scores* as the column's header.
2. Below *Z-Scores*, use the function `= STANDARDIZE(C4, mean, stand_dev)` to convert the value in cell C4 in the data set. In the formula, replace mean with the cell address containing the mean value from calculations in *Section* ***7.4.1***, and replace `stand_dev` with the cell address with the standard deviation calculated. Include `$` in the mean and standard deviation addresses since they are fixed for all data values.
3. Adjust the z-score to have two (2) decimal places.
4. Select the cell containing the formula in Step 2.
5. Position the mouse pointer in the lower right corner of the selected cell until it becomes a `+` sign and click-drag downward across the range that covers all the *Hawaii measures*.

As a result, the z-scores range from -1.55 to 1.69.

7.4.3 Practice 2

Use Excel to calculate the z-scores of the values in the column *South Pole, Antarctica* in *Lab 7 Carbon Dioxide Concentrations.xlsx*.

Chapter 8

Box and Whisker Plot

Quartiles are quantiles (also percentiles) that divide the data into approximately four equal parts.

A *box and whisker plot* is a graph showing the data distribution using the *five-number-summary*: minimum value, first quartile, median (second quartile), third quartile, and maximum value. This type of graph is also known as a box plot.

The *interquartile range* (*IQR*) represents the spread of the middle half of the data and is calculated as the difference between the third and first quartiles.

8.1 Preparation

In this lab, you will use a data set containing NY temperatures in September of 2013.

Download the data set called *Lab 8 NY Temp Sept2013.xlsx* from here **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-8-NY-Temp-Sept2013.xlsx**.

Open the Excel file *Lab 8 NY Temp Sept2013.xlsx* and follow the instructions below.

8.2 Five-Number-Summary

Using Excel, you will compute the five-number summary, generate a box plot, and compute the interquartile range (IQR) for *Lab 8 NY temperatures in the September 2013* workbook.

Use Excel to compute the following in an empty part of the worksheet.

1. Choose a cell and type *Min*, and in the cell next to it, enter `= MIN(B3:B32)` to find the minimum value of the data set.
2. Below *Min*, type *Q1*, and in the cell next to it, enter `= QUARTILE(B3:B32,1)` to find the first quartile.
3. Below *Q1*, type *Q2*, and in the cell next to it, enter `= QUARTILE(B3:B32,2)` to find the second quartile.
4. Below *Q2*, type *Q3*, and in the cell next to it, enter `= QUARTILE(B3:B32,3)` to find the third quartile. 5, Below *Q3*, type *Max*, and in the cell next to it, enter `= Max(B3:B32)` to find the maximum value of the data set.
5. Below *Max*, type *Median*, and enter `= MEDIAN(B3:B32)` next to it to compute the median.

Be sure that your five-number summary is organized to display the results. For example:

Example of a five-number-summary

Measures

Values

Min

18.00

Q1

24.50

Q2

38.00

Q3

44.25

Max

58.00

Median

38.00

Note: The five-number summary for the NY Temperatures will differ from those shown in Table 1.

8.3 The Interquartile Range

Using Excel compute the *IQR* of the data set. Recall that the formula for *IQR* is `IQR = Q3 - Q1`.

Use an Excel formula to calculate the *IQR*. Your formula must use cell addresses in the calculation. Place your formula in empty cells below your five-number summary.

8.4 Box Plot

1. Select the cells with the temperature values in Column B.
2. Go to the Insert tab. Select `Box and Whisker` in the `Charts` ribbon (Figure **8.1**).

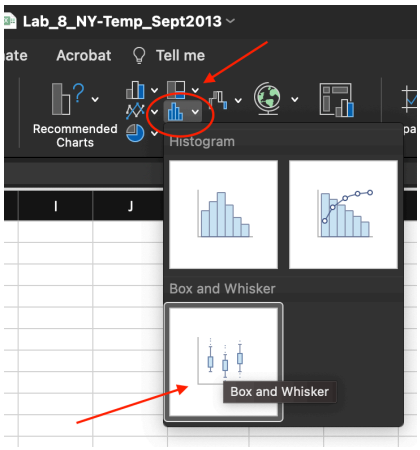


Figure 8.1: Box and Whisker option in the Chart ribbon of the Insert tab.

The resulting plot is shown in Figure **8.2**.

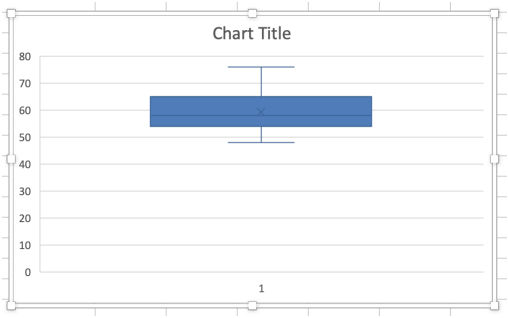


Figure 8.2: Box and Whisker created.

3. Replace *Chart Title* with *NY Temperatures in September 2013*.
4. `RIGHT-click` on the vertical axis and then select `Format Axis` to change the axis format in the graph. Set the *Minimum Bound* as 45. Adjusting the axis removes the unnecessary space in the chart.

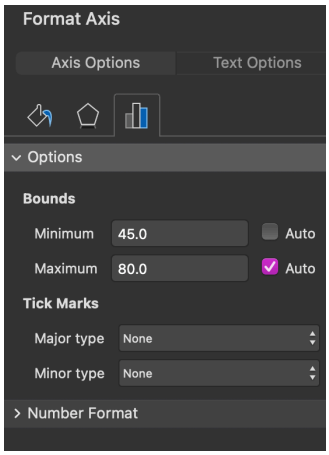


Figure 8.3: Format Axis options.

Use the `Chart Design` to customize the look of your chart.

5. Select the graph, and the tab `Chart Design` appears in the menu. Click on the `Chart Design` tab.
6. Click `Add Chart Element > Data Labels > Right`.
7. Click `Add Chart Element > Axes`. Uncheck the option `Primary Horizontal`.

The adjusted box plot is shown in Figure **8.4**.

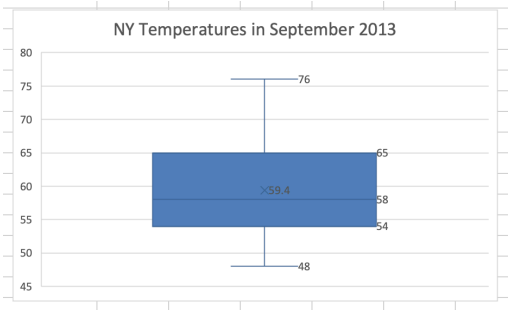


Figure 8.4: Adjusted box plot.

8.5 Practice

Use Excel to generate the five-number-summary, *IQR*, and box plot for the financial data you collected for *Lab **5** Time Series and Measures of Central Tendency*.

Chapter 9

Linear Regression and Correlation

Linear regression (also known as linear regression) is a statistical method used to determine whether a linear relationship (correlation) between variables exists. It is used to predict a variable's value based on another variable's value.

You will use Excel to determine the correlation of a data set with a pair of variables, graph its scatter plot, and determine the best-fitting line (linear regression equations).

9.1 Preparation

In this lab, you will use a file containing data about the eruptions of the Old Faithful Geyser in Yellowstone National Park. Each row in the file represents an observed eruption of the Old Faithful Geyser. The *Eruptions* column represents the duration of the eruption in minutes, and the *Waiting* column represents the duration in minutes until the next eruption.

Download the data set called *Lab 9 Old Faithful.xlsx* from here **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-9-Old-Faithful.xlsx**.

Open the Excel file *Lab 9 Old Faithful.xlsx* and follow the instructions below.

9.2 Creating a Scatter Plot

1. Select the cells in the Columns *Eruptions* and *Waiting*.
2. Go to the `Insert` tab. Select `Scatter` in the `Charts` ribbon. Select the leftmost chart in the top row by clicking on it (Figure **9.1**).

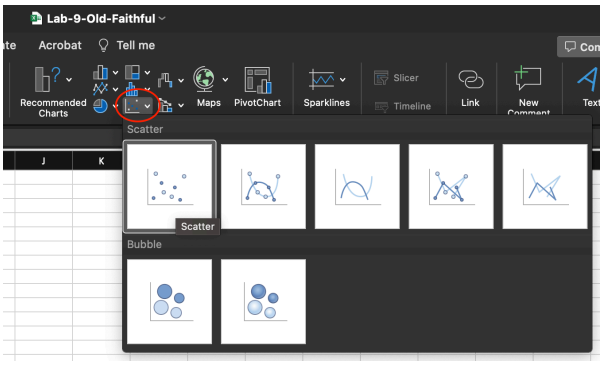


Figure 9.1: Scatter option in the Chart ribbon of the Insert tab.

The resulting plot is shown in Figure **9.2**. The `Chart Design` tab shows that the leftmost diagram style is selected (Style 1).

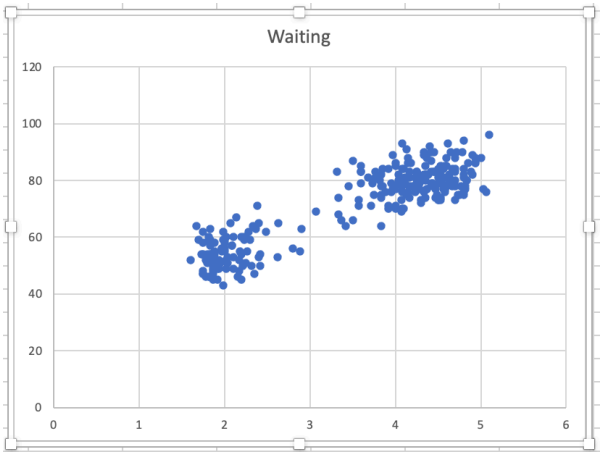


Figure 9.2: Scatter plot created.

3. For the title, replace *Waiting* with *Old Faithful Data*.
4. Click `Add Chart Element > Axis Title > More Axis Title Options`. This will add an automatic x-axis title and y-axis title.
5. Click the x-axis title and replace it with *Duration of Eruptions (in Minutes)*.
6. Click the y-axis title and replace it with *Time Until the Next Eruption (in Minutes)*.
7. `RIGHT-click` on the horizontal axis and then select `Format Axis` to change the axis format in the plot. Set the *Minimum Bound* as 1. Adjusting the axis removes
8. `RIGHT-click` on the vertical axis and then select `Format Axis` to change

the axis format. Set the *Minimum Bound* as 30 and *Maximum Bound* as 100.

The adjusted box plot is shown in Figure **9.3**.

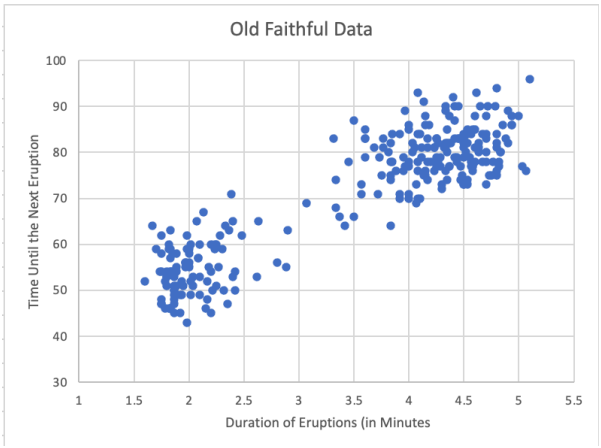


Figure 9.3: Adjusted scatter plot.

9.3 Calculating the Correlation Coefficient

1. Type the label *Correlation* in an empty cell in the worksheet.
2. In the cell below *Correlation*, go to the `Formulas` tab and click the `Insert Function` ribbon.
3. In the dialog window, search for the `CORREL` function. Click `OK`.
4. For `Array 1`, select the cells in Column A of the data set.
5. For `Array 2`, select the cells in Column B of the data set.

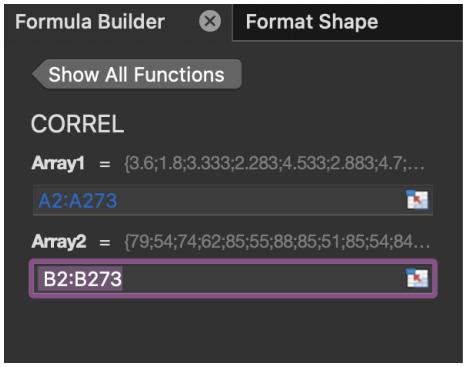


Figure 9.4: Dialog window of the Correl function on MAC.

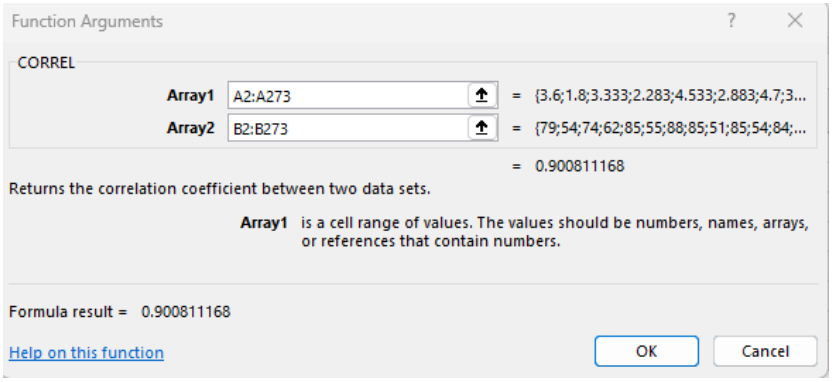


Figure 9.5: Dialog window of the Correl function on PC.

Up to 4 decimal places, the correlation coefficient is 0.9008.

##Determining the Regression Line Equation

The *regression line equation* (or *least squares line*) is a straight line that best fits the data set. The equation of the line has the form $\hat{y}-bx+a$ where b is the slope and a is the y-intercept. The regression line explains the relationship between the independent variable (explanatory variable) and one or more dependent variables (response variable(s)).

The `LINEST(known_y's, known_x's, [const], [stats])` function calculates the slope and y-intercept of the regression line. The last two arguments in the formula are optional.

1. Type the label *Regression Line* in an empty cell in the worksheet.
2. Type the formula '`= LINEST(B1:B273, A1:A273)` in the cell below *Regression Line*.

Note: Alternatively, go to `Formulas > Insert Function` and search for the `LINEST` function.

The results output by the `LINEST` function is a slope equal to 10.7296414 and a y-intercept equal to 33.47439702.

9.4 Graphing the Regression Line

1. `RIGHT-click` on any point plotted in the scatter plot (blue points).
2. Select `Add Trendline`.
3. Select the `Linear` trend.
4. Check the box to `Display Equation on chart`.

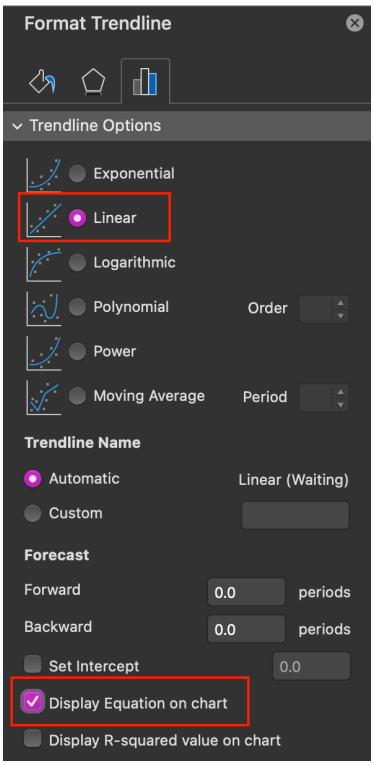


Figure 9.6: Trendline options.

9.5 Practice

For this exercise, you will switch the explanatory and response variables. That means changing the roles of x and y values for the same data set. Following the guidelines above, use Excel to:

a. Construct the scatter plot.
b. Calculate the correlation coefficient.
c. Determine the regression line's slope and y-intercept.
d. Draw the regression line in the scatter plot.

Chapter 10

Coin Toss Simulation

In *Lab 2 Random Number Generators and Tables*, the function `RANDBETWEEN(Bottom,Top)` was used to obtain a random integer between (and including) the bottom and top numbers.

You will use the random number generator to simulate probability experiments such as tossing coins. The goal of this lab is to verify the Law of Large Numbers. Recall that the *Law of Large Numbers* states that in the long run, as the sample size increases and increases, the relative frequencies of outcomes get closer and closer to the theoretical or actual probability values. The goal of this lab is to verify the Law of Large Numbers.

10.1 Simulating a Coin Tossing Experiment

To simulate the experiment, assign to the outcome *Head* the digit *1* and to *Tail* the digit *2*.

You will draw a random sample of size 30 in Column A from the distributions of integers with a minimum value of 1 to a maximum value of 2.

1. Write the label *Coin Toss* in cell A1.
2. Type the formula `= RANDBETWEEN(1,2)` in cell A2 and press `Enter/Return`.
3. Select cell A2 again, position the mouse pointer in the lower right corner of the selected cell until it becomes a `+` sign, and click-drag downward across until cell A31.
4. Ensure that Excel randomizes all the values by going to the `Formula` tab in the menu bar and clicking the `Calculate Now` ribbon to apply the randomization. Column A should now have 30 random values.

10.2 Counting the Number of Heads and the Number of Tails

Next, you will use Excel to count the *1*'s and *2*'s in column A. You will set up a table in Columns C and D to display the results.

1. Type the label *Outcome* in Cell C1.
2. Type the label *Heads* in Cell C2.
3. Type the label *Tails* in Cell C3.
4. Type the label *Frequency* in Cell D1.

You will use the formula `COUNTIF(data range, condition)` to count the *1*'s and *2*'s as the formula's conditions. Recall that we assigned the number *1* to the outcome *Head* and the number *2* to the *Tail*.

1. Select cell D2, and enter the formula `= COUNTIF(A2:A31,1)`.
2. Next, select cell D3, and enter `= COUNTIF(A2:A31,2)`.

10.3 Calculating the Empirical Probability

Empirical Probability is based on experience or historical data to determine the likelihood of outcomes. It is calculated by determining the relative frequency. You will calculate the observed probabilities of the outcomes *Heads* and *Tails* you simulated.

To calculate the relative frequency of an outcome, calculate the number of times the outcomes appear divided by the total number of all outcomes.

1. Enter the label *Rel Freq* in cell E1.
2. In cell E2, enter `= D2/30`.
3. In cell E3, enter `= D3/30`.
4. Select the cells E2 and E3 and go to the `Home` tab, and then adjust the results to have four (4) decimal places in the `Number` ribbon.

10.4 Practice

1. Repeat this process to generate another simulation with 100 *Coin Tosses* in any empty space of the *Lab 9* worksheet.
2. Repeat this process to generate another simulation with 500 *Coin Tosses* in any empty space of the worksheet.

Chapter 11

Normal Distributions

A *normal distribution* is a continuous probability distribution for a random variable that describes the frequency of the outcomes of the variable. The mean and the standard deviation parameters determine the normal distribution.

In this lab, you will use the `AVERAGE` and `STDEV.S` functions to calculate the data's mean and standard deviation, respectively.

The `NORM.DIST(x,mean,standard_dev,cumulative)` function returns the normal distribution for the specified mean and standard deviation. In this case, we will use the *false* argument for *cumulative* since we will get the corresponding probability for the respective x values.

The `NORM.INV(probability,mean,standard_dev)` function returns the z-score in a normal cumulative distribution for the specified mean and standard deviation.

The `STANDARDIZE(x, mean, standard_dev)` function returns a normalized value from a mean and standard deviation distribution.

11.1 Preparation

You will use a data set on baby birth, including the variables birthweight, mother's age, mother's weight, mother's height, and whether the mother is a smoker. There are one thousand (1000) cases included in this data set.

Download the data set called *Lab 11 Babies.xlsx* from here **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-11-Babies.xlsx**.

Open the Excel file *Lab 11 Babies.xlsx* and follow the instructions below.

11.2 Calculating the Mean and Standard Deviation

First, you will make a copy of the data on birthweight and sort the values from the smallest to the largest. Then you will calculate the mean and standard deviation. These Excel functions were covered in *Lab* **4** *Measures of Central Tendency and Variation*.

1. Select the cells in Column B (cells B1:B1001).

Note: A shortcut to select data in a range, click on the first cell in the range and press the `Shift` + `Arrow Down` keys to select a range of data.

2. `Copy` (`CRTL-C` on PC or `Command-C` on MAC) the selected data, click on cell J1 (or any empty cells of the worksheet), and `Paste` (`CRTL-V` on PC or `Command-V` on MAC) the data.
3. Select the cells J1:J1001 (or the cells that contain the copy of the data). Go to `Sort & Filter` ribbon in the `Home` tab and select `Sort Smallest to Largest`.
4. Type the label *Mean* in cell H1.
5. Select cell H2, then enter the formula `= AVERAGE(J2:J101)`.
6. Adjust the mean value to one (1) decimal place.
7. Type the label *Stand Dev* in cell H3.
8. Select cell H4, then enter `= STDEV(J2:J101)`.
9. Adjust the standard deviation value to two (2) decimal places.

The data's mean is 119.9, and the standard deviation is 17.99.

11.3 Graphing the Normal Distribution

The `NORM.DIST(x,mean,standard_dev,FALSE)` function calculates the normal probability density function that can be used to plot the bell-shaped curve. To graph the normal curve, we need x and y values. The x values correspond to the birthweights. The y values are values of the probability density functions.

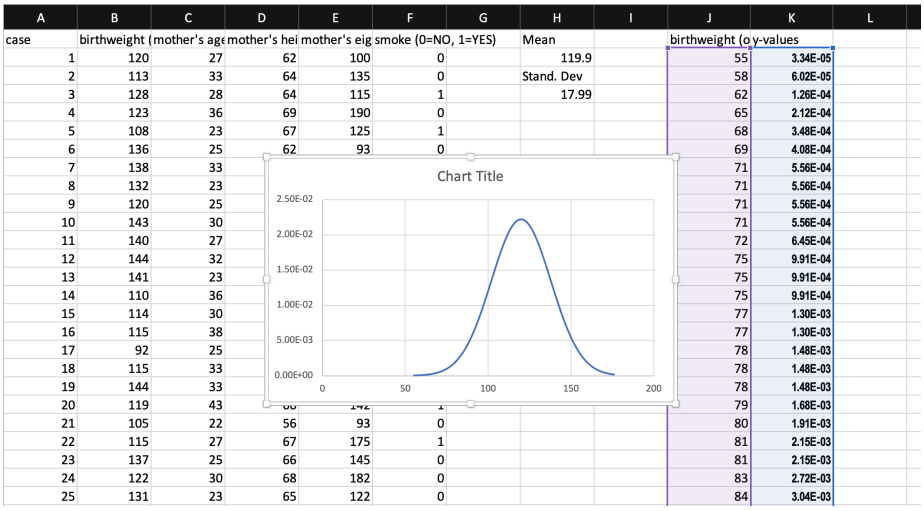
First, you will generate the y-values and then use a scatter plot to display the bell-shaped curve.

1. Select the cell K1 and type the label *Y-Values*.
2. Below *Y-Values*, enter the formula `= NORM.DIST(J2,$H$2,$H$4, FALSE)`.
3. Select cell K2 and place the mouse pointer in the lower right corner of the selected cell until it becomes a `+` sign, and click-drag downward across the range that covers all the values in column K.

The y-values range from 3.34 x 10^{-5} to 1.71 x 10^{-4}.

4. Select all cells in the *birthweight* and *Y-Values* columns (select the first cells in columns J and K, then press the `Shift` + `Arrow Down` keys to select the whole range of data).

5. Click `Insert > Scatter > Scatter with Smooth Lines`.



case	birthweight (	mother's age	mother's hei	mother's eig	smoke (0=NO, 1=YES)	Mean	birthweight (o	y-values
1	120	27	62	100	0	119.9	55	3.34E-05
2	113	33	64	135	0	Stand. Dev	58	6.02E-05
3	128	28	64	115	1	17.99	62	1.26E-04
4	123	36	69	190	0		65	2.12E-04
5	108	23	67	125	1		68	3.48E-04
6	136	25	62	93	0		69	4.08E-04
7	138	33					71	5.56E-04
8	132	23					71	5.56E-04
9	120	25					71	5.56E-04
10	143	30					71	5.56E-04
11	140	27					72	6.45E-04
12	144	32					75	9.91E-04
13	141	23					75	9.91E-04
14	110	36					75	9.91E-04
15	114	30					77	1.30E-03
16	115	38					77	1.30E-03
17	92	25					78	1.48E-03
18	115	33					78	1.48E-03
19	144	33					78	1.48E-03
20	119	43					79	1.68E-03
21	105	22	56	93	0		80	1.91E-03
22	115	27	67	175	1		81	2.15E-03
23	137	25	66	145	0		81	2.15E-03
24	122	30	68	182	0		83	2.72E-03
25	131	23	65	122	0		84	3.04E-03

Figure 11.1: The normal curve created.

The resulting graph needed to be adjusted (Figure **11.1**). Use the instructions in Steps 3-8 in *Section **9.2** Creating a Scatter Plot* to make the necessary adjustments to the graph. Adjust the title, axis titles, and horizontal axis.

11.4 Calculating Probability or Cumulative Area

We can use the `NORM.DIST(x,mean,standard_dev,TRUE)` function to calculate the probability that variable x falls below or at a specified value. Recall that a percentile in a normal distribution also corresponds to the area to the left (cumulative area) of a specific value.

1. Select cell L1 and enter the label *Percentile*.
2. Go to cell L2 and enter the formula `= NORM.DIST(J2,$H$2,$H$4,TRUE)`.

Note: Remember to include the `$` symbol.

3. Select cell L2 and place the mouse pointer in the lower right corner of the selected cell until it becomes a `+` sign, and click-drag downward across the range that covers all the values in column K.
4. In the `Home` tab, select `Percentage` (%) in the `Numbers` ribbon to convert the values in Column L to percentages.
5. Adjust the probability (or percentiles) values to two (2) decimal places.

The data's percentiles should range from 0.02% to 99.91%.

11.5 Finding Raw Scores from Probabilities

The function `NORM.INV(probability,mean,standard_dev)`, where *probability* is in the percentage form, will provide the raw value corresponding to the probability (or percentile).

1. Select cell M1 and type the label *90th Percentile*.
2. In cell M2, type the label *95th Percentile*.
3. In cell M3, type the label *99th Percentile*.
4. In cell N1, type `= NORM.INV(90%, $H$2,$H$4)`.

Note: Remember to include the `$` symbol.

5. Repeat Step 4 above to obtain the other percentiles, 95th in cell N2 and 99th in cell N3.

11.6 Converting to Z-Scores

We can use `STANDARDIZE(value, mean, standard_dev)` to convert any value to a z-score when the mean and standard deviation are given.

1. Select the cell K1 and type the label *Z-Scores*.
2. Below *Z-Scores*, enter the formula `= STANDARDIZE(J2,$H$2,$H$4)`.

Note: Remember to include the `$` symbol.

3. Select the cell with the z-score calculated, K2.
4. Place the mouse pointer in the lower right corner of the selected cell until it becomes a `+` sign, and click-drag downward across the range that covers all the values in column J.
5. Adjust the z-scores values to two (2) decimal places

11.7 Practice

For the variable *mother's age*, perform the following tasks. 1. Graph the normal curve for the distribution. 2. Find the distribution's probability. 3. Find the *88th Percentile*. 4. Find the z-scores.

Chapter 12

The Central Limit Theorem

The *sampling distribution* of sample means is a distribution formed by means of all samples of size n repeatedly taken from a population. If a population has mean μ and standard deviation σ, and all possible samples of size n are taken from the population then sampling distribution satisfies the following properties $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

The *Central Limit Theorem* states that the sampling distribution of sample means approaches a normal distribution as the sample size gets larger no matter the population distribution's shape. This is especially true for sample sizes over 30. The greater the sample size, the better the approximation.

If the population itself is normally distributed, then the sampling distribution of sample means is normally distributed for any sample size n.

In this lab, you will simulate the sampling distribution by generating some of the samples of a specific size and comparing the results with the theoretical values for actual sampling distribution.

12.1 Preparation

You will use the *Lab 12 FB Daily.xlsx* that contains some of the daily Facebook stock prices.

Download the data set from here **https://github.com/bsosnovski/Intro-Stats-Excel-Lab-Manual/blob/main/Data_Sets/Lab-12-FB-Daily.xlsx**.

Open the Excel file *Lab 12 FB Daily.xlsx* and follow the instructions below.

12.2 Mean and Standard Deviation of the Population

Assume that the data contains all the *Close Prices* of the FB stocks (it is data from a population). You will calculate the mean and standard deviation of the populations of FB *Close Prices*.

1. Go to an empty column in the *Lab 12 FB Daily* worksheet, and type the label *Population*, *Mean* and *Stand Dev* in separate cells.
2. Select the cell next to the cell *Mean* and enter the formula `= AVERAGE(F2:F389)` to compute the mean.
3. Select the cell next to the cell *Stand Dev* and enter `= STDEV.P(F2:F389)` to compute the data's standard deviation of the population.
4. Adjust the mean and standard deviation values to two (2) decimal places.

12.3 Generating Random Samples

You will simulate the sampling distribution by generating only 10 random samples of *Close Prices* with a sample size equal to 20.

1. Go to an empty column and type the label *Sample 1*.
2. Below *Sample 1*, click `Data > Data Analysis > Sampling`.
3. Click `OK`.
4. In the window that appears, enter the following:
   - Input Range: Select cells `F2:F389`.
   - Sample method: Check `Random`.
     - Number of Samples: `20`.
   - Output Options: Click `Output Range`.
     - Select below *Sample 1* to paste the sample created..
5. Click `OK`.
6. Repeat the process for *Sample 2* through *Sample 10*.

12.4 Sample Means

Now you will calculate the mean for each sample you generated.

1. For each sample, place its sample mean below the column that contains the sample (use a similar formula to the one in *Section **12.2** Mean and Standard Deviation of the Population*).
2. Highlight the results to distinguish them from the values in the sample.

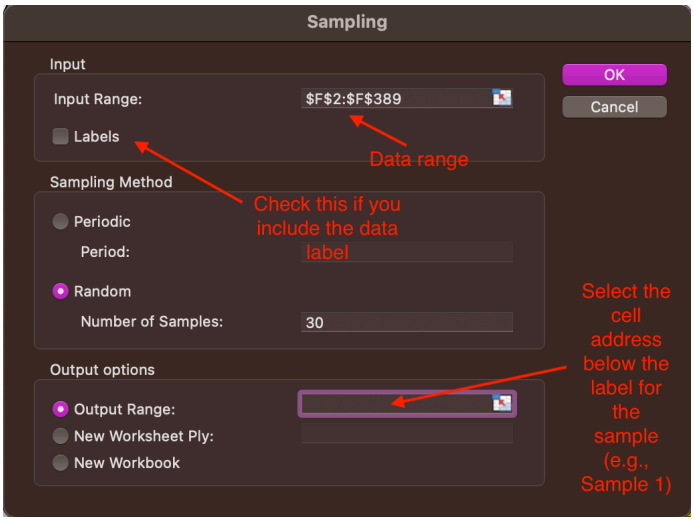


Figure 12.1: Sampling window in Data Analysis ribbon

J	K	L	M	N	O	P	Q	R	S	T	U	V
Population			Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	sample 8	Sample 9	Sample 10
mean	30.30		27.4	24.1	52.68	23.85	19.52	23.15	28.31	28.75	52.68	44.31
stand deviation	9.25		29.73	21.09	25.73	45.83	21.6	20.38	50.28	37.08	50.23	46.43
			27.1	37.08	19.75	27.4	39.64	27.57	19.21	27.98	20.07	20.07
			38.32	37.63	28.26	19.32	26.28	47.14	24.55	22.56	38.55	38.55
			27.77	27.71	19.99	20.47	47.49	27.72	25.87	20.07	24	45.83
			26.81	28.45	25.81	21.95	45.98	52.68	29.06	21.09	18.81	18.81
			30.98	21.92	21.11	23.15	25.86	25.13	24.31	44.31	44.31	44.31
			29.85	28.45	26.31	22.92	24.32	26.25	27.4	31.6	18.81	18.81
			18.81	30.59	30.1	21.83	49.01	34.36	45.23	31.47	31.47	31.47
			20.04	26.32	25.16	23.73	21.2	24.31	19.52	30.01	45.04	45.04
			28.75	19.34	30.59	21.87	38.22	27.43	53.85	27.27	25.16	25.16
			25.58	41.28	25.8	27.4	21.6	24.21	49.01	30.79	50.42	50.42
			28.76	38.32	33.05	25.76	25.76	19.52	25.42	25.8	29.06	29.06
			41.34	49.01	31.73	28.75	18.98	30.82	28.76	50.39	26.85	26.85
			28.75	21.18	30.59	38.41	28.25	31.84	19.05	23.73	29.42	29.42
			20.38	44.31	44.75	28.19	19.32	20.4	46.77	25.86	33.1	50.23
			19.32	20.88	25.53	26.85	24.03	27.08	26.85	46.61	26.32	26.32
			27.32	31.73	32.06	52.68	30.59	26.85	27.98	24.32	25.42	25.42
			54.22	50.28	34.01	26.05	30.01	28.25	49.12	24.21	50.21	44.82
			25.87	26.18	32.47	19.48	28.45	41.78	28.75	47.01	25.94	25.94
		sample means	28.86	31.29	29.77	27.29	29.31	29.34	32.47	31.05	33.29	34.36

Simulation of the sampling distribution

Figure 12.2: An example of the sampling mean distribution generated.

12.5 The Mean and Standard Deviation of the Sampling Distribution

1. After *Section **12.4*** above, all the ten (10) sample means should be on the same row (Figure **12.2**). From these sample means calculate the mean and standard deviation (use `STDEV.S`) for the FB *Close Prices* in an empty part of the worksheet.
2. Calculate the population standard deviation value in *Section **12.2*** divided by $\sqrt{20}$ (square root of the sample size).

Recall that the theoretical value of generating all the samples of size 20 should be $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Compare the results in Step 1 with the results obtained in Step 2.

12.6 Practice

1. Repeat the process above with 10 samples of size 50 for the FB *Close Prices*.
2. Repeat the process above with 20 samples of size 50 for the FB *Open Prices*.