Intro to Statistics — Excel Worksheet

Warning: We have chosen to use Excel because it is easily available on campus and because the data is easily visible on the screen as you use it. But it has errors (different versions have different errors) in its statistical functions, so for any serious purpose we urge you to learn and use a real statistical package, such as SPSS, SAS, Minitab, etc., etc.

1. Spreadsheet **cells** (those little rectangles) can hold numbers, words, or formulas (which then produce numbers). Try filling in some cells with numbers and others with words (click on it to highlight it and start typing your entry).

   - Use "Enter" or "Return" to finish an entry and move down. "Tab" moves you to the right. "Shift-Enter" or "Shift-Return" moves you up, and "Shift-Tab" moves you left. You can also use the mouse to click on a cell that you wish to change, or you can use the arrow keys to move around in the spreadsheet.
   - What happens if your words are longer than a cell? Try this with the cell to the right being empty and again with it having some entry.
   - Select a cell and change its format with the "Format" menu.

2. Scrolling works as you might expect. Enter the value 23 into cell D104 (use the column and row headings to find cell D104). Then come back to the top of the spreadsheet.

3. **OK... Now we're ready to try entering formulas.**
   All formulas start with an equals sign "=". Select cell D4 and enter the formula "=3+2*D104" (without the quotes). The value 49 (which is $3 + 2(23)$) should appear. When you select a cell, the formula that is actually in the cell appears in the line at the top of the screen. (What appears in the cell is the present value of that formula.) Select cell D4 again and click in the line at the top to edit it to be "=5+2*D104". The value should now appear as 51.

4. **Copying/Cutting and Pasting:**

   - Clear out column A. (Select and "Edit→Clear→All" or "Edit→Delete" the cells you don't need). Enter the number "1" in cell A1 and the formula "=A1+1" in cell A2.
   - Select cell A2 and move the cursor to the lower right corner of that cell. The cursor (which is usually a white plus sign) should change to either a "black plus sign" or a "box with arrows" depending on your version of Excel. Using that special cursor, drag the cell down to A10 so that the numbers "1" through "10" appear. This dragging "copies" the formula to the cells below. Check the formulas in this column and figure out how Excel changed the formula during the copying.
   - Now select the entire column of ten numbers and, using the special cursor from the lower right hand corner, copy them to the two columns to the right of the first. What happened to the formulas this time?
   - To move a column, you should "cut" and "paste". Select all three columns. Select the "Edit→Cut" menu. The border of the cells starts to shimmer. Select cell A12 and choose the "Edit→Paste" menu. What happened to the formulas?

5. **What if you don't want the formulas to change?**

- You can protect a cell name in a formula from being changed when the formula is moved or copied by using a dollar sign in front of the column letter or row number or both. For example, select cell C13. Change the formula to "=C12+1", and then copy it down the column. All the numbers in the column after C12 should now be the same; the formula doesn't change when you copy.
- Enter the value 2 into cell C10. Change the formula of C13 to "=C12+C$10" and copy it down the column. Now the column counts by twos. Changing cell C10 automatically changes all the other cells to count by whatever value you want. Try counting by threes.

6. **Formulas on Ranges:**
Some functions use a range of values. For example, if we want to sum the ten values in the first column. Enter the formula "=sum(A12:A21)" into cell A10. It should sum those ten numbers.

   - You can also use the "Paste Function" button at the top of the screen. It looks like f_x in fancy letters. Select cell B10. Push the "Paste Function" button. A list of functions should appear. Choose the function you want (sum). (Actually, you will see two lists. The first is a list consisting of categories of functions; the default is the functions that were Most Recently Used, so if Sum hasn't be used recently, it may not be in the second list. But other categories are All and Statistical, either of which should include all the Excel functions needed in the course.) Excel will then prompt you to enter the range of cells you wish to use. You can enter them by hand, **OR** click on the red arrow symbol to the right of the blank. This allows you to simply select the cells you want in the range. Highlight the cells B12 through B21 and then find the red arrow again to get out of selection mode. Click OK and the function is pasted in.
   - In a new column, enter five arbitrary numbers. Find the average, median, max, min, stdev and stdevp(=SD) of those numbers. (The stdev is what our text calls the SD^+ — it will appear late in the course — and other texts call the sample standard deviation. It is a little larger than the SD, which other texts call the population standard deviation; hence the "p" in stdevp.) It is a good idea to label the functions you use, so type the labels "average", etc. next to the values you just computed.
   - Go back and change one number of the original column and observe how that affects the average and the SD.

7. **Using Web Data:**
Go to the website http://math.colgate.edu/math102/examples/econlie.html Copy the data (and labels) and then go to Excel and Paste it into a column (say, column A) of your worksheet. (Use the Edit menu.)

   - In Excel, move the Unemployment data (with label) next to the Inflation data (say, into column B).
   - Create a third column labelled Misery Index (say, in column C) which is the sum of the other two. (If the inflation data began in cell A3 and unemployment in B3, then you might put "=A3+B3" into cell C3 and copy down.)

8. **Making a Histogram Frequency Table and Chart**
We now want to make a histogram of the Misery Index data. To make a histogram, Excel needs to know the edges of each bar so it can sort the numbers into the correct "bins". Excel calls this list of edges the "Bin Range". For this exercise, we want a histogram with interval ranges 5-10,

10-15, 15-20 and 20-25, so the "Bin Range" will be a column of the numbers 10, 15, 20, 25. (We'll assume that none of the numbers will be below 5, so we don't need the 5.) Create this "Bin Range" (say in cells D2:D5). Then we can proceed in one of two ways; the first is easier to use but doesn't change if the data changes, while the second is harder to use but gives results that change with the data. Here is the first way:

- Use the "Tools→Data Analysis" menu to select the "histogram" tool. (If "Data Analysis" does not appear under your tools menu, then proceed as follows: Use "Tools→Add-Ins", and in the resulting window check the boxes for the two Analysis Toolpaks — probably the first two in the list — and click on Okay. There may be a short pause before the window disappears, but when it does, "Data Analysis" should appear in the Tools menu. If Excel asks you to put the Excel program disk into the CD drive, and you don't have it, you'll have to use the other method to create histograms.) The histogram tool asks for two ranges: The "Input Range" and the "Bin Range". The Input Range is the location of the Misery Index values, and the Bin Range is the location of the 10,15,20,25 you typed in (say D2:D5). It also asks you for an Output Range. Choose an output range on the current worksheet (not a new ply) — it is enough to put in the label of the upper left cell of the region where you want your results to appear. Create a frequency table for the Misery Index data using the histogram tool. Follow the directions on the screen.
- Where does the entry 20 get counted? Where does the entry 27 get counted? Why is there one more "frequency" value than "bin" value?
- Note that you could now plot the histogram using the "Chart Wizard"; the method is outlined in the description of the second method of creating a histogram, below. But you don't need to do it here — see the next point.
- Use the "histogram" data analysis tool again to create the frequency table and a histogram chart at the same time. To produce the chart, check the "Chart output" option before pressing the "OK" button — otherwise only the frequency table will be created.

Here is the second way to make a histogram:

- Highlight a range of cells, one more than in the Bin Range. (For example, if you typed 10, 15, 20, 25 into cells D2:D5, you might highlight cells E2:E6.)
- Now, supposing that the Input Range (the Misery Index) is in cells C2:C18 and the bin range is D2:D5, type "=frequency(C2:C18,D2:D5)" but <u>don't</u> just hit Enter. Instead, hold down the Control and Shift keys <u>while</u> hitting Enter. This step is why the frequency command is harder to use than the Histogram command. You can't use the Paste Function button or click-and-drag, because for some reason they don't work correctly with the frequency function; you must actually type it all out — and if you make a mistake, you must start over with the highlighting. (You may need to hit the Escape key to get out of a frustrating loop: "There is an error in the array function", "You cannot change part of an array".)
- The last step is much easier: Use the Chart Wizard (the blue/yellow/red "column graph" icon at the top of the spreadsheet) to make a column chart of the frequency values. To specify that the Misery Index column is to be graphed in the Y direction and that the Bin Range is to be the labels on the X-axis: The second page in the Chart Wizard includes a tab labelled "Series". Click on it, and add the location of frequency values as Series 1 (it may already be there, if it was highlighted when you entered the Chart Wizard) and the Bin Range as the "Category (X) axis labels", using the boxes with red arrows. (It would actually be better to

label the bars as, for example, "5-10" instead of just "10"; but you would have to type the list of such labels separately — and before typing them, you would first have to format the cells as Text, because Excel's default is to assume that "5-10" means the date May 10th.)

9. **Other Types of Charts:**
   Use the Chart Wizard to insert a Chart of type "(XY)-scatter" with the Misery Index as the X–data series and Unemployment as the Y–data series.

10. **Print Only What You Want:**
    When you want to print out your results on paper, for example to hand them in, you rarely want to print all the data; rather, you just want the results. To print part of the spreadsheet, highlight a rectangle of cells that contains what you want, and use "File→Print Area→Set Print Area". To see what will print before actually printing it, use "File→Print Preview". If your data is wide but not tall, it may be helpful to use "File→Page Setup" and change the layout from Portrait to Landscape to get it onto fewer pages.