

## Session I: Introduction to Excel

In this session, we will open, build and read worksheets, operate functions and formulas, do basic statistical analysis, and chart, edit, and print data. These data are from the 1996 *Economic Report of the President* on civilian unemployment rates and consumer prices.

Year	Unemployment Rate (%)	Inflation Rate (% , Dec. To Dec.)
1975	8.5	6.9
1976	7.7	4.9
1977	7.1	6.7
1978	6.1	9.0
1979	5.8	13.3
1980	7.1	12.5
1981	7.6	8.9
1982	9.7	3.8
1983	9.6	3.8
1984	7.5	3.9
1985	7.2	3.8
1986	7.0	1.1
1987	6.2	4.4
1988	5.5	4.4
1989	5.3	4.6
1990	5.5	6.1
1991	6.7	3.1
1992	7.4	2.9
1993	6.8	2.7
1994	6.1	2.7
1995	5.6	2.5

**I. Click Start button, choose Programs, click Microsoft Excel, and when Excel is running, enter the data in the table above. (In some PC-Labs, you must enter Excel from Office97 to get the latest version). After entering the text or number in a cell, press Enter or use the arrow key to move to another cell. If you start in location (cell) A1, your last CPI number, 2.5, should be in C22.**

**Toolbars:** The top left corner is the Excel window. Click here and choose Close to end Excel, or use Windows tools like File, Edit, etc. The toolbar beneath is the Standard toolbar, where 1 click opens files, saves, prints, cuts, copies and pastes, sums, and sorts. Underneath the Standard toolbar is the Formatting toolbar, whose buttons select fonts, add bolding, format numbers, etc. Under the Formatting toolbar in about the right 3/4ths is the formula bar, where you edit formulas and data. You can also edit in individual cells. A cell (the location column A, first line is cell A1, etc) is where you enter numbers, text, dates, formulas, and all kinds of data. Columns are letters, rows are numbers.

**File:** Save the worksheet by clicking on the Save (diskette) button on the toolbar, which will pop up a Save As dialog box. Type a name (best to choose something that will describe the contents such as Unemp) in the File Name text box, check to see that the disk and directory settings are correct, then click Save. To save the file Unemp.xls to your diskette in the A drive, you may type A:\unemp.xls in the File Name text box. Now you could close the workbook or exit Excel from File. To reopen a closed workbook, click the Open button (which looks like an opening manila folder) or use the File menu, and click on a file name to open it.

**II. Save your worksheet. Regular backups are important.**

**Autocalculate:** Values appear in status bar on bottom right of page. Right-click anywhere on the status bar to make the Autocalculate shortcut menu appear. It displays Average (mean of values in a selected range), Count (number of filled cells in a selected range, including text), Count Nums (counts number of cells containing numbers, but not blank cells or text), Max or Min (for highest or lowest values), or Sum (total of values). Note: these values are not inserted in the worksheet. This is a good way to check for errors before you print or calculate with your data.

**Fill Handle:** The lower right corner of an active cell (heavy border) is a little square called the fill handle. Aim at the fill handle, and the pointer turns into a black cross. You can drag the fill handle through a range to select cells. **Copy Across Columns:** Select the cell you wish to copy from, point at the fill handle and drag it across the other columns to copy the formula. The range automatically adjusts to reflect the cells in each column.

**III. Click C2 and drag through C25 to select the range. Find the Fill Handle and drag it across to D25. It repeats the column C values in Column D.**

**IV. Select the range A2 through A25. Click on A35, then click on the Autosum (summation sign) on the toolbar. Cells A2 to A25 are automatically totaled, and the result put in A35.**

**V. Point at the fill handle in A35 and drag through D35. Press Enter when the range is correct.**

**Formulas:** Click the Edit Formula button on the formula bar (=). This automatically inserts the equal sign, which must always begin a formula.

Arithmetic operators: + Addition, - Subtraction, \* Multiplication, / Division, ^ Exponents ( $=5^2$  gives 25), % Percentage ( $=%$  for .5 gives 50%). Note: without parentheses to indicate what order for operations, multiplication and division are done first, so  $=10+15*3$  would be 55, but  $=(10+15)*3$  would be 75. Numbers can be used in formulas, as well as single cells or a range of cells.

Comparison operators: =, >, <, >= (greater than or equal to), <= (less than or equal to), <> (not equal to)

Reference operators: Range operator is colon (:), so A1:C30 is the range that includes all the cells we have used so far. For cells in the first 2 rows use 1:2. For all cells in columns A and B use A:B. Union operator is comma (,) so A26,B2 means cell A26 and cell B2. A26:A30,B2:B25 means the range A26 to A30 plus the range B2 to B25.

Mouse: You can also select ranges with the mouse as above. **To select non-adjacent ranges with the mouse**, drag through the first part of the range, move the mouse pointer to the beginning of the second part, hold down the Ctrl key and click the first cell in the range, then Ctrl+drag through to the last cell.

**VI. Create a new variable, the change in the unemployment rate, in column E. Write the label in the first row. Then select E3 for the first observation. Use a formula to make this value the difference between B3 and B2 ( $=B3-B2$ ). Then use the fill handle to copy the formula down the rest of the observations. Do the same for changes in the inflation rate.**

**Paste Function button:** If you click on the Paste Function button ( $f_x$ ) you get a list and descriptions of functions, help in editing functions, and prompts for the right type and number of arguments in each step. Select the range, then click the Paste Function button. Find and click on the function you want to get prompts.

**VII. Calculate the correlation between inflation rates and unemployment rates. Put it in the B36 cell. Click Paste Function on the Standard toolbar or the formula toolbar. Click Statistics, Correl. The first range is B3:B22, and the second range is C3:C22. (We are dropping the first data point because the changes data start in the 3<sup>rd</sup> row.) Separate the arrays by a comma. Do the same for the changes in inflation and changes in unemployment (put it in D36). Which has a higher correlation? Why? Save your worksheet. It is a good idea to save it more than one place, such as on your A drive on a diskette and on your H drive, your personal space on the network, in case the disk is damaged or lost.**

**VIII. In Row 37, put the means (AVERAGE) for columns B through E and the standard deviations (STDEV) underneath in Row 38.**

**Trend line:** draws a best-fit straight line using OLS, in other words, a regression. The dependent variable is your data, and the independent variable is 1,2,3 . . ., in other words, a straight line trend. First select the range enclosing the data series. At the bottom of the selected range, grab the fill handle with the right mouse button and drag to the end of the range enclosing the blank cells you want filled in. Release the mouse button to get the shortcut menu and click on Linear Trend, and Excel fills in the blank cells with a trend forecast. But be careful! If you clicked Edit, Fill, Series to get the Series dialog box, selecting Linear, or selecting the Trend check box will replace your data with the trend line from the data. You could lose your original data. (If you want to try this, first save your series in another place on your worksheet.)

You will use this later:

**RAND:** This function produces a new random number between 0 and 1 every time you recalculate the worksheet or press F9. To get numbers greater than 1, multiply the RAND function, such as by 100 to get numbers from 0 to 100.

**IX. Fill in a trend projection for 1996, 1997, 1998, 1999, and 2000 unemployment and inflation rates.**

**Charts:** First, select the range of the data you want to chart (nonadjacent ranges are ok). Include row and column titles for the chart legend and labels. Then click the Chart Wizard (bar chart) button on the Standard toolbar. We will usually have too many data points for many kinds of charts. We will use mostly the scatter (or xy chart). The vertical arm of the chart, the y axis, is the dependent variable on a chart. The horizontal arm, the x axis, is time or the independent variable. The chart may be embedded in the worksheet right next to your data, or on a separate worksheet. Select the type of scatter chart, then click Press and Hold to View Sample button to see your chart. If this is what you want, click Next. Here is where you can add a legend, label axes, and title the chart if this hasn't been done already by Excel. Click Next again to choose whether the chart will be in its own worksheet (labeled Chart 1, better if you want to turn the chart into a slide or overhead) or embed it in the current worksheet (better if you want to present data and chart together or if you want to see the results of changes in data). Click on Finish to

end. To change the chart afterwards, right-click anywhere inside the chart area to pop up a shortcut menu for editing the chart.

- X. CHART YOUR DATA! Chart a line chart (a type of scatter chart -- DON'T USE LINE CHART!) with the unemployment rate and the inflation rate as dependent (y axis) variables against time (on the x axis). You may connect the dots to make the data more legible, but use the scatterplot plot type. Do unemployment rates and inflation rates appear to move together over time?**

**Second Chart: Make a scattergram of the unemployment rate versus the inflation rate. Excel requires that the inflation rate data be in a column to the *left* of the unemployment rate data in order to put the inflation rate on the x axis. (Inflation is supposed to cause unemployment, so it belongs on the x axis. Unemployment rates are on the y axis.) Do they look correlated?**

**Print:** Click File, Print to check the default printer. Then click the Print Preview button (magnifying glass and sheet of paper). This gives a full-page view of the worksheet, and buttons to change margins, get rid of grid lines if you want, etc. Click the Setup button to pop up the Page Setup dialog box, then alter margins, click Center on Page Horizontally and Vertically check boxes, deselect by clicking the Print Gridlines if you don't want gridlines. Click OK in the Page Setup dialog box, then click the Print button on the toolbar. If you want 1 copy, click OK.

If you just click on the Print button, Excel will print whatever is in the current window all the way from the first cell with something in it to the last cell with something in it. If the data are too much for 1 page, you can print only part of the worksheet. Select the part you want to print, then click File, Print, to open the Print dialog box. Click Selection under Print What, and then click OK. You can click the Print Preview button to see how the printout will look.

## Session II: Regression Analysis in Excel

There are 2 ways to run regressions in Excel. In forecasting inflation and unemployment rates in 1996 and 1997, we use the TREND function. Using this same function, we can put a line of regression in a scatter (XY) diagram as well. The other method is through the Data Analysis Tool, Regression option.

### TREND Function

**To superimpose the regression line on the scatter diagram** right click on any single point. (Many of the points will be highlighted.) Select the command Insert|Trendline. When the Trendline dialog box appears, select the Type Tab, and then select Linear as the Trend/Regression Type. Now click the Options Tab. In the Options Tab, select the Automatic Trendline Name option button and select the Display Equation on Chart and the Display R-Squared Value on Chart check boxes. Click the OK button, and Excel will add a regression line, the regression equation, and the R-squared value to the chart. You can reposition the equation and r-squared value by dragging them.

- I. Create a scatter diagram (X axis: inflation, Y axis: unemployment) using data from 1975 to 1995. Add the regression line, equation, and R-squared value. Create a scatter diagram for the changes in inflation and unemployment from 1976 to 1995, and add the regression line, equation and R-squared value. Print your result. How do your results compare with the correlations you performed between inflation and unemployment, and between the changes in inflation and the changes in unemployment?**

The TREND function is a useful way to see how a regression changes as the data change. For example, the inflation and unemployment data series start in 1975, a recession year. Starting or ending a data series at extreme points in the business cycle (troughs or peaks) can tilt the regression, causing the slope (coefficient multiplying X) to be larger or smaller.

- II. Remove the equation text and R-squared value by selecting them and commanding Edit | Clear | All. Then remove the 1975 observations on the Data sheet. Find the new values for the Y intercept and regression equation by repeating the Insert | Trendline Command. What changes occurred?**

### Data Analysis Tool for Regression

Select Tools | Data Analysis, then select Regression in the Analysis Tools list box, and click the OK button. If Data Analysis does not appear at the bottom of the selections under Tools, select Add Ins, then click (check mark) for both Data Analysis Toolpak and Data Analysis Toolpak VBA (the first boxes), and click OK. Now Tools should have the Data Analysis | Regression.

In the Regression dialog box, enter the input range for the Y variable. Enter the input range for the X variable.

Select the Labels check box, **BUT ONLY IF YOUR SERIES LABEL IS INCLUDED IN THE DATA RANGE YOU SELECTED!** Otherwise, Excel will use the first data point as a label, not in the regression. If you don't select Labels, then Excel will call your variables X Variable 1, X Variable 2, and so on. You can edit the cells and chart labels afterwards to insert the correct series names.

Select the Confidence Level check box and set the level to 90%.

Select the New Worksheet Ply option button and enter Regression as the name.

Select the Residuals, Standardized Residuals, and Residual Plots check boxes. Select the Line Fit Plots check box if you wish to see a scatter diagram with a fitted regression line as above.

Click the OK button, and the text part of the output (4 sections) will appear in the left side of the sheet, with the Residuals plot on the right side.

- III. Using the Data Analysis Tools, regress changes in unemployment (y variable) on changes in inflation (x variable) from 1976 to 1995. How does this regression compare to the changes regression above?**

### **Multiple Regression**

For multiple regression, the X variables must be together in a single continuous range, so move or copy the series to place all X variables together. For example, copy the Inflation Rate to column H (to the right of the Changes in Inflation). Then your X variables data are G3:H22. Select Tools | Data Analysis, select Regression from the Analysis Tools list box, and click OK. In the Regression dialog box:

Enter the Y range and X range in the appropriate edit boxes, select the Labels check box, the New Worksheet Ply option button, and enter Multiple Regression as the sheet name, and select the Residuals, Standardized Residuals, and Residual Plots check boxes. Click the OK button to obtain the output.

- IV. Run a regression with the Unemployment rate from 1976 to 1995 as the Y variable and the inflation rate, and changes in the inflation rate as the 2 X variables. Is this a better regression than the simple regression above?**

## Extra Credit: Random Numbers

To create a series with random numbers use the **RAND** function. This function produces a new random number between 0 and 1 every time you recalculate the worksheet or press F9. To get numbers greater than 1, multiply the RAND function. For example, =100\*Rand to get numbers from 0 to 100.

- I. Create a series containing 2500 random numbers between 0 and 10 using the Rand function, in A1 through A2500. Put the "class" variable, the frequency bin upper bounds (1,2,3,4...), in B1 through B10.

### **Frequency Distribution** (First with Functions | Frequency, then with Tools | Data Analysis | Histogram)

Create a series (column) whose cells contain the values for the upper boundaries for each class you want to count the frequencies in. Call it "Class". For example, if you have random numbers from 0 to 10, and you want them separated into 10 classes, then the upper boundaries are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

The next column is the cumulative frequency. Use the function **=Frequency(cell range, upper boundaries for each class)**, where the cell range is the data to be analyzed and the upper boundaries is the range of cells that contain the values for the upper boundaries. You can use the paste function or type the function command in cell C1. (**=frequency(\$a\$1:\$a\$2500,b1)**) and use the fill handle to copy it down through C10. The number in C10 should be 2500, your total sample size.

The next column is the relative frequency. Create the relative frequency for each frequency subtracting each cumulative frequency from the previous one (except for the first, in C1) and dividing by the sum. Since this number's location would not change if we moved the relative frequency column to another location on the worksheet, we will use the **absolute address** of the cell. To do this, put a dollar sign before the letter and again before the number (**\$C\$10**). So the command for the relative frequency for cell D1 is **=C1/\$C\$10**. For the second, it is **=(C2-C1)/\$C\$10**, which can be copied down the column.

The next column is percent. To present the first cell E1, in percent form, enter **=D1** and so on and make them percent form by selecting the range E10:E10 and press the Percent Format button. You can adjust the decimals displayed by the Increase Decimal button. Then click OK.

You can now use Chart Wizard to make a chart of the relative frequencies. Or . . .

### **Histogram**

**Use Tools | Data Analysis | Histogram. In the Input Area, put the data (A1:A2500 in the above example) in the Input Range box. Put the range with the upper class boundaries (B1:B10 in the above example) in the Bin Range edit box. In the Output Area, you may select (click) New Worksheet Ply button (name the new sheet Histogram) then select Cumulative % and Chart Output boxes, then click OK.**

The chart may be improved by **removing gaps between bars**. To edit the chart, click on the white area in the chart. Right click on Column. Select Format Data Series. In the dialog box, Options tab, set Gap Width=0, and click OK.

There is also an extra class displayed, and an extra bar and cumulative point. To **remove the extra bar**, right click on the white area of the chart. Select the Source Data dialog box. Click the Series tab. In the Values edit box, change the ending cell for each of the 2 series. (In the example above, change the ending cell from \$B\$11 to \$B\$10 and change \$C\$11 to \$C\$10) Click the OK button. The chart should now have the correct number, 10 classes.

**To enlarge a chart:** Select a cell outside the chart, then single click in the white area inside the chart. A set of 8 square “handles” appears on the border of the chart. Move the mouse pointer directly over the lower left-hand corner of the chart box. The mouse pointer changes to a small double-sided arrow. Drag the mouse pointer, which will change to a plus sign. As you drag, the chart border expands. Release the mouse button when the chart is big enough.

**II. Create a histogram of the random series above. What distribution do you think Excel uses to generate the random series?**