

Kenyon College
ECON 375 Introduction to Econometrics
Spring 2011, Keeler

Introduction to Excel

January 27, 2011

1. Open Excel

From the START button, go to PROGRAMS and then MICROSOFT OFFICE, and then EXCEL.

2. Import the data set

A) Data from an Excel File

When Excel is opened, there will be a blank spreadsheet grid. Click the Office Button in the upper left, then OPEN. In the Dialog Box, click on the drive (Look In:) until you navigate to the file. Click on the data file name and then click Open (or double-click on the file name).

The data set is from Stock and Watson, Ch. 6 p. 211, is available as an Excel file at:

P:\Class\ECON\Keeler\ECON 375\Growth.xlsx

The data set has information on 65 countries for variables on economic growth. The Data Description is attached.

If you open another type of file, not an Excel file with a .xls or .xlsx extension, it will open the **Import Wizard**, which will ask you to make some choices about how to import the data. Files may be “delimited” with some marker between columns of data series, or they may be “fixed width” where a certain column width is specified for each data series. Usually data files are “delimited”, so select that option. The data are previewed in the window. Click NEXT. Step 2 is to tell Excel what delimiters are used - what separates the data. TAB is the default and you can leave that. Some files from government web sites will use commas. Many files use “spaces” as delimiters, so check that box. When you do, vertical lines will be inserted between the columns of data series. If that looks right for separating the variables, then click NEXT. Step 3 asks you about the format of the data. “General” is the default, and you can leave that and change it later to better options. It will also let you select columns to skip and not import. Click FINISH. The data should appear in the spreadsheet.

B) Importing data from a web site can be done by copy and paste, but more are offering data in Excel format. For example, open an internet browser and go to the Federal Reserve Bank of St. Louis data source, FREDII at:

<http://research.stlouisfed.org/fred2/>

Locate a data series that you are interested in such as the unemployment rate, an interest rate, real GDP growth rate or an exchange rate for example. Click on the series, and it will show a chart. Click on “Download”, and it will show a screen that has options including an Excel format. Select that, click Download Data, and you have the option to open the file in Excel or to save the file to a folder to open later. If you open it with the default Excel, the data should appear in the spreadsheet.

If the date and the data series, or several data series are ever all in one column – leave it highlighted that way. Now click the **DATA** tab in the second row at the top in Excel, and Text-To-Columns and that opens the Import Wizard that will let you put in the delimiter and move the data series to a separate column. The next time you do this, Excel will automatically put the data and the data series in different columns.

3. Variable Names

If the Excel file does not already have names for the variables at the top of each column, you can type in your own names for the variables. You can insert a row at the top of the data set (click on Row 1 on the left border, then click Insert from the Menu Bar, then click Row) to identify the data set more clearly. The names should be brief but descriptive, and one word. You can use letters or numbers or both, but don't use special characters. If for some reason you want to change the order of the variables in the spreadsheet, you can do that. Select an area by highlighting it, or select the whole column by clicking on the letter at the top of the column, and then grab the column at any place on the border and drag it to the new area

4. Construct new variables

To create a new variable, select a cell at the top of a new column. Every formula begins with the equals sign, "=", and then Excel will then expect a formula to be entered which it will use to calculate the new variable. For example,

$$= b2 + c2 \qquad \qquad \qquad = (b2/c2)*100 \qquad \qquad =b2^2$$

Create a new variable for the square of initial income, rgdp60. Click the cell in a new column, in the top row. Type a name for the new variable. Determine a formula that creates the squared value of the salary. Type the equals sign, =, and then the formula using cell names in the same row, as with the examples just given. When the formula reads as you want it to, click the green check mark in the row just above the spreadsheet. That will apply the formula and put the value into the active cell. The formula appears in the box whenever you select a cell that uses a formula, so you can check how the value was calculated.

Use **AUTOFILL** to complete the data for the new variable. You now have a formula in one cell at the top of the column. Click on that cell to make it the active cell if it isn't already. In the lower right corner of the active cell, there is a square. Place the mouse over the square, press down on the mouse button, and hold it down as you move the mouse down the column to the last row of the table. Then release the mouse button and that will apply the formula to the entire column. Alternatively, put the cursor over the lower right corner and double click – that will also autofill the column.

When a cell address is written as "b2" for example, Excel will update the column and row. As you drag the mouse down to row 3, the formula changes for that cell and updates it to "b3". If you click on another cell in the column, you will see how the formula has been "updated" by Excel. So you should only have to type a formula once and then you can either drag it or copy and paste it.

Another use of **AUTOFILL** is to continue a pattern in the data. Select a new column and type in a pattern of numbers in the first three or four rows. Select the rows of that column and grab the handle at the lower right and drag it down a few more rows. That should continue to enter

numbers that extend the pattern - you only have to select enough cells for Excel to identify the pattern and it will continue. You can delete this by clicking the blue arrow that curls around to the left in the toolbar - it is “**Undo**” - very handy.

5. Functions

Calculate the average value of the Growth variable in the data set. Select a cell below the data column for that variable. Click on the icon for “**Function Wizard**” which is just above the spreadsheet grid, and it is a script letter f followed by x, **fx**. Click on AVERAGE from the pull down menu, and click OK. That opens a dialog box. You can enter the data range in several ways. Since you located the active cell at the bottom of a column, the range will be written already as all the cells above the active cell (until it hits a blank cell or letters). In the box for “Number 1” is written the address for the range of cells, such as B2:B66. An address for a range, in a column, row or rectangular area, is always the name of the cell in the upper left, followed by a colon “;”, and then the name of the cell in the lower right. If you want to change it, click the red white and blue spreadsheet icon at the right of the box and that returns you to the spreadsheet. There you can use the mouse to select the range you want, and then click the spreadsheet icon again. Click OK. That should take the mean of the column that you selected and write the value in the cell. When you click on the cell, the function you used will show up in the Formula Bar at the top of the page.

6. Data Analysis

Click on the **DATA** tab at the very top of the page, and the last item on the right should be **DATA ANALYSIS**. If it is not, click the Office Button, and then at the bottom of the list, click the box for Excel Options. At the left, click “Add-Ins” and select Data Analysis to add to Excel. When Excel comes back, it will have DATA ANALYSIS at the DATA tab. The menu provides a list of statistical procedures that you can do in Excel.

Click **DESCRIPTIVE STATISTICS**. In the Dialog Box, you need to tell Excel the data you want to analyze. Click the red white and blue spreadsheet icon at the right of the box and that returns you to the DATA spreadsheet where you can select a range. Use the mouse to select data for one or more variables, and include the variable name at the top of the column. Click the box for “New Worksheet Ply” for Excel to write the results to a new page. Click the box for “Summary Statistics”, and then click OK. When the results appear, you will need to widen the columns to read all of the column titles. Move the mouse to the vertical bar between two column letters at the top, and double click – it will automatically widen enough to show the widest entry. The value for mean should be the same as you calculated earlier. Do this for one of the main variables, such as Growth.

A **HISTOGRAM** will make a table of the number of observations in several ranges of values, and show the distribution of the variable in a graph. Before you start with this one, you have to tell Excel how many bars you want in the graph, which are ranges and the midpoint of each range. Try specifying about 10 “bins” that cover the range of data for one of the variables that you chose for the Descriptive Statistics. In a new column of the spreadsheet, type in numbers that represent midpoints for the bins, covering the range of values for the Growth variable for example. Put in the name Bins at the top of the column. From the Table produced by Descriptive Statistics, you know the range of values of the variable such as Growth, from minimum to maximum - split that into about 10 even

groups so that you cover the full range of values with 10 numbers in even steps. Type the values of the bins into a column of numbers, which you can locate anywhere that's convenient. Click on DATA, DATA ANALYSIS and HISTOGRAM. In the dialog box, enter the input range for the values of one of the variables, or click the spreadsheet icon and select the variable in the DATA spreadsheet. Click on the icon again when you have selected the data. Back in the HISTOGRAM dialog box, at Bin Range, enter the cell range for the bins that you just created, or again click on the spreadsheet icon and specify it that way. Check the box for Labels if you included the variable name at the top of the column. Click on "New Worksheet Ply", then Chart Output, and then OK. The results should show 10 rows that list how many observations are in each bin and the percent of the total number of observations that represents (frequency), as well as the cumulative percent in bins from the smallest value through that bin. It will give you a vertical bar chart with a bar for each bin. If the result is not very informative, you may want to create either fewer or more bins.

Variance and Covariance - One way to work with a variable or a relation between variables is using a building block such as "deviation from the mean". Let's start with one variable and calculate the variance. We know the mean of Growth. Create a new variable that is the value of the variable minus the mean of the variable. Type in that formula and drag the cell to apply it to all the rows. What do the values in that column represent?

Use a build-in formula to take the sum of the column (on the HOME tab, the sigma Σ symbol at the right has the formula for sum. What do you get as the result? Why?

One way of preventing the values from canceling is to square the deviation from the mean. Change the formula to square the deviation and apply that to the column. Now take the sum again - actually the sum function should update automatically. What is the result? How is that value affected by the number of observations we have? Let's take an "average squared deviation from the mean" by dividing by $(n-1) = (66 - 1) = 65$. Compare that to the Covariance table results.

Skewness and Kurtosis – Once you have calculated the deviation from the mean as a variable, you can construct skewness and variance. For Skewness, raise the deviation to the power 3. Sum that column and then divide by $(n - 1) = 65$. Divide that mean of the sum of the cube by the cube of the standard deviation. Compare that to the skewness measure provided by Excel.

7. Create a chart of the data

Construct an XY Scatter graph of the growth rate on the vertical and initial income or rgdp60 on the horizontal axis.

Click on the **INSERT** tab. Click to choose a chart type. It is easy to try out different chart types and change it after you have made the chart, so you can always come back and do the steps differently. Let's use the Scatter chart. Click on that chart type, and then click on one of the versions of that chart.

The **Select Data** dialog box opens and lets you select the variables to be graphed. You can add or remove variables, and you can change the data range. At any later time you can come back to this easily. Click ADD and then the Edit Series dialog box opens. Click the spreadsheet icon at the right end of the Name box, and that puts you back in the data in DATA. Click the variable name and then the spreadsheet icon again. Click the icon at the right of the Series X values box and select those, and the same for the Series Y values. Click OK, and that should show the chart.

You can enter labels for the chart. Make sure the chart is selected, so it has a light blue border (if not click somewhere in the chart). The tab under **Chart Tools** should be for **Layout**. Click Chart Title and you can put in a Title for the Chart with some options. Click Axis Titles and put in names for the X and Y axes in the graph. Excel will update the chart to show you what it looks like and you can change it then if you want. Try the other tabs to see what options you have there to control the appearance of the chart.

You have two choices about where you want to save the chart. Currently it is embedded in the spreadsheet page and you will be able to resize it or move it. You can also move the chart to its own page. On the **Chart Tools Design** tab, click the icon in the upper right for Move Chart Location. If you select “new sheet” it will store the chart by itself on a new worksheet, and create a new tab for it. Choose one of these and then Click Finish.

After the chart is made, you can always edit it. Go to the chart and click anywhere in the middle of it. That should put a light blue border around it. It also changes the TOOLBAR at the top and adds CHART TOOLS to the menu items. Click on Design Layout or Format, and it will show you options to edit the chart type, data sources and more.

8. Draw a line in the graph

At **CHART TOOLS** and **LAYOUT**, there is an icon for Shapes. Click on the straight line that angles down to the right. Then click in the chart at one end of the line you want to draw, say a vertical line for the mean values for rgdp60. Drag the mouse to vertically locate the line, and then release the mouse. Click somewhere else in the chart to complete the line. You can always click UNDO or click on the line and then Delete if you want to try again. Draw a horizontal line also for the mean of Growth. That splits the scatter graph into 4 quadrants. You can see how each data point can be expressed as the deviation from the mean of x and the mean of y.

ADD TRENDLINE to a chart. Use this only with an XY Scatter chart (NOT a Line Chart). The variable *Growth* is on the vertical (Y) axis and rgdp60, is on the horizontal (X) axis. Click anywhere in the middle of the Chart to select the chart. Then at the **CHART TOOLS** and **LAYOUT** tab, choose TRENDLINE, and click Options. Try the linear trendline which is the default, but other forms are available. Check the boxes for Show Equation and R-Squared in the Chart. Excel is performing a regression estimate and adding that to the chart.

9. Ordinary Least Squares OLS

Does the trendline, which is a regression line, minimize the sum of squared errors? Use your estimated intercept and slope values from the line that you drew to calculate residuals. In a new column, type in the formula:

$$= B0 + (B1*C2)$$

Where B0 is the numerical value for your constant or intercept, and B1 is the numerical value for your estimated slope. C2 refers to the cell address of the first observation in the rgdp60 variable. Click the green check mark, and then autofill. In the next column, calculate the residuals. Use Autofill. In the next column, square the residual, and at the bottom of that column, take the sum. How does that compare to the sum of squared residuals for the regression?

10. Regression in Excel

Excel will calculate the estimates of the intercept and slope parameters, along with other statistics. Select the DATA worksheet at the bottom of the screen, which has the spreadsheet of the data. Click on the **DATA** tab and then **DATA ANALYSIS**, and then click Regression and OK. Fill in the boxes. When you select the variables, include the variable names in the range and also check the box for Labels. Check the boxes for residuals and plots. This produces a lot of tables and plots, so leave the location as a New Worksheet Ply, and click OK.

In the section for ANOVA (analysis of Variance) the column at the left lists the variable name, and then the next column is for coefficient. Those are the estimates of the B0 intercept and the B1 slope parameters and they should be the same values (more precise) as the values in the trendline equation. Above that in the column labeled SS are Sums of Squares. Total Sum of Squares = SST is in the row for Total, the Explained Sum of Squares SSE is in the row for Regression, and the Residual Sum of Squares SSR is in the row for Residual, and that is the 5980.682 in this example.

11. Save the Data Set

Click on FILE and SAVE AS. Type in a new file name.

DOCUMENTATION FOR GROWTH DATA

Growth contains data on average growth rates over 1960-1995 for 65 countries, along with variables that are potentially related to growth. These data were provided by Professor Ross Levine of Brown University and were used in his paper with Thorsten Beck and Norman Loayza: "Finance and the Sources of Growth" *Journal of Financial Economics*, 2000, Vol. 58, pp. 261-300.

Variable	Definition
<i>Country_name</i>	Name of country
<i>growth</i>	Average annual percentage growth of real Gross Domestic Product (GDP)* from 1960 to 1995.
<i>rgdp60</i>	The value of GDP* per capita in 1960, converted to 1960 US dollars
<i>tradeshare</i>	The average share of trade in the economy from 1960 to 1995, measured as the sum of exports plus imports, divided by GDP; that is, the average value of $(X + M)/GDP$ from 1960 to 1995, where X = exports and M = imports (both X and M are positive).
<i>years school</i>	Average number of years of schooling of adult residents in that country in 1960
<i>rev_coups</i>	Average annual number of revolutions, insurrections (successful or not) and coup d'etats in that country from 1960 to 1995
<i>assassinations</i>	Average annual number of political assassinations in that country from 1960 to 1995 (per million population)
<i>oil</i>	= 1 if oil accounted for at least half of exports in 1960 = 0 otherwise