Partial Solutions to the Some of the Exercises in *Analysis of Economic Data*, by Gary Koop

General Comments

Most of the questions in the book require you to use Excel (or an econometrics software package, although this set of notes is written for the Excel user). Although I am reasonably explicit about the steps involved in solving the questions, it is very useful if you already knows the basics of Excel. There are an enormous number of books about Excel which you might want to take a look at *Computing Skills for Economists* by Guy Judge (John Wiley & Sons, 2000) discusses Excel (and many other computer-related issues) in the context of what economists do, so this book may be a good place to start if you are unfamiliar with Excel. The solutions I provide in these notes mainly focus on the mechanics of how to use Excel; there is relatively little about interpretation of results. Interpretation of results is, of course, of great importance and I would encourage you to think carefully about the economic implications of your results. The book itself discusses in great detail how to interpret results.

Here I offer a list of thoughts about general issues I have noted students have had trouble with.

* Most of the questions require selecting the "Tools/Data Analysis" option. However, depending on how Excel was installed on your computer, this may not be there. If it is not you should be able to get "Data Analysis" by clicking on "Tools/Add Ins" and then adding the "Analysis Toolpack". If this is not an available option, you will have to get the CD-ROM for Excel and re-install it (using the "Complete Installation" option).
* In Excel, there is often more than one way of doing something. For instance, to specify a column of data you can either type out its location (e.g. A1:A15) or highlight it (e.g. start at A1 and hold down your left mouse button and then drag the mouse down to cell A15). In these notes, I will describe one way of doing things, but others are, of course, perfectly acceptable.
* If you have made a mistake or want to start over, note that deleting a chart or a worksheet in Excel can be done by going to the offending sheet/chart (or the place where its title is at the bottom of the screen) and clicking on the right mouse button. This brings up a list of options, one of which will be "Delete" or "Clear". Clicking on one of these allows you to remove your mistakes and start over.
* In these notes I often use quotation marks (" ") to indicate what you should type. For instance, I might say: type "= A1+A2". You should NOT actually type the quotation marks, only the bits inside.
* You may wish to make back-up copies of all the Excel data files provided with the book. The exercises will involve you in altering the Excel spreadsheets, so you may want to keep the original spreadsheets just in case you make a mistake. However, all the data used in the book is available on the accompanying web site so you can also get new copies of the data files from the web should need arise.

Chapter 2: Basic Data Handling

General comments: This chapter discusses basic graphical methods and descriptive statistics. It is quite brief and only covers a few of the most important tools and techniques. With regards to the graphical methods, in particular, I would encourage you to experiment with Excel's ChartWizard which has many types of graphs, including some not mentioned in the book. I also note that ChartWizard allows you to make your graphs look nice (e.g. make titles, label axes, make legends, etc.). I do not discuss these things here, but I would encourage you to experiment with these options.

**Exercise 2.1**

a) Recreate Figure 2.1.

Solution:

i) Open the data file EXRUK.XLS (use File/Open and then find the file in whatever folder it has been placed).

ii) Click on the ChartWizard icon (it is in the upper right area of the screen and looks like a little bar graph).

iii) Choose chart type to be "Line" then click on "Next".

iv) In the box labelled "Data Range" enter B1:B598. Then, where it says "Series In", click on "Columns". Then click "Next".

v) The menu which appears at this stage allows you to make titles, legends, etc. I will ignore it, although you may wish to experiment with it. Click "Next".

vi) The menu which appears at this stage asks where you want to put the graph you are making and lets you put it on the same sheet as the data or on a new sheet. Choose whichever you want and click "Finish". A graph similar to Figure 2.1 in the book should appear.

b) Make one time series plot containing two time series, the logs of consumption and income.

Solution:

i) Open the data file INCOME.XLS.

ii) Highlight all of the data (i.e. move mouse to A1 and then move to C165 while holding down the left mouse button).

iii) Click on ChartWizard, choose chart type of "Line" and then just keep on clicking "Next" until you get to the last menu and click "Finish".

Note that this should produce a graph with both consumption and income on it. By highlighting all of the data Excel automatically labels each of the series (using the name in finds in the top row) and labels the X-axis with the dates it finds in the first column.

c) Transform the logged consumption and income series to growth rates and then plot them.

Solution:

i) Move to D3 (i.e. move the mouse to this cell and then click, cell should become surrounded by a bold border). Then move to the formula bar (the blank row just above the spreadsheet that follows the "=" sign) and type "=100\*(B3-B2)" and press enter. The number -.55684 should appear. This is the percentage change in income between 1954Q1 and 1954Q2 (i.e. income fell by a little more than half a percent in this quarter).

ii) Now move the mouse *precisely* to the bottom right hand corner of cell D3. Hold down the left mouse button and drag all the way down to D165 before releasing the left mouse button. A whole list of numbers, which are the percentage change in income during every quarter should appear.

iii) Graphing this series is done in the same manner as Exercise 2.1 a).

**Exercise 2.2**

a) Recreate the histogram using data set GDPPC.XLS which contains GDP per capita for 90 countries.

Solution:

For the remainder of these notes I will not explicitly say "Open the relevant data file" as I did above. I will assume you have done this. Note, in this case, that the GDP per capita data is in cells D1:D90.

i) Create the bins you wish somewhere on the spreadsheet. For instance, type 2,000 in cell E1, 4,000 in E2, 6,000 in E3, 8,000 in E4, 10,000 in E5, 12,000 in E6, 14,000 in E7 and 16,000 in E8.

ii) Click on Tools/Data Analysis and then select Histogram and click OK. In the box labelled "Data Range" type D1:D90. In "Bin Range" type E1:E8. Click on the box which says "Chart Output". Click "OK".

Note: Excel seems to squash the resulting histogram and you may have to re-size it to get it looking nice (click the mouse in the body of the graph, little black boxes should appear on the border of the graph. If you then hold down the left mouse button on these boxes and drag the mouse you can re-size however you like.)

b) Experiment with different bins.

Solution:

The steps in a) will create a frequency table and histogram using the bins you specified in cells E1:E8. By simply typing up another set of bins (e.g. in column F) and then telling Excel to use these in the "Bin Range" box you can make a histogram with any bins you want. If you leave the "Bin Range" box blank, Excel will make a default choice for you.

c) What does the "Cumulative Percentage" box in the Histogram Menu do?

Solution:

The frequency table tells you the number of countries which have GDP per capita in a certain interval. The "Cumulative Percentage" box lets you calculate cumulative percentages. For instance, if you use the bin choices of part a) you will find the bin labelled 6,000 has a frequency of 7 and a cumulative percentage of 68.89%. These two figures mean that there are 7 countries with GDP per capita *between* $4,001 and $6,000 and that 68.89% of the countries have a GDP per capita of $6,000 *or less*.

**Exercise 2.3**

Create *XY*-plots using the data in FOREST.XLS and see if you can find any relationship between the variables.

Solution:

I will just write an answer for the deforestation series (in A2:A71) and the change in cropland series (C2:C71).

i) Highlight all of the data then click on Chart Wizard and choose for chart type "XY(Scatter)" then click Next.

ii) The menu that comes up will contain a graph which looks a bit of a mess (because Excel is plotting all of the variables in the data set at the same time). Do not worry, just click the tab which says "Series" and remove the variables you do not want. That is, in the box labelled "Series" click on "Pop dens" then click on "Remove". Do the same with "Pasture ch". Click Next.

iii) The resulting menu will allow you to title or label your graph as you want. Click Next. Then Click Finish at the following menu and re-size the chart which results if necessary.

Although the pattern is not perfectly clear, there does seem to be a positive association between deforestation and the expansion of cropland. Countries with low values for "Crop ch" tend to have low values for deforestation. Countries with high values for "Crop ch" tend to have high values for deforestation. There are, however, several countries which do not fit this pattern.

**Exercise 2.4**

Obtain descriptive statistics on all variables in FOREST.XLS.

Solution:

i) Select Tools/Data Analysis then Descriptive Statistics and click OK.

ii) In the box labelled "Input Range" type in A2:D71 (or simply highlight all the data). Then click on the box labelled "Summary Statistics". Then click OK.

A table will result which contains all sorts of descriptive statistics (including some not covered in the book) for the four variables. Note that the title "Column 1" refers to the variable in the first column of the spreadsheet (deforestation), "Column 2" refers to the variable in the second column (population density), etc.