TEACHING STATISTICS USING DEMONSTRATIONS IMPLEMENTED WITH EXCEL

Rodney Carr
Deakin University
Australia

Most people realise that Excel can be used to carry out statistical analyses of data but what is not so well known is that it provides an extremely flexible and versatile teaching tool for constructing computer demonstrations. In this paper we first show a number of pre-built Excel workbooks for demonstrating statistical ideas - these illustrate the range of things that can be achieved. Then we show that building such workbooks is quite easy by showing how to construct a useful computer demonstration "from scratch" in just a few minutes, starting with a blank workbook. The aims of the paper are to outline benefits of using Excel for teaching statistics and to encourage other teachers to experiment with their own demonstrations and ideas.

INTRODUCTION

Most people will know that Microsoft Excel comes equipped with statistical functions and tools that enable the application to be used as a (simple) data analysis package. At the most basic level, it can be used by simply entering formulas and drawing charts, etc. Workbooks tailored for particular needs can be built this way. Or users can select from one of many add-in packages such as the Data Analysis Toolbox that comes with Excel itself, or XLStatistics (see http://www.deakin.edu.au/~rodneyc/XLStats.htm). But in this paper we do not wish to illustrate any particular package or to show Excel can be used as a data analysis package – instead the aim is to show how Excel itself can be used to construct powerful demonstrations of statistical ideas for teaching purposes. In the first section below we give descriptions of some such workbooks to provide an overview of what is possible. The last section in this paper explains how one particular demonstration workbook can be built.

DESCRIPTIONS OF SOME DEMONSTRATION EXCEL WORKBOOKS

In this section we describe some Excel workbooks that can be used to demonstrate statistical ideas. The workbooks we describe are designed to be used in a lecture-type situation by a teacher demonstrating particular concepts, or by students themselves in lessons designed to help them discover concepts. The workbooks are all highly interactive and are intended to used and viewed on a computer, so unfortunately they lose much of their impact when simply described and static images presented. However, all of the workbooks we describe are available from http://www.deakin.edu.au/~rodneyc/PS.htm, in the 'Practical Statistics' text ('Practical Statistics' aims to provide a complete introduction to statistics using computer (Excel) demonstrations to convey all the important statistical ideas).

The first demonstration workbook we describe is a simple one used to illustrate various shapes in single-variable data, along with properties of the corresponding numerical summaries. It is called ShapeX.xls. When opened, users see a number of sheets in ShapeX.xls, named according to the type of shape they illustrate. Figure 1 shows an image of the sheet named 'Skew', used to teach students what skewness is, to show relationships between the actual shape of data and the skewness index, and to illustrate properties of various numerical measures such as the mean and median in the presence of skewness.

On this sheet the scrollbar labelled 'Skewness' (on row 3) provides a simple means for the user to vary the skewness in data. The data is shown in Column A. A frequency chart of the data is provided and the skewness index is shown in cell H6. Students essentially always immediately start moving the scrollbar (it is almost impossible to stop them from doing so!) and as they do, the data in column A changes and the chart and skewness index is updated. By moving the scrollbar, they quickly see that skewness is a measure of asymmetry and that data such as that pictured in Figure 1 (with a tail going out to the right) has a positive skewness index. By clicking the boxes below the chart students can experiment with effects of skewness on the measures such as the mean or median, or get to see that the 2-standard deviation rule fails to hold

when data is seriously skewed, or "test" Chebyshev's inequality by using the spinner, etc. There are many lessons that can be build around this spreadsheet.
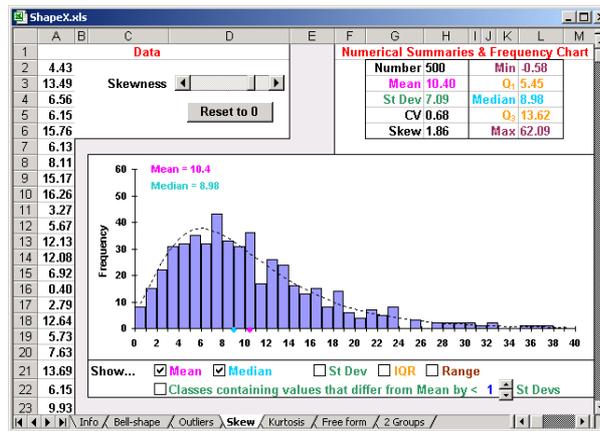

Figure 1. Image of the 'Skew' Sheet in ShapeX.xls

Other sheets in the ShapeX.xls workbook provide other lessons. The sheet labelled '2 Groups', for instance, is designed to show students how various shapes in data can arise if the data comes from two underlying different populations, see Figure 2. On this sheet, basic characteristics of the two groups can be set using the scrollbars in the 'Data' area. In the lesson pictured, for example, the two groups have quite different means, but there are many fewer observations from the group with the larger mean. The result is skewed data. By clicking the button for the groups to be shown in different colours, the underlying groups can be seen, as in Figure 3.
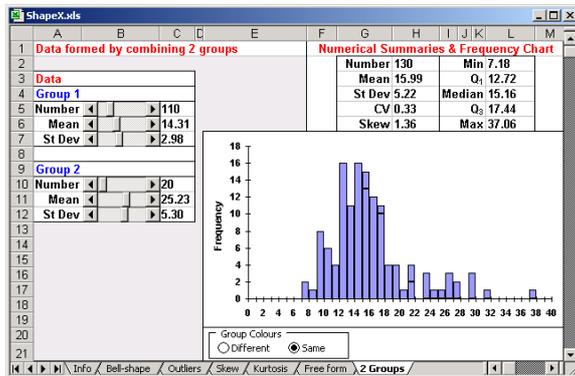

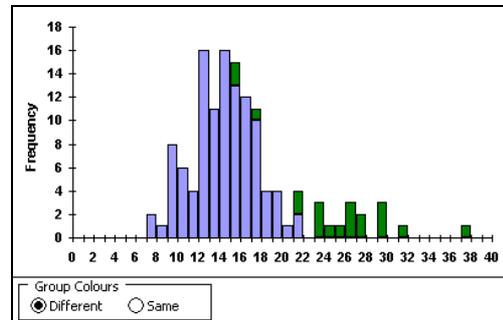*Figure 2*. Image of the '2 Groups' Sheet in ShapeX.xls

Figure 3. Frequency Chart on the '2 Groups' Sheet with the Groups Coloured Differently

The basic ideas used in ShapeX.xls can be used to construct demonstrations illustrating many other statistical ideas. Just to give one more example, in Figure 4 is shown the image of a sheet named 'Measure', from a workbook named Lies.xls. This sheet has scrollbars for showing the effect of invalid or unreliable measuring instruments – as the bar labelled Validity is moved away from the middle, the measured values become more and more "off-target"; as the bar labelled Reliability is moved to the right the measured values gain an unpredictable random error component.
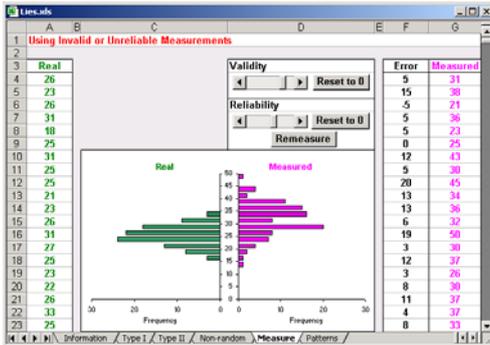
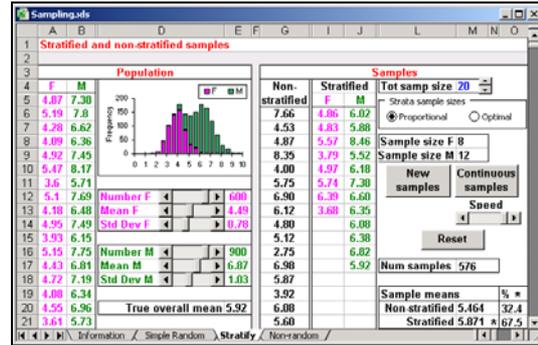*Figure 4*. Worksheet Illustrating Reliability and Validity

Figure 5. Illustrating Advantages of Stratified Random Sampling

The above examples are all interactive, the interactivity implemented using scrollbars, spinners, command buttons, etc, from the Forms or Control Toolbox toolbars in Excel. There is usually no programming required to implement such features. However, with all of these demonstrations, the user must vary things "by hand" to see the effects. This is appropriate in many situations, but others require the demonstration to be "animated" on the computer. For example, for many lessons it is useful for many samples to be drawn from a population, so that the distribution of sample statistics can be seen, for example. With a (usually small) amount of programming in visual basic it is possible to construct workbooks that are animated and show processes happening. For example, shown in Figure 5 is an image of a sheet named 'Stratify', contained in a workbook named Sampling.xls. It is used to illustrate properties of stratified random samples.

On the left of this sheet are controls for altering the structure of a population, similar to controls seen in workbooks in the previous section. The population data is shown in columns A (strata "F") and B (strata "M"). The population mean is shown in E20. On the right of the sheet a non-stratified sample is shown in column G and a stratified sample (of the same total size) is shown in columns I and J. The non-stratified mean is shown in M20 and the stratified mean in M21. The * in column N (which appears in either row 20 or 21) indicates which of the two means in M20 or M21 is closest to the population mean in E20 - for the particular sample shown, the stratified mean is closest to the population mean, so the * is next to the mean in M21. Of course, the stratified mean is not always closer to the true mean than the non-stratified mean, but you need to take many samples to illustrate how much "better" it is. The values in O20 and O21 show the percentages of the time that a particular mean "wins" as more samples are taken. To take many samples the user can repeatedly click the button labelled 'New samples' or the spreadsheet can be "animated" by clicking the button labelled 'Continuous samples'. Then samples are taken one after another, with the * and the % updating continuously. The image in Figure 5 was captured after 576 such samples had been taken and apparently with this particular population structure the stratified means seem to "win" about 67% of the time.

MAKING EXCEL DEMONSTRATION WORKBOOKS

The examples in the previous section hopefully give you some idea of what is possible in Excel. In this section we describe how one such workbook can be built, giving step-by-step instructions. The point here is not the actual workbook that we describe, and the actual steps are certainly not important – the point is to show you that, with a little training, it is very easy for people to construct their own demonstration workbooks. This workbook contains no programming in visual basic. The workbook we build is called Power.xls. It can be used to illustrate how the power of a simple one-sample z-test is determined. The final product is pictured in Figure 6.
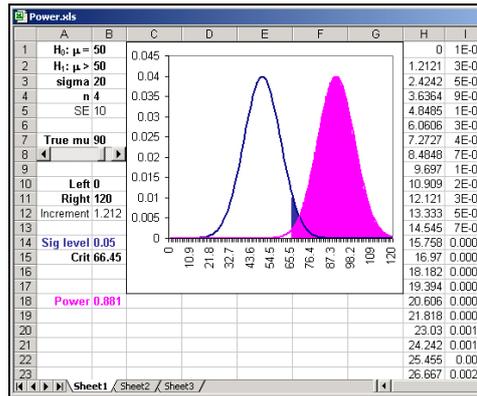
Figure 6. The Power.xls Worksheet

Steps to build Power.xls
1. Set up the statements for the hypothesis test in A1:B2.
2. Set a value for the population standard deviation in B3 and an initial value for the sample size in B4.
3. Compute the standard error of the mean in B5 by entering the formula =B3/SQRT(B4)
4. Draw the corresponding normal curve:
   a. Set values for left- and right-hand endpoints of the curve in B10 and B11.
   b. We'll drawn the curve using 100 points between the endpoints, so set the horizontal increment between points in B12 by entering the formula =(B11-B10)/99
   c. Values for horizontal axis: In H1 enter the formula =B10+(ROW()-1)*B12 and fill down to B100.
   d. The normal values for the vertical axis: In I1 enter the formula =NORMDIST(H1,B1,B5,0) and fill down to I100.
   e. Highlight H1:I100 and draw a line chart based on this data.
5. We will determine the power of the test to (correctly) reject the null hypothesis (H_0) if the true mean is something greater than 50, say 90 initially. Set this value in B7. We'll need to be able to vary this quantity quickly to see the effect, so place a scrollbar (from the Forms toolbar) linked to B7 just below the cell, on row 8.
6. Draw the normal curve corresponding to the true mean set in Step 5:
   a. We use the same horizontal values. For the vertical-axis values, enter the formula =NORMDIST(H1,B7,B5,0)
   in J1 and fill down to J100.
   b. Add these values as another series to the chart drawn in Step 4.
7. Let's set the significance level to 0.05 (in B14). The corresponding critical value (in B15) is then given by the formula =NORMINV(1-B14,B1,B5). We'll shade in the area corresponding to this on our chart. Also, we'll shade in the area corresponding to the power of the test.
   a. Enter the formula =IF(H1<B15,NA(),I1) into K1 and fill down to K100. Enter the formula =IF(H1<B15,NA(),J1) into L1 and fill down to L100.
   b. Add the data in columns K and L to the chart as two new series.
   c. To shade the areas, make the two new series into bar charts by altering the Chart type (on the Chart menu).
   d. Make the bars for the first series the same colour as the line chart drawn in Step 4 and the bars for the second series the same colour as the line drawn in Step 6.
8. Finally, determine the power of the test by entering the formula =1-NORMDIST(B15,B7,B3,1)
   into B18.
Now, as the scrollbar is moved (or other parameters altered) the spreadsheet (including the chart) will be updated to show how the corresponding power is determined.

Note: The Power.xls workbook, together with more complete details for its construction, can be obtained from http://www.deakin.edu.au/~rodneyc