USING INTERACTIVE VISUALISATION TO DEVELOP STATISTICAL UNDERSTANDING

Richard Castle
The University of Brighton
United Kingdom

This paper illustrates how Excel can be used by students to develop their statistical understanding. The student can vary data values by simply dragging data points on graphs and charts and seeing how this affects statistical estimates; thus, by visually exploring the effects of changing data values, students can get a feel for statistical concepts. Excel spreadsheets have been developed to explore both univariate, bivariate and inferential statistical topics. It is important when teaching statistics to non-statisticians that new statistical ideas are presented in a familiar and relevant context. The flexibility of Excel spreadsheets means that tutors can download relevant examples into the spreadsheet. The spreadsheets and some sample data sets are available on the World Wide Web.

INTRODUCTION

The teaching of statistics in UK post school education to non-statisticians generally involves a combination of traditional lectures together with tutorials and practical sessions using statistical software such as Minitab. The use of statistical packages has improved the quality of the teaching experience. It has enabled the students to investigate statistical ideas using realistic and relevant data without the recourse to laborious calculations on calculators. It has thus been possible to develop and emphasise the importance of statistical concepts in a familiar context.

Packages, such as Minitab, only allow an investigation of the impact of changing data values on a statistical measure (e.g. the mean or correlation coefficient) by changing values in the data worksheet. A more immediate and visual impact would be obtained by changing the data on the graph by means of the mouse, but, unfortunately, most statistical packages do not have this facility at present. However, Excel does have this capability.

Packages exist which use this visual technique. For example, "Statistics for the Terrified" is a useful package that teaches many statistical ideas using this method (Morris & Szuscikiewicz, 2001). However, it is not sufficiently flexible and does not allow different data sets to be used. The author has developed a number of Excel spreadsheets, each of which looks at a different statistical idea. Sample data are available from a variety of disciplines and these can be simply pasted into a worksheet. Alternatively, the tutor can use their own data as an example. This paper will put the case for "visual" learning and the use of appropriate data sets in developing students' understanding of statistical concepts. It will suggest different ways in which the material can be used as a learning aid and give brief examples from two of the spreadsheets.

VISUAL LEARNING

The development of statistical literacy is not achieved by using a single learning method. Snee (1993) points out the importance of giving students a variety of learning methods and strongly advocates the use experiential learning. "Visual" understanding is one important skill that students need to develop and this is best achieved by students actively participating in their learning. For example, we encourage students to carry out an initial exploratory data analysis and to comment on key features in the data such as outliers but, we seldom spend much time in saying how to recognise an outlier and whether it is likely to cause us a problem. By visually altering the data of a dotplot the student can see the impact of an outlier on the mean, standard deviation and the confidence interval.

This is illustrated in Figure 1 which shows a spreadsheet containing univariate data; it also includes summary statistics and graphical output. By dragging a data value, the student can see the effect of an outlier on both the summary statistics and the graphical summaries (boxplot and confidence interval). This method allows the student to be an active learner, able to investigate the effect of changing data values. The results are immediate and have a strong visual impact.

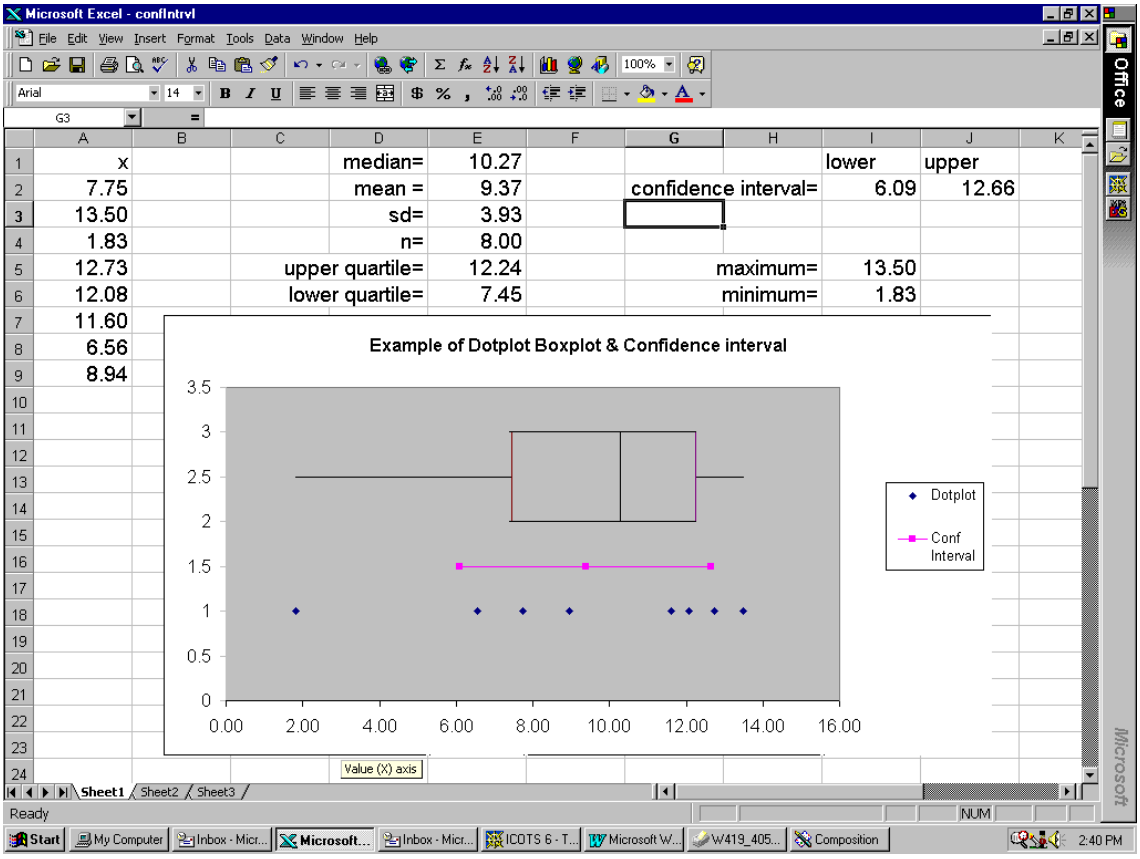

Figure 1. Spreadsheet Containing Univariate Data.

APPLICABLE PROBLEMS

When teaching students from other disciplines it is important that they see the relevance of the statistical techniques to their own subject. Singer and Willett (1990) advocate the use of real data and real problems. Using examples from a different discipline or simulated data that is not placed in an applicable and familiar context is often seen as a "turn-off" by students. Students who see the usefulness and relevance of a statistical technique are more likely to be motivated to become statistically literate. Tutors can download the spreadsheets from the Internet (www.it.bton.ac.uk/staff/rcc2) and simply paste alternative data sets into the spreadsheet at the appropriate place. Some useful data sets are provided on the web site or alternatively tutors can use their own data sets. This flexibility means that course tutors can quickly tailor the spreadsheets to suit their own courses.

FLEXIBILITY IN LEARNING

The spreadsheets can be used in a variety of ways to aid learning. In a traditional setting they could be used within a statistics lecture to explain and illustrate a new statistical idea. Alternatively, students can use the spreadsheets in a practical class. They can visually see the effect of changing the data values – they take an active rather than a passive role in their learning.

The material will also be used by distance learning students in a new master's programme that is being developed by the University. It is important to provide such students with a variety of approaches for developing statistical literacy; and especially a means to carry out learning in an active rather than a passive manner. The spreadsheets will be one of the techniques used to provide this.

STATISTICAL TECHNIQUES

Although both Hunt (1996) and Nash and Quon (1996) point out the limitations of using Excel spreadsheets in teaching statistics, the Excel package has built-in functions for many of the basic statistical techniques. Additionally features such as dotplots and confidence intervals have

been developed for particular topics although they are not standard options within Excel. Spreadsheets have been developed for the following topics that are typical of the kind of topics that are included in many introductory statistics courses:

- Univariate analysis
  - summary statistics
  - dotplots and boxplots
  - confidence intervals
- Bivariate Analysis
  - regression
  - correlation
  - model validation
- Inference
  - single sample t test
  - 2 sample paired and independent t tests
  - 1 way and 2 way ANOVA

Another example is shown in Figure 2. Experience of teaching Anova to non-statisticians has shown the importance of using confidence intervals to illustrate the significance of the results. The effect of changing data values by dragging and how this impacts on both the confidence interval and the Anova table (Table 1) has a strong effect on students' experiential learning.

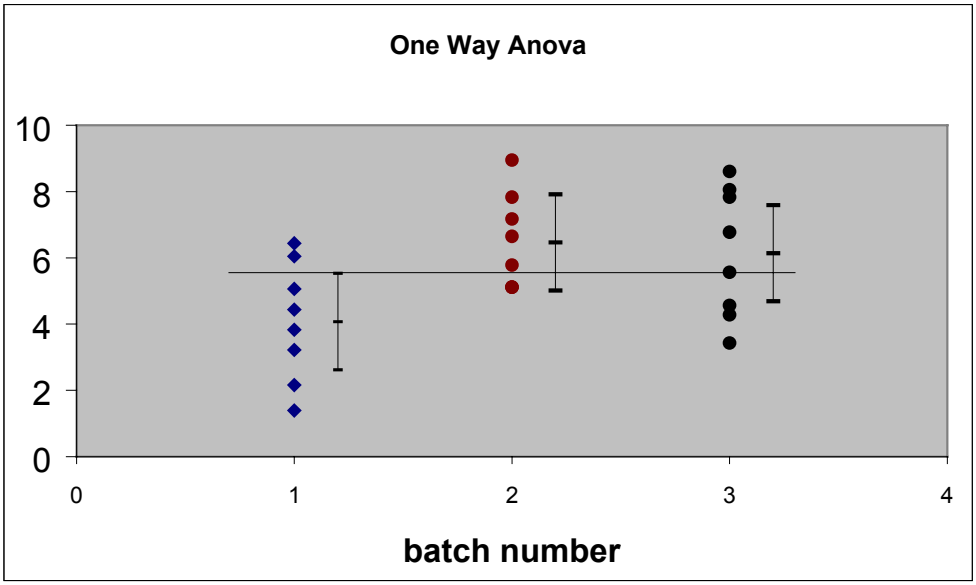

Figure 2. An example of a One way Anova.

Table 1
ANOVA Table

Analysis of Variance Table					
	df	SS	MS	F	pvalue
batch	2	26.87	13.44	4.44	0.025
error	21	63.49	3.02		
total	23	90.36			

DISCUSSION

The use of interactive visualisation can play an important role in developing statistical understanding. It is an active learning method. Snee (1993) states the value of "right brain"

learning styles. These are activities that encourage the student to discover how things are put together and organised - they require the active participation of the student. Using spreadsheets in the way proposed in this article is such an activity and allows the student to learn by practising and experimenting.

Bradstreet (1996), Hawkins, Joliffe and Glickman (1992) emphasise the need to use real data in a relevant context. If students see the relevance of a statistical idea, they are more likely to be motivated to learn about it. Excel is noted for the ease with which data can be transferred into a spreadsheet and, it was with this in mind, that the spreadsheets were developed to allow alternative data sets to be placed in the spreadsheet.

Excel is a standard package on most PCs and therefore students can benefit from this portfolio of pre-prepared spreadsheets without having to buy additional software. The spreadsheets can be downloaded from the author's website at (www.it.bton.ac.uk/staff/rcc2).

Nash and Quon (1996) stress the limitations of spreadsheets in teaching statistics. The author does not propose that Excel should solely be used within a statistics course, but Excel does offer this very useful facility of being able to graphically change data values. It should therefore be used as part of the statistics teachers' armoury to improve students' learning experience. Hopefully this will help lead to a more statistically literate society.

REFERENCES

Bradstreet, T.E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *The American Statistician*, *50*, 69-78.

Hawkins, A., Joliffe, F., & Glickman, L. (1992). *Teaching statistical concepts.* London: Longman

Hunt, N. (1996). *Teaching statistics with Excel 5.0* [Online]. Available at http://www.stats.gla.ac.uk/cti/activities/reviews/96_05/excel.html

Nash, J. C., & Quon, T. K. (1996). Issues in teaching statistical statistical thinking with spreadsheets. *Journal of Statistics Education* [Online], *4(1).* Available at http://www.amstat.org/publications/jse/v4n1/nash.html

Morris S., & Szuscikiewicz J. (2001). *Statistics for the terrified.* Concept Stew Ltd Educational Software Development

Singer, J.D., & Willett, J.B. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. *The American Statistician*, *44(3),* 223-230.

Snee, R.D. (1993). What's missing in statistical education? *The American Statistician*, *47*, 149-154.