Excel as an Aide in Teaching Measurement and Research Methods

Dr. Larry R. Nelson

Faculty of Education, Language Studies, and Social Work

Curtin University of Technology

Western Australia

At last year's conference, I spoke of the richness of the resources now on hand for students and teachers of research and measurement, suggesting that we've "never had it so good".

This year I will continue to explore this theme, looking particularly at some of the recent Excel-based resources now readily available over the Internet.

Background

I have taught measurement, statistics, and data analysis courses for a large portion of the past 35 years. By and large, my students have come from the social sciences, particularly from education and psychology. I generally find that students in the social sciences will often have a negative attitude toward quantitative methods, a fact that has at times added more than a small bit of extra challenge.

My teaching has always had a practical bent. I make an effort to give students real-life examples, authentic-like exercises, and I fully integrate technology in my classes. I cannot always assume that students are as technology-literate as I would like; at times it's necessary to set aside instructional time in order to deal with specific technology topics, as I will mention below.

Using sample data sets

The data analysis programs we use these days have shown a steady drift to table-based data designs, that is, to using a two-dimensional matrix to house data. These matrices, or tables, usually have columns representing variables, or measures, and rows representing observations.

Both of the major data analysis systems used in the social sciences, SPSS and SAS, use table-based data designs. So does Lertap, the classical item and test analysis system developed at Curtin University of Technology. (For references, see www.spss.com, www.sas.com, and www.larrynelsonstuff.com/lertap/.)

Accordingly, it has become one of my teaching goals to have students become familiar with table-based data designs. To do this, I rely on two sample data sets. One of these, the "CEQ" set, involves the responses of twelve school principals to a ten-item, Likert-style survey, measuring attitudes towards the use of computers in schools. The other sample data set, "TenItemCognitive" is based
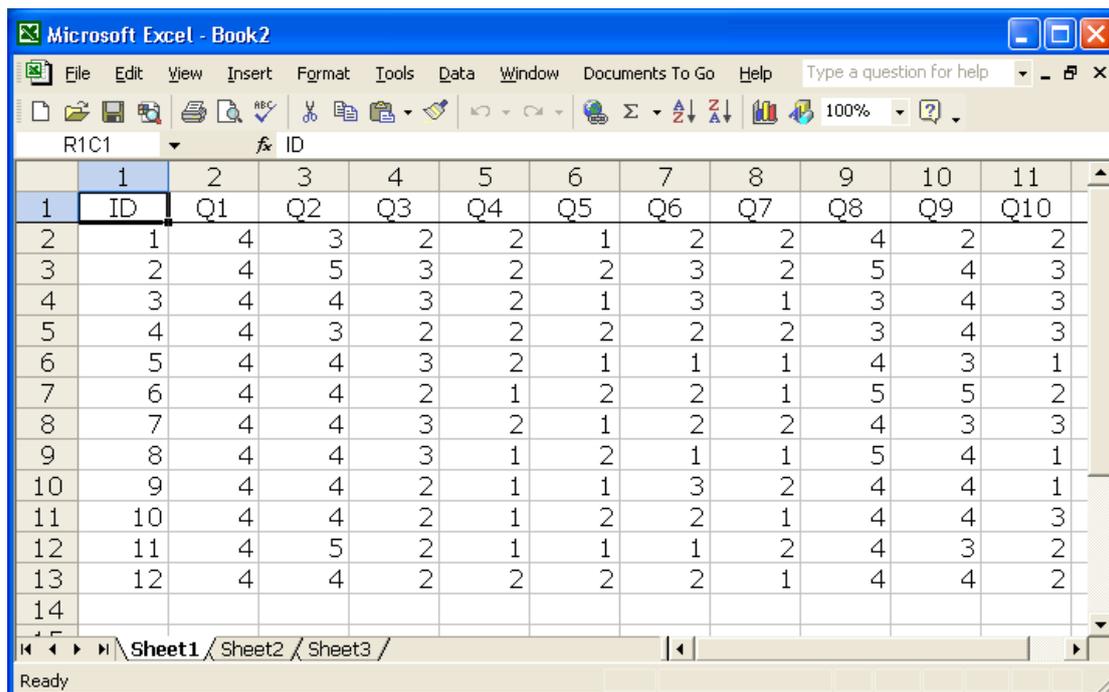
on an example involving the responses of fifteen students to a ten-item multiple-choice test.

I have placed both data sets on a web server. Their respective URLs are:

www.larrynelsonstuff.com/Documentation/Samples/CEQ/CEQ1.DOC

www.larrynelsonstuff.com/Documentation/Samples/TenCogs/TenItemCognitive.doc

A visit to these URLs will reveal that the sample data sets are just that: I have put on the web server actual answer sheets. I give students the task of entering item responses into a suitable table, and I have them use Excel to do so. A typical student-entered Excel data table for the CEQ data set looks like this:

Above I mentioned that learning how to use technology forms part of many of my classes. In this case, I have students actually use the URLs in order to access the sets of answer sheets, giving some of the students their first opportunity to download something from the Internet. I suggest to students that they print the sets of answer sheets in order to best simulate what happens in real life (we administer a test or survey, collect answer sheets, and then process data).

Basic statistics and graphs

Once item responses have been placed in Excel, there are often a number of illuminating exercises which students can undertake. In the process they become more familiar with Excel, and with fundamental basic statistics.

For example, I ask students to find indices of central tendency for the items, such as the mean and median. They're also directed to find indicators of variance, and to determine the covariance and correlation between any two items of their choice. How are students to do these things? By using Excel's in-built functions, such as AVERAGE, MEDIAN, STDEVA, STDEVP, as so on.

Finally, I introduce students to the use of graphs by asking them to plot item means. The first of the screen snapshots below shows the item means; the

second snapshot is a graph of an Excel line chart made to indicate the distribution of means:

	1	2	3	4	5	6	7	8	9	10	11
1	ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
2	1	4	3	2	2	1	2	2	4	2	2
3	2	4	5	3	2	2	3	2	5	4	3
4	3	4	4	3	2	1	3	1	3	4	3
5	4	4	3	2	2	2	2	2	3	4	3
6	5	4	4	3	2	1	1	1	4	3	1
7	6	4	4	2	1	2	2	1	5	5	2
8	7	4	4	3	2	1	2	2	4	3	3
9	8	4	4	3	1	2	1	1	5	4	1
10	9	4	4	2	1	1	3	2	4	4	1
11	10	4	4	2	1	2	2	1	4	4	3
12	11	4	5	2	1	1	1	2	4	3	2
13	12	4	4	2	2	2	2	1	4	4	2
14											
15	Mean	4.00	4.00	2.42	1.58	1.50	2.00	1.50	4.08	3.67	2.17
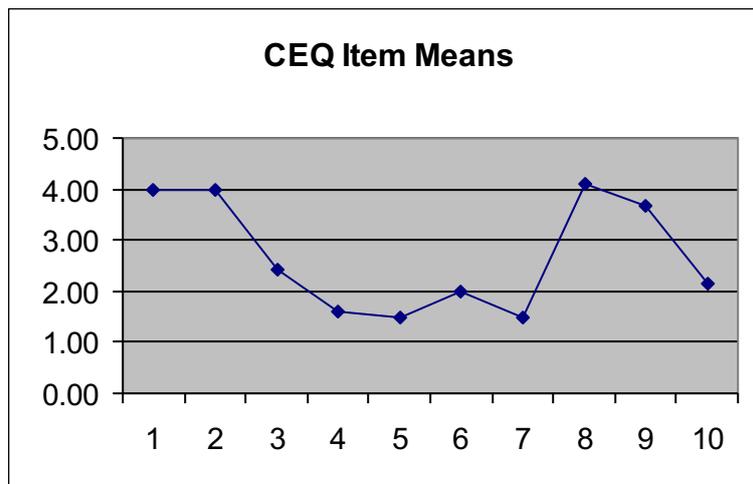
These are practical, introductory exercises. For many in today's audience, these may be seen as rather simple and straightforward – but I have found that social science majors are often challenged by them, and, almost as often, pleasantly rewarded when they see something as visually meaningful as the graph above.

The CEQ data set involves ten Likert-style questions of mixed polarity. Some of the ten items have stems which are positive statements (e.g., *Computers provide motivation for students to learn.*); others have negative stems (e.g., *A computer is an unnecessary luxury in most schools.*). I give students the task of reversing the negative items, and making a new plot of item means. Again this is something which may seem easy enough, but I find my beginning students will often find the going slow.

After looking at the CEQ data set, I turn to the other sample, the one involving ten cognitive items. Again I have students download, print, and enter item

responses in Excel. And then what? Do I have them get item means? No – I can't. Unlike the CEQ items, the multiple-choice items have to be scored first.

The matter of scoring cognitive items provides excellent material for discussion in beginning measurement classes. I pose the possibility of simply scoring the items on a right/wrong, one/zero basis, and then entering item scores in Excel, after which I can of course get item means, and plot them. But this process, often found when preparing for IRT[1] processing, omits the critical matter of looking at the effectiveness of item distractors, a fact that leads us to look for an item analysis program of some sort.

Basic Excel-based item analysis

At Curtin University we've developed a classical item and test analysis system called "Lertap"[2]. It has been programmed as an Excel application – essentially, it's nothing but a collection of Excel macros, written in VBA[3].

Here's how the item responses from the TenItemCognitive data set look when entered in a Lertap-compatible Excel workbook:


In the CEQ example, data were recorded in a worksheet with a default Excel name: "Sheet1". In a Lertap-compatible Excel workbook, data are in the "Data" worksheet, with the top two rows of this sheet reserved for a title, and for column headers.

---

[1] Item Response Theory
[2] Laboratory of Educational Research Test Analysis Package
[3] Visual Basic for Applications

All Lertap-compatible Excel workbooks have a "CCs" worksheet, such as the one shown here:
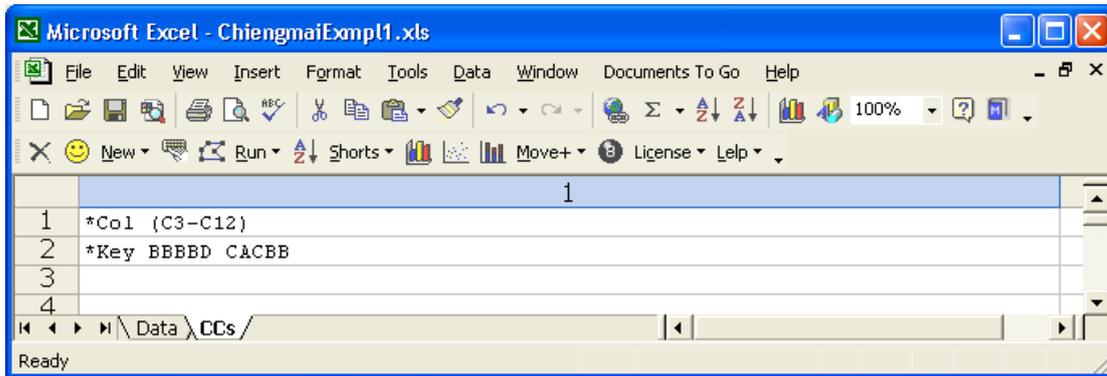
The purpose of the CCs worksheet is to provide basic job control information. In this example, the first line tells Excel that item responses are found in columns 3 through 12 of the Data worksheet, while the second line gives the keyed-correct answer for each of the ten items.

Students have little trouble following this example; they are generally able to enter data and set up the two CCs lines without requiring help – the jump from the CEQ sample, with affective data, to the TenItemCognitive example is an easy, intuitive one.

What next? I instruct students to use the options which drop-down from the Run menu:
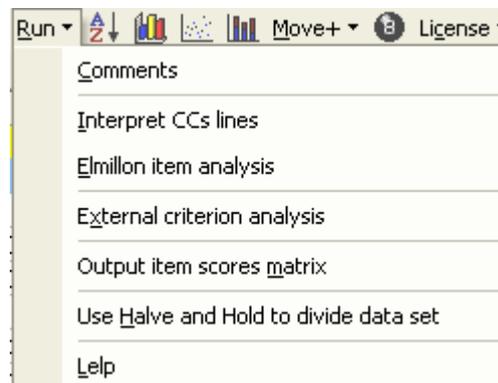
The Elmillon item analysis option gets Excel to execute several macros, a process which involves the creation of three "reports" with classical item and test statistics. The reports, which are Excel worksheets, include several indices of item discrimination[4], the usual index of item difficulty, indicators of distractor adequacy, and a variety of summary test statistics. Examples of these reports may be seen at this URL:

www.larrynelsonstuff.com/Documentation/Chapter2.pdf

I very much feel that graphics are a core component of data analysis, and, at this point, I ask students to use Excel's numerous chart options to plot item difficulties and discrimination, using the "Stats1b" item statistics report output by Lertap.

---

[4] Point-biserial, biserial, and the upper-lower index.

Lertap's Elmillon macros also create an Excel worksheet with test scores. I have students plot these – the current version of Excel does not have an adequate histogrammer, so students use a special histogramming macro available from Lertap's toolbar. (An interesting exercise with more capable students is to have them search the Internet for guidance on how to make boxplots in Excel.)

These straightforward exercises have as much to do with data analysis proce-dures as they do with actual measurement topics. By the time students have worked through the CEQ and TenItemCognitive samples, they will have learned much about Excel and the use of data tables, an experience which prepares them well for subsequent work with SPSS, SAS, and Lertap. They will have garnered an appreciation of graphics, and gained experience with making charts. And, not least by any means, they'll have had ample opportunity to study many of the fundamental measurement concepts pertaining to classical item and test analysis.

I have students use the free student version of Lertap for this work. It may be obtained by visiting the following URL[5]:

<div align="center">

www.larrynelsonstuff.com/lertap/

</div>

More Advanced Excel Applications

The Excel-based Lertap system I have mentioned in this paper is new, having gone into general circulation in 2002. Not only is it new, but additional features are being added frequently. I myself have not yet had the opportunity to use some of its advanced features with students who have survived my initial test and measurement course.

I can however mention some of the things I have in mind. There are two papers on our Lertap website which exemplify more advanced measurement topics. One of these involves a presentation of how to use Lertap, and Excel, in a parallel-forms reliability study. The respective URL is:

<div align="center">

www.larrynelsonstuff.com/Documentation/PFExample1.htm

</div>

Matters related to test validity, and, especially, concurrent validity are addressed in a paper which may be found at:

<div align="center">

www.larrynelsonstuff.com/Documentation/ValidationExample1.htm

</div>

Given sufficient time, I hope to make freely available data sets which will let students replicate the work seen in these two papers. Note, however, that the non-student version of Lertap will be required for such work – the student version is programmed to process data sets with a restricted number of cases. At present, the cost of the unrestricted version for bona fide students is on par with the price of a contemporary text book. In order that the system falls within the economic reach of more students, next year we expect to make it possible to rent the software on a monthly or semestral basis.

Lertap has the ability to randomly divide a data set into halves. An interesting exercise would be to have a group of students take copies of the same data set, divide it into halves, and generate classical indices of item difficulty and discrimi-nation. The class would then meet to compare notes – how much sample vari-

---

[5] In 2013 the Student Version was replaced by the "Mini" version.

ance is there in these statistics? We always used to think that there'd be a lot, but recent research suggests this is probably not the case.

(See: www.larrynelsonstuff.com/Documentation/IRTandCTTComments.htm.)

Excel is, or course, a widely-used application. It has been around for several years, long enough for many people to have created a variety of macros, some of which are now freely available via the Internet.

For example, the Foxes Group in Europe offers free macros in the form of an Excel Add-In. Some of these will support a principal components analysis, and a principal factors analysis, using one of Lertap's standard correlation matrices as input. We've now enhanced Lertap so that it will look to see if a Foxes Group Add-In is present, and, if so, it'll access its eigenvalue decomposition routine to output the latent roots of the correlation matrix, as exemplified below:
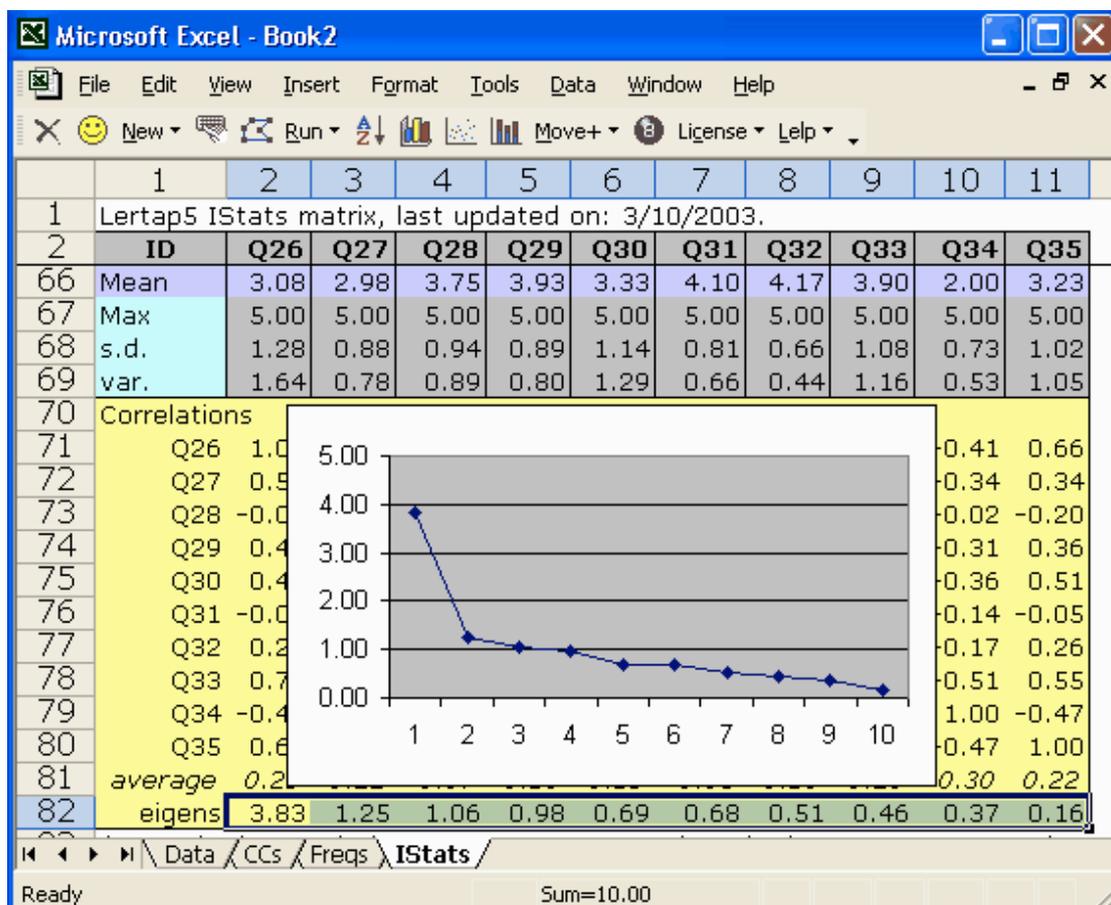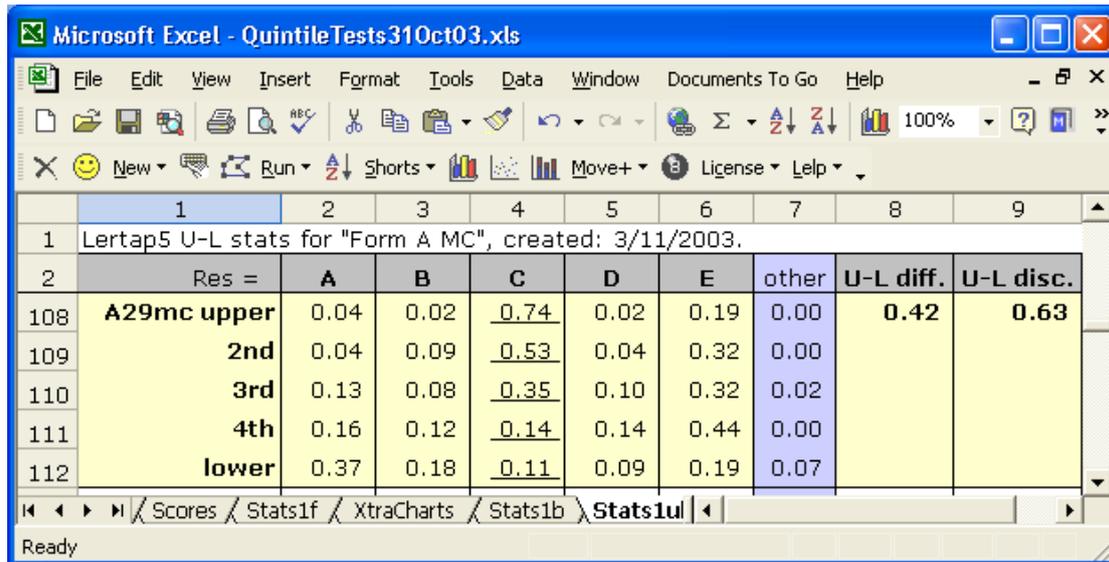

The screen snapshot above shows part of an "IStats" matrix (created by Lertap), with a row of "eigens" showing at the bottom. I've had Excel overlay the matrix with a line graph of the eigenvalues. Degree of difficulty? Dead easy.

Advanced classes could make more use of the Foxes Group Add-Ins. Exercises might include using the Add-Ins to output a complete components or p-factors analysis, perhaps followed by a rotation of some sort (this would not be at all difficult to do; the Add-Ins are sufficiently documented to make this reasonably simple).

Item option trace lines

When speaking of advanced topics, let me insert this special small section in order to draw particular attention to one of the very latest additions to Lertap: quintile plots. This addition was suggested by a user in the United States, and I'm especially pleased with the potential of the results.

Consider the following Lertap-generated Excel worksheet, and data related to an item, "A29mc", which formed part of an aptitude test developed years ago in New Zealand:


The item had five options, with C being the keyed-correct answer. Lertap has divided the test takers into quintiles, into five achievement groups, each one housing 20% of the sample size. The table clearly shows that the "lower" group, those in the bottom 20%, were not very capable of identifying the right answer – in fact, only 11% of the respondents in this group, the lowest achievers, got the item right, compared with 74% in the top, or "upper" group.

Is this clear? Could it be made clearer? Sure thing – have a look at this Excel chart:
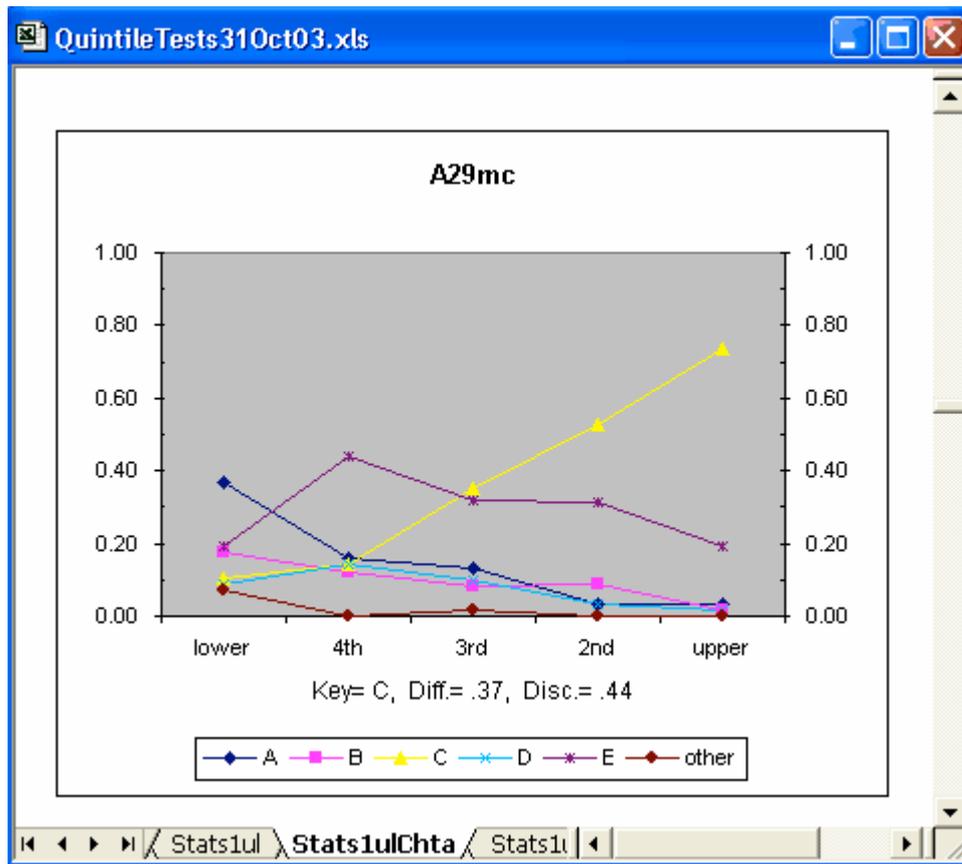
Isn't this Excel chart a beauty? It graphically captures the action. Each of the chart's lines corresponds to one of the five options. The yellow line, the one with the small triangles, represents the correct answer. It rises as we go from left to right, that is, as we go from the lower achievement group to the upper one. The graph clearly indicates a strong relationship between achievement level and ability to identify the item's correct answer, and it shows how successful the distractors are in foiling weaker students – for example, option A pulled off many in the lower group, while option E functioned as a good foil in the three middle quintiles.

Now have a glance at the chart below, again based on item A29mc:
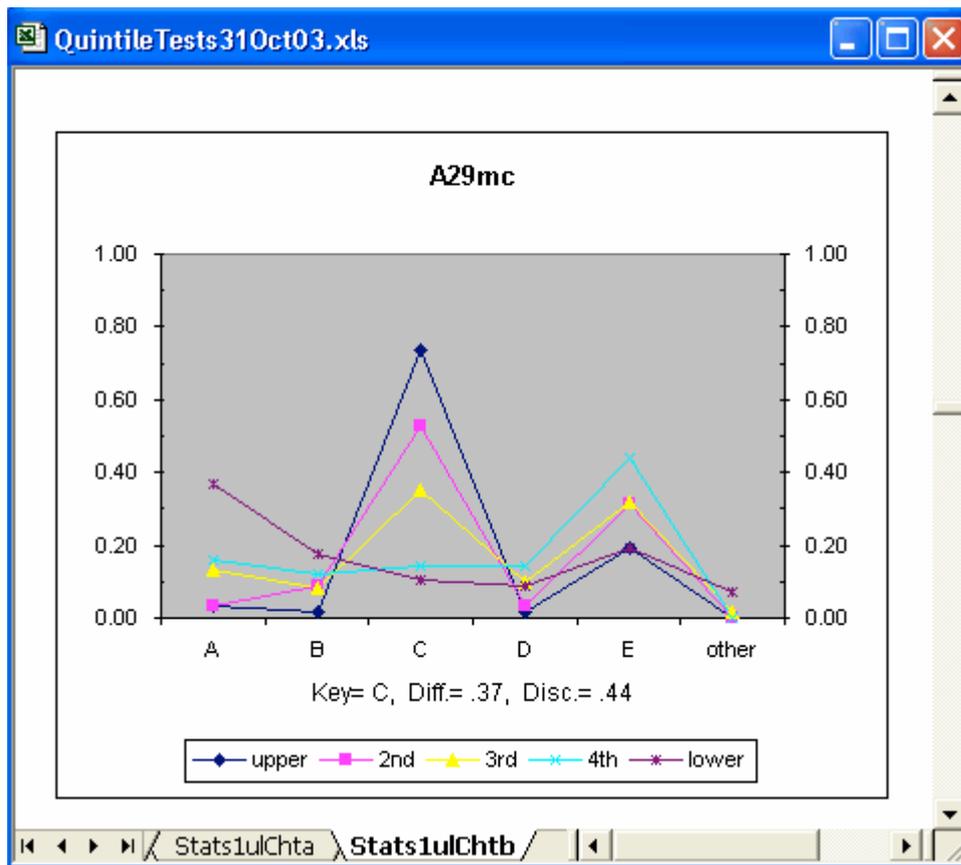
This chart's lines correspond to the quintiles; there's one line for each of the five achievement levels. We see that option A was most popular with the three lowest quintiles, which is how we want a good distractor to work. It also clearly shows how the proportion of students able to identify the correct answer, C, steadily decreases as we drop down the achievement levels.

Graphs such as these are similar to the "option trace lines" proposed by Wainer (1989). Given an appropriate data set, they may be used to look for "DIF", differential item functioning – the example I've given here uses achievement quintiles to define the groups, but Lertap will allow the groups to be defined on the basis of something else: perhaps a categorical variable, such as gender or regional area.

I have plans to make a suitable "DIF-able" data set available via the Lertap website. I could use help here; at the moment I have only one potential data set, and it's not the best – perhaps someone in the audience might be able to contribute a data set?

From Lertap to IRT

We all know that there's more than one way to assess how a test and its items are functioning; we've all heard of item response theory, and no doubt many of us have studied IRT methods in some detail.

Lertap does not currently provide support for any of the IRT models. But we've started to build in interfaces which will ease the task of copying data out of Lertap and into an IRT program. Thus far, for example, Lertap has been programmed to

output a text-based item-scores data file in a format ready for input to Bilog-MG (du Toit, 2003). An example is shown below:
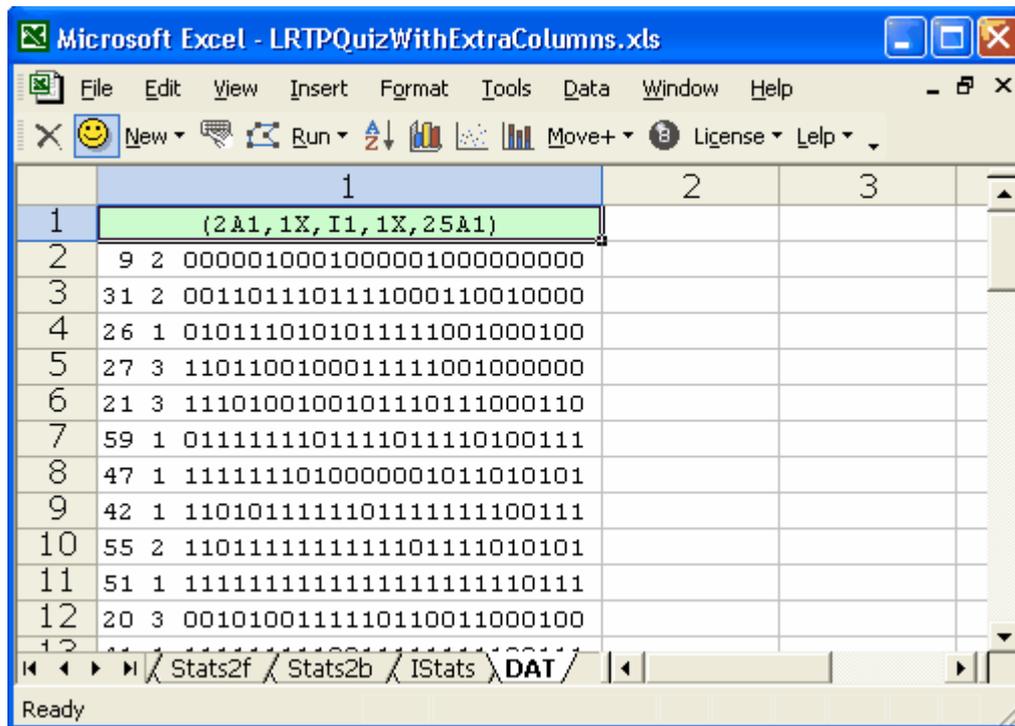
Some other Excel resources

I am sure that there would be several websites offering similar Excel-based resources. In the process of preparing this paper, I communicated with two colleagues in particular: Gail Delicio at Clemson University, and Glenn Fulcher at the University of Dundee.

Professor Delicio's website for measurement classes is at this URL (refer to the "Spreadsheets" topic once into this site):

http://www.gcd.clemson.edu/Main808/Notes808/

The Excel resources on Professor Fulcher's website are still much in the process of construction, but will (I understand) be built onto the following site:

http://www.dundee.ac.uk/languagestudies/ltest/ltr.html

I apologise to colleagues here in Thailand who may also have made some of their teaching resources available over the Internet, but with whom I was unable to establish contact prior to the conference.

Summary

I have found Excel to be eminently suitable for use in my measurement and data analysis classes. It's not only suitable, but a very effective and readily-available tool for introducing students to contemporary data analysis methods. Excel's fundamental data table design, coupled with useful chart capabilities, easily leads students down paths which will pave the way for their later application of such systems as SPSS and SAS.

In recent years I have used Excel to develop a new version of Lertap. This version is simply a collection of Excel macros useful in classes dealing with classical item and test analysis. Our Lertap website at Curtin University (www.lertap.curtin.edu.au) contains a number of resources designed for use in test and measurement classes. Most of these resources are available at no cost, and I most cordially invite you to use them. I also request your participation in this work – we have a need for new data sets, and I keenly welcome suggestions as to how our resources may be improved.

References

Du Toit, M. (Ed.) (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT.* Lincolnwood (IL): Scientific Software International (www.ssicentral.com).

Wainer, H. (1989). The future of item analysis. *Journal of Educational Measurement, 26*, 191-208.