How to make statistics interesting, relevant and fun!

Anne Morris
Department of Information Science, Loughborough University, Loughborough, UK
E-mail address: a.morris@lboro.ac.uk

Copyright © 2013 by Anne Morris. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: http://creativecommons.org/licenses/by/3.0/

Abstract:
This paper describes how statistics are taught in the Department of Information Science (Loughborough University) and particularly focuses on step-by-step SPSS guides and data sets that have been specifically produced for LIS students. The teaching of statistics for Masters’ students forms part of a ”short-fat” Research Management module which prepares them for dissertation research. Prior to statistical teaching, the practical relevance of collecting and manipulating quantitative and qualitative data in the sector is illustrated. This helps to inspire LIS students to develop the necessary skills to use statistical methods with confidence and competence. Descriptive and inferential statistics are first taught in lecture and tutorial format to deliver the basic theories and to provide hands-on experience of manipulating data. Students are then introduced to SPSS which reinforces statistical theory while at the same time being fun to use. The guides (one for Masters’ students and a more in-depth version for PhD students) enable the weakest students to obtain success while stretching the more able students. Relevant ”real” data sets are used. Examples includes the top 100 best-selling books from 1989 to 2010; survey responses to students’ use and views of VLEs; UK library survey data on Open Access; and IT piracy worldwide. The guides and data sets have been successfully used by two cohorts and have been adopted by other disciplines in UK Universities. Feedback of the module has been excellent – “I have always hated numbers, I now know why they are important and I am much more confident about statistics”.

Keywords: Statistics teaching, SPSS guides, Statistics for Information Science

1 INTRODUCTION
1.1 Negative perceptions about statistics
Making statistics relevant, interesting and fun is very difficult because it is well recognised that many students, no matter what discipline they are from, have negative attitudes towards statistics (Green et al 2001; Lazar, 1990; Onwuegbuzie & Wilson, 2003; Maschi et al 2007). Common attitudes include, fear, anxiety, cynicism and contempt (Hopkins et al, 1996). Often, students think that statistics are too difficult to learn or just not worth putting the effort in to master them (Schau, Millar and Petrocz, 2012).

Several surveys have been devised to measure student attitudes to statistics. However, only four of these have proved to have validity (content, substantive, structural and external
validity), as well as internal consistency: Statistics Attitude Survey (SAS); Attitudes Towards Statistics (ATS), Survey of Attitudes Toward Statistics -28 (SATS-28); and Survey of Attitudes Toward Statistics -36 (SAT36), (Nolan et al, 2012). According to Nolan et al (2012) there are three common themes that are measured: students’ feelings about statistics, perceived value or usefulness of statistics and perceived ability to understand statistics. SAT36 is perhaps the most commonly used survey tool. It contains 36 items that assesses six components: Affect (positive and negative feelings), Cognitive Competence (knowledge and statistical skills), Value (usefulness, relevance and worth), Difficulty (perceived difficulty), Interest (perceived interest), and Effort (amount perceived to learn statistics) (Schau and Emmiolgu, 2012).

Studies using these attitude instruments have reported that negative attitudes towards statistics affect students’ achievement, whether they persist with their studies, the development of their statistical skills, the use of statistics beyond their course, and whether they study statistics in the future (Gal, Ginsburg, & Schau 1997, Hilton, Schau & Olsen 2004; Waters, Martelli, Zakrjasek & Papovisch 1988). Consequently, reducing negative attitudes towards statistics is extremely important.

1.2 Measures to reduce negative attitudes towards statistics

Students will only use statistics in life, at work or while studying if they believe they are useful, can use them effectively, find them interesting and they like the subject (Schau and Emmiolgu, 2012). So how can we improve students’ attitudes towards statistics? They have been a few studies that have examined the effect of various interventions. For example, Harlow, Burkholder and Morrow (2002) investigated the effect of peer mentoring, small group problem-solving, hands-on applied projects and student feedback on lecture clarity, and reported a decreased anxiety and increased self-efficacy. Posner (2011) compared two groups, one using his PARLO System (Proficiency-based Assessment and Reassessment of Learning Outcomes) and a control group. Lecturers for the PARLO students stressed learning objectives, assessed students’ performance using proficiency-based scoring, and allowed students to resubmit coursework when the first submission did not meet the required grades. Mean attitudes, as measured by the SAT36 survey, were higher for the PARLO group compared to the control group. The intervention used by Carlson and Winquist (2011) was a workbook curriculum. Students were expected to read recommended material before and after classes and work on problem-solving activities during supervised sessions. Using the SAT36 survey before and after the intervention they found significant mean increases in Affect and Cognitive Competence scores and a decrease in Effort scores. The Difficulty scores also moved from somewhat difficult to neutral.

Meletiou-Mavrotheris et al (2007) compared student attitudes and knowledge between a group having a technology-based statistical course with a group that were conventionally taught. The study found that students on the technology based course were more motivated, enjoyed the course more and recognised the role technology can play in statistics. However, no differences were found between the groups on students’ grasp of fundamental concepts in statistics. Indeed they found that some students were frustrated at their failure to grasp the underlying theory underpinning the computer activity. This shows the obvious need to combine underlying theories of statistics with the practical aspects of computing data with or without technology. Marson (2007) investigated a number of intervention strategies and found that giving students immediate feedback to enable them to correct errors and allowing them to work at their own pace was the most effective in reducing anxiety. Lalayants (2012) surveyed 195 graduate social work students and identified the following strategies to reduce
negative perceptions of statistics: instructors should be attentive to students’ anxieties, the pace of the course should be slow, and additional tutoring should be provided where needed. The author also suggested three areas for enhancing statistics learning: “(1). Fostering an environment that is nonthreatening, friendly, and conducive to learning; (2). Clarifying statistical concepts in plain terms; and (3) offering practical skills relevant to students’ field of specialization.” The latter should reflect the usefulness of statistics to the students’ field of study using practical case studies as examples.

1.3 Other factors that influence students’ statistical performance
Previous research have shown that a number of factors influence statistical performance apart from attitude towards statistics. These include gender, spatial ability, pedagogical approaches, prior knowledge, and mathematical ability (Griffith et al 2012). Griffith et al also report findings from a survey of 684 undergraduate students and found that their performance in statistics was affected by their teachers. They advocate the need for teachers to reduce fear at the beginning of the course, and stress the usefulness of statistics both for their immediate studies and in their future career. In terms of pedagogical approaches, Strangfeld (2013) advocates that statistical courses should be organized around a student led research project whereby students are expected to generate topics, develop research tools, collect the data and conduct the analysis. Gordon and Gordon (1992); Moore (1992, 1997) showed that student learning of statistics is improved when instructors use active learning, real data, conceptual understanding, and technology. Research has also shown that materials that utilise different learning styles such as auditory, visual, and kinaesthetic, are more likely to interest students and result in improved performance (Searson and Dunn 2001; Uzuntiryaki, Bilgin,and Geban 2003; Clark and Mayer 2008).

Aliaga et al (2010) reported the outcome of The Guidelines for Assessment and Instruction in Statistics Education (GAISE) project carried out between 2003 and 2005. They emphasised the need to move away from a traditional lecture only format to a more activity based approach and from just using technology for calculation purposes to one that helps students understand basic concepts (MacDaniel and Green 2012). The project made six recommendations for the teaching of statistics including:

- Emphasise statistical literacy and develop statistical thinking.
- Use real data.
- Stress conceptual understanding, rather than mere knowledge of procedures.
- Foster active learning in the classroom.
- Use technology for developing concepts and analysing data.
- Use assessments to improve and evaluate student learning.

These and the other recommendations in the literature are, therefore, the key to making statistics courses relevant, interesting and fun.

2 RESEARCH MANAGEMENT MODULE
The teaching of statistics for Masters’ students forms part of a “short-fat” Research Management module which prepares them for dissertation research undertaken between May and September. A “short-fat” module is one where the students are taught in-depth in a short space of time. In this case the Research Management module is taught for six full days over a period of two weeks by the author and other lecturers in the Department during January and
February. Students are then given another three weeks to complete an assignment – a research proposal on their intended research.

The Research Management has therefore to cover a lot of ground to give them the confidence, skills and knowledge to enable them to tackle the dissertation module effectively. Specifically, the aims of the Research Management module are to:

1. Provide an overview of the research process in terms of theory, topic generation, proposal writing, literature searching, data collection, analysis, interpretation of results and the communication of findings;
2. Prepare students for work on their dissertation on an information, knowledge management or related topic;
3. Provide an understanding of the role and value of research as applied to information services;
4. Enable students to develop an appreciation of the principal methods and strategies employed in information research;
5. Enable students to analyse, interpret and evaluate research results;
6. Provide transferable skills for subsequent employment.

On completion of this module students should be able to: demonstrate knowledge and understanding of the research process; formulate research strategies and write research proposals; select appropriate research methods; select and apply statistical techniques; interpret data; evaluate research reports/papers; undertake information research with supervision; use simple statistics’ present data in a meaningful way; write research reports and manage simple research projects.

The key transferable skills include demonstrating an ability to plan and execute projects and work to deadlines within groups and individually; evaluating and synthesising a body of knowledge; selecting between, and use of, a variety of printed and electronic information resources; and communicating in a clear, systematic and concise way for a range of audiences. The skills are not unlike those for required for other Research Methods modules in the UK (Morris, 2005).

2.1 Content of the Research Management module

Each of the six days is split into lectures and tutorials and practical classes. Included are sessions on introduction to research methods, the process of research, how to undertake a systematic literature review, project planning, qualitative methods (research interviews, focus groups, historical research, diaries, observation, Delphi method, Discourse Analysis, Repertory Grid), analysis of qualitative data, quantitative methods (surveys, questionnaire design, sampling, social network analysis, content analysis, performance data, usability testing, experiments, and bibliometrics). In addition, there are sessions that cover research ethics and current research undertaken in the Department. A wide variety of real research examples and case studies are provided by all the lecturers which stress the relevance of various methods and the need for, and usefulness of, systematic qualitative and quantitative analysis thereby helping to reduce any negative attitudes students might have towards statistics. The tutorials also help ground the lecture material in reality and give them practice in applying research theory. For example, students are divided into groups and given real case based problem senarios to discuss and to formulate a plan of action research. In another
tutorial, groups of students are asked to critique a “poor” questionnaire asking for information about the University’s virtual learning environment. Discussion with the whole class then takes place highlighting the problems, giving immediate feedback. They are then asked to redesign the questionnaire. A prepared “improved” questionnaire is handed out at the end of the session and they are asked to complete to enable the data to be entered for use in the statistics component. This gives them ownership of the data, although, in practice the data used in the statistics practical sessions is that from a previous cohort. Throughout, students are recommended reading material and encouraged to be critical and reflective of published research; a necessary skill for undertaking literature reviews and for improving their understanding of the research process. It is stressed that any research methods used need to be valid, reliable and practical for the particular problem being investigated. It is also emphasised that it is extremely important to use appropriate sample sizes and to get the analysis right, because the use of incorrect statistics can result in meaningless conclusions.

The amount of time that can be devoted to the teaching of statistics in the module has been vastly reduced over the last two decades. Currently, students only have nine hours instruction; six hours devoted to theory and manual hands-on problem solving and three hours for SPSS training and use. The first two hours are devoted to sampling techniques; the first, a lecture, covers the theory, and the second, a practical session, requires students to answer set questions. The correct answers are given at the end of the session and students are encouraged to ask for clarification of any answers they have wrong. Another lecture covers basic statistical annotation, data types and descriptive statistics. Again this is followed by a practical session providing students with hands-on practice and immediate feedback. This same format is used for inferential statistics. Here, the theory and general process behind inferential statistics is provided together with the basics of the Chi-square test, t-test and correlation. Stress is put on why inferential statistics is needed, how it is used, the difference between parametric and non-parametric tests and their assumptions, and the meaning of significance. An example of the Chi-square test is worked through in the lecture. Students are provided with guided instruction sheets which enable them to complete a Chi-square test and a t-test manually. By the end of the session they understand the process but are ready to accept that such laborious calculations are better undertaken with the use of technology!

3 SPSS ENFORCEMENT – THE REWARD!
Once the students feel more confident about the theory of basic statistics and have had the opportunity of analysing data manually they are then required to use SPSS to analyse real data sets. Two guides have been created, one for Masters or Undergraduate Students in Information Science and the other for PhD students.

Both guides are split into two parts: Part 1 Reference Sections and Part 2 Tutorials. Part 1 is common to both guides and is 38 pages long. After an initial introduction, the main SPSS windows are discussed. This covers the data and variable view of the Data Editor window and the Viewer window used for displaying and handling output (see Figure 1).

Following this, step-by-step instructions, with graphics, are provided for obtaining on-screen help, creating variables and entering data, loading a data file and editing data in a data file and saving a data file. A brief description is also provided on how to sort and select cases and how to compute and recode variables. Part 1 concludes with an introduction to creating charts and graphs, notes of the Chi-square test and a list of data files for the tutorials.
Part 2 differs in the number, coverage and depth of tutorials; the PhD student version is more complex. The more basic guide aimed at Masters or undergraduate students in Information Science has 17 tutorials that can be worked through by students individually. These cover the following topics:

- Introductory basics: starting the SPSS program, loading and saving SPSS data files, creating new data files and inputting data.
- Descriptive statistics: finding the mean, median, mode, standard deviation and variance, creating one-way, two-way and three-way variable frequency tables, creating simple bar, clustered bar, stacked bar, pie and line charts and making histograms.
- Inferential statistics: Crosstabs and the Chi-square test, the t-test and correlation tests (Pearson correlation co-efficient and the Spearman correlation test for non-parametric data).
An example of part of one of the tutorials is provided in Figure 2:

**Tutorial T8: Descriptive Statistics**

The Descriptives procedure provides basic statistical measures of location and dispersion ('average' and 'spread') for ordinal and scale data similar to that available in Frequencies.

1. Load data file: **File → Open → Data → DATA03_Ls questionnaire.sav** (if not loaded)

2. Select **Analyze → Descriptive Statistics → Descriptives**
   - This opens the Descriptives window →

3. Using the blue arrow (or by double-clicking each) move the following variables into the **Variable(s) box**: Marks_modA, Marks_modB, Marks_modC, Marks_modD.

4. Click on **Options**.
   - This opens the Descriptives Options window shown here →
   - This shows the default statistics to be: Mean, Standard deviation, Minimum and Maximum →

5. Select **Skewness** →
   - (this is a measure of how 'flattened' to one side is the distribution of the data. N.B. positively skewed means having a longer tail to the right).

6. Click on **Continue** and click **OK** to produce the output below:

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std Deviation</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marks_modA</td>
<td>161</td>
<td>45</td>
<td>71</td>
<td>62.47</td>
<td>8.131</td>
<td>-.390</td>
</tr>
<tr>
<td>Marks_modB</td>
<td>155</td>
<td>42</td>
<td>67</td>
<td>60.40</td>
<td>6.922</td>
<td>-.429</td>
</tr>
<tr>
<td>Marks_modC</td>
<td>141</td>
<td>48</td>
<td>71</td>
<td>64.28</td>
<td>5.444</td>
<td>-.443</td>
</tr>
<tr>
<td>Marks_modD</td>
<td>161</td>
<td>42</td>
<td>61</td>
<td>62.96</td>
<td>14.512</td>
<td>.226</td>
</tr>
</tbody>
</table>

- The N column indicates that there is one missing value for Module C, i.e. 149 valid entries.
- Module D marks are very different from the others – with a much lower mean, a much larger standard deviation and a much wider range (Max – Min).
- The standard deviations are similar for Modules A, B and C, so they could be combined, and a parametric test such as a t Test (or ANOVA) could be used to compare them.
- Significant skewness occurs when the Skewness Statistic lies outside +/- 2 x Std. Error.

Figure 2: Example of a tutorial
There are four data sets required for the use of this guide:

1. 100 top-selling books during the period 1989 to 2010 as produced by the Guardian, Nielsen. The variables comprise: author, publisher, publisher group, number of books sold, sales value, recommended retail price, average selling price, type of binding, month of publication, year of publication, genre, and type.

2. The survey results to the questionnaire investigating student’s use and opinions about a University’s use of a VLE. The variables cover the demographics of the students, whether they accessed the VLE, for what purpose and for what modules, their opinions about the VLE, and the student results of three modules.

3. Survey results that identifies the effects and assesses the impact of Open Access to research outputs on pay-to-publish and self-archiving publishing models (conducted by SQWconsulting and LISU (Loughborough University)). Just under half of the questions in the survey are included as variables in this data set. These cover types of institutions, whether they have their own repository, if they include open access publications in their library catalogues, if authors are mandated to deposit material, the number of different types of items deposited and cost of running institutional repositories.

4. Data on IT piracy for 109 countries, published by Business Software Alliance, who define the IT piracy rates as the percentage of all software in use which is pirated. The variables included cover country, world regions, population, GDP and IT piracy rates and values for 2005 to 2010.

These are used in the tutorials mainly to help explain the procedural aspects of SPSS. There are also sets of questions relating to each of the data sets at the end of the tutorials to stretch students and get them to think about the data and how SPSS can be used to manipulate it to find answers to specific questions. The list of questions are devised to get students to interpret the data and are typical of the types of research questions they might have to devise themselves if they were having to undertake the research. See Figure 3 for an example.

The guide for PhD students takes a similar format but covers additional material such as nonparametric alternatives to the t Test, Analysis of Variance, Kolmogorov-Smirnov One-sample Test, Linear Regression, Logistic Regression, Reliability Analysis and Factor Analysis. It makes use of the four data sets above and others based on data about Facebook users worldwide, internet users in Europe, internet users worldwide, and demographics worldwide.

4 EVALUATION OF THE MODULE AND STATISTICS COMPONENT

The module, as a whole, is evaluated by each cohort. It consistently receives high scores. Last year, for example, it had a Likert Score of 4.38 out of 5 on a scale from strongly disagree to strongly agree overall. Based on standard questions students, in particular, thought that:

- there was an opportunity to participate in class;
- the coursework supported the module objectives;
- the module had developed their understanding of the subject;
- the assessment requirements were clear;
- criteria for marking was clear;
- the module was well organised.
Figure 3: Example of Data set questions

There were many optional positive comments made by students about the module. Several mentioned the fact that they thought it provided a very useful grounding for the dissertation such as: “Very helpful preparation for the dissertation”, “Full of useful content to help me
with my dissertation”; “Insightful information about requirements/expectations for dissertation” and “Good module – encourages head start on dissertation”.

Several students specifically mentioned that they thought the tutorials were very interesting and helpful: “The tutorials and practical statistical sessions were very useful and reinforced teaching”; “They were useful in applying the knowledge learnt” and “Reading a previous dissertation was helpful to see what is expected”. Some students commented that they thought interesting examples were used throughout, including the teaching of statistics.

Overall, the students thought that module was well organised, interesting, pitched at the right level and provided a thorough knowledge of techniques and methods, for example: “The module was very thorough and in-depth and provided a wide ranging insight into many research management techniques and approaches”; “Each method was explained clearly and logically and was easy to follow – useful for judging range of options available”; and “It gave me a good understanding of analysing both qualitative and quantitative data”.

Several students said that they enjoyed the assessment - the writing of a research proposal - while others said that the module has really motivated them to undertake their dissertation research. What was really gratifying was that students specifically mentioned the statistics component. One student said “I have always hated numbers, I now know why they are important and I am much more confident about statistics” while another said “Really, really useful, I know loads more about dissertations and statistics, in particular. Stats was bête noir in GCSE maths but now I actually understand them and their applications”. Several students mentioned that they thought the SPSS component, in particular, was informative and fun.

There were only a few negative comments mostly related to other student behaviour: “Do not let students who are more than 5 or 10 minutes late into the lecture room” and “More discouragement [needed] of mobile phone use during lectures”. Students are asked to switch off mobiles but they still hide them under tables!

5 CONCLUSION

Several approaches have been used to reduce students’ negative attitudes towards statistics and improve their understanding and use in the Department of Information Science. These concur with recommended practice as described in the literature and include:

- Stressing the learning objectives of the module including the statistics component.
- Being attentive towards students’ underlying feelings and attitudes towards statistics.
- Stressing the usefulness of statistics for their dissertations (both in applying statistics for analysing their own data and for critiquing and understanding the use of statistics by other researchers), for use in their future careers and in their lives generally.
- Developing statistical skills relevant to their field and stage of study.
- Fostering active learning by giving students lots of real, meaningful, relevant, practical examples and case studies illustrating the use of various methods and analyses.
- Providing a non-threatening environment by encouraging students to ask questions if they do not understand points.
- Clarifying statistical concepts using plain language.
- Stressing conceptual understanding of statistics before the use of SPSS.
• Reinforcing theory with tutorials, for example, getting students to calculate statistics manually before using SPSS.
• Providing immediate feedback in tutorials and when using the SPSS guide.
• Giving students relevant, useful real data sets to work on.
• Varying the approaches used: lecture, case studies, role playing, class discussions, repetition and reinforcement, small group problem solving, and step-by-step self-paced guides.
• Encouraging peer mentoring when choosing dissertation topics, methods and analyses.
• Providing self-paced SPSS guides for the technological based taught component.
• Requiring appropriate assessment, a research proposal with a requirement to discuss methods and intended data analysis.
• Informing students where they can get further help with statistics, if required.

While the statistics component is short it does cover the basics as needed by Masters’ students. Any student wishing to utilise more complicated statistics in their dissertation do have the option of using the more comprehensive PhD-level SPSS guide.

As shown above, the feedback showed that students found the statistics component useful, relevant, interesting and fun, praise indeed when many of the students come from an arts background with little previous knowledge of this topic. However, more evaluation of the statistics component is needed. It would be interesting, for example, to conduct a before and after SAT36 survey (Schau and Emmiolgu, 2012) to determine shifts, in any, in attitude towards statistics. Another measure of their success is that the SPSS guides have been adapted for use in four other Departments within the University. The Mathematics Education Department at the University also recommend the guide to students from any discipline who want to know how to use SPSS. Further, the guides have been adopted in other universities.

ACKNOWLEDGMENTS
I would like to thank the Higher Education Academy (Subject Centre for Information and Computer Sciences) for their contribution towards the cost of producing the SPSS guides and to David Green who worked tirelessly with me to write the guides.

REFERENCES


