THE PREVALENCE OF STATISTICS AND DATA MINING IN MANAGEMENT JOURNALS

Shirley Y. Coleman[1], Alex Douglas[2] and Jiju Antony[3]
[1]Industrial Statistics Research Unit, Newcastle University, UK
[2]Liverpool John Moores University, UK
[3]Heriot-Watt University, UK
shirley.coleman@newcastle.ac.uk

Postgraduate students tackling Business and Management Master's degrees are told they will need to understand statistical terms when reading research journals. Some of the journals their lecturers recommend are philosophical or conceptual in nature but others are experimental and evidence-based. Many of the articles are purely qualitative and those that include quantitative analysis are either very basic or are rather sophisticated; it could be interpreted that there is a divide rather than a continuum between authors who are confident in statistics and those who avoid them. This paper reviews relevant journals and examines the prevalence of statistics and suggests some explanations for the findings. The quantitative methods lecturer can use this information when deciding what to teach and how to help students appreciate the importance of statistical analysis.

INTRODUCTION

Students increasingly need to be competent in understanding figures and the insights claimed from analysis of all types of data. Data science is a growing profession (a Google search for "Data scientist" returned 280 million hits in February 2014 and a trend graph showed fast increase in the use of the term since 2011) and the interest in statistical analysis is widespread. Yet in the UK at least it is possible for highly educated business students not to have studied mathematics or statistics since they were 16. This is a travesty addressed in a study chaired by Vorderman (2011). Their report noted that:

> *"As our society becomes more complex, so the level of mathematical knowledge every citizen requires has deepened. Critical personal finance decisions, from choosing a mortgage to thinking about pensions, depend on individuals being secure in the basics of numeracy. Understanding the decisions governments make; on debt, deficits and the design of taxes is impossible without a grounding in mathematics."*

Hopefully the situation will be rectified in the near future. Visionaries such as Hal Varian of Google and Nate Silver (Wijlaars, 2012) have extolled the virtues of statistics and the message is gradually filtering through to all walks of life. Newcastle University Business School realises the importance of briefing students in quantitative methods and has instigated a module to address the quantitative methods that Master's students are likely to encounter as they complete the second half of their studies in the Netherlands and carry out their research and dissertations. The question arises as to what statistics they need. This paper looks at the business and management research journals that postgraduate students studying Advanced International Business Management are likely to read and reviews an example for the statistical content. The next section considers the relevant journals, followed by a section which gathers the relevant statistical terms. The results from a review of the TQM journal in 2013 are subsequently presented. It is apparent from this cursory study that not much statistical analysis is published. Near the end of the paper, the contrast with medical journals is considered with a commentary on the lost opportunities for statistical analysis. The paper closes by addressing the question of what statistics should be taught, gives conclusions and offers an agenda for further research.

RELEVANT JOURNALS

Students studying the Advanced International Business Management Master's program at Newcastle University attend modules delivered by academic staff who recommend a wide range of publications including some written by themselves. A list of recommended journals is given in Table 1.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Table 1 List of recommended journals

Academy of Marketing Science	Journal of Management Studies
Accounting, auditing & accountability journal	Journal of Marketing Management
Administrative science quarterly	Journal of Operations Management
American journal of sociology	Journal of Quality Technology
American political science review	Journal of Radical Statistics
British Journal of Management	Journal of Relationship Marketing
Business history	Journal of Retailing
Business Publications	Journal of World Business
California Management Review	Leadership
Critical Discourse Studies	Lean Management Journal
Employee Relations	Long range planning
Entrepreneurship and Regional Development	Management Learning
Entrepreneurship theory and practice	Managing Services Journal
European management journal	Organization
Gender, Work & Organization	Organization Studies
Harvard business review	Perspectives on Global Development and Technology
Human Relations	Psychology & Marketing
Industry and innovation	Public Administration Review
International Journal of Human Resource Management	Sloan management review
International Journal of Service Sciences	Sociology
International Marketing Review	Strategic management journal
International Small Business Journal	Strategic organization
International Studies of Management & Organization	The Academy of Management executive
Journal of business venturing	The Academy of Management review
Journal of general management	The Journal of business strategy
Journal of International Development	The TQM Journal
Journal of Management	World Development

It can be seen that students are recommended to read a diverse range of journals. The rankings for the journals vary from the higher ranked Journal of Operations Management with 6.01 citations per document to the lower ranked Industry and Innovation with 1.00 citations per document (SCImago, 2007).

REVIEW OF CONTENTS

The TQM journal was selected for an initial review of articles. The TQM journal aims to publish research articles and case studies about the theory and practice of quality management. It was set up in 1988 and publishes 6 issues per year often with guest editors. It currently has 1.15 citations per document and its influence has been increasing over years. The 6 issues in 2013 are reviewed as quality management covers a wide spectrum of research and the TQM journal offers a wide range of short to medium length articles of the type that may be encountered by business and management students. In 2013, the TQM journal included one issue dedicated to quality improvement in East Africa, one issue with selected papers from the 4th Canadian Quality Congress and one special issue for the TQM journal's 25th anniversary. The articles can have different style, content and focus.

Initial categorisation of the content was carried out from the point of view of the quantitative methods lecturer. It is important when preparing a course to know what kind of numerical, graphical and statistical content students are likely to encounter and the prevalence of this type of content. Articles were analysed for their use of numbers, tables and %; use of statistical tests based on t, F, chi-square, correlation, regression and corresponding p values and confidence intervals; use of surveys and Cronbach's alpha reliability; and use of more sophisticated statistical analysis like factor analysis, structural equation modelling, partial least squares, design of experiments and Taguchi; data mining.

RESULTS

The 6 issues of the TQM journal in 2013 contain 43 articles. A review of their content showed that around half of the articles included some kind of numerical, statistical or data mining indications as shown in Table 2.

Table 2 Overview of journal content

Article includes:	Numbers, tables, %	Graphics	Statistics and data mining
Number of articles (out of 43)	22	18	20

The most common type of graphics was time or line plots, followed by bar charts and then scatterplots as shown in Table 3.

Table 3 Overview of graphics used

Article includes:	Bar charts	Time or line plots	Scatter plots
Number of articles (out of 43)	6	9	4

Nine of the articles included surveys and 4 reported Cronbach's alpha reliability statistics. Regarding the articles that included statistics and data mining, six included basic summary statistics and 16 included hypothesis testing and linear modelling. Six of the articles included more complex multivariate analysis as shown in Table 4. The only data mining mentioned was text mining. Designed experiments featured in 1 article and another article reported Taguchi studies.

Table 4 Overview of statistics used

Article includes:	Means and sds	P values, t, F, correlations, regression	Factor analysis, SEM, PLS, text mining
Number of articles (out of 43)	6	16	6

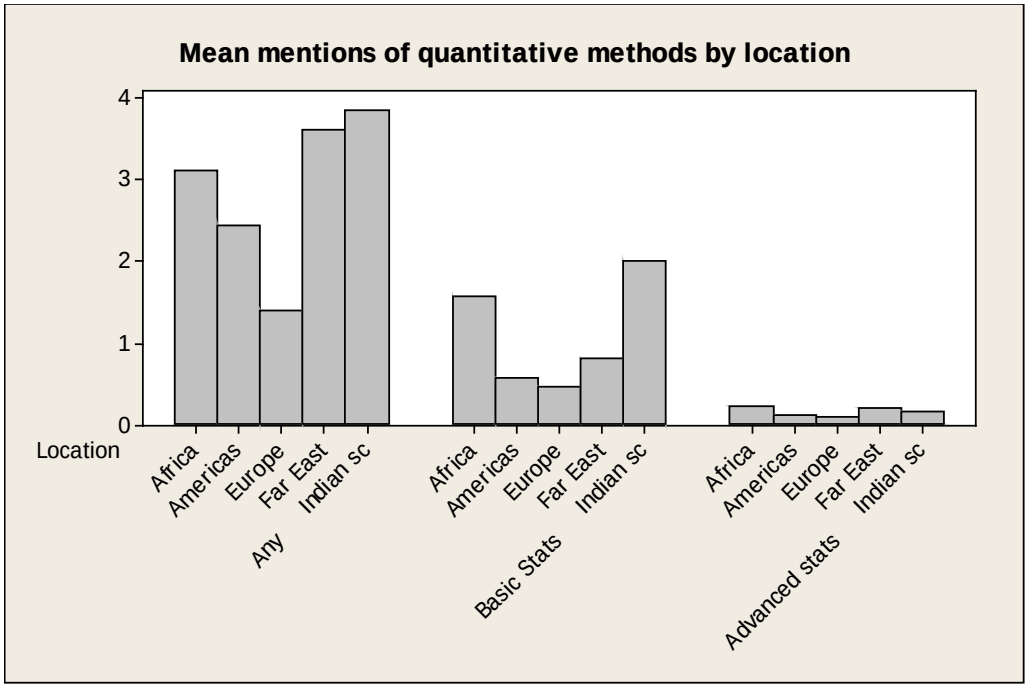

Figure 1 Mean number of times per article that quantitative methods are mentioned

The geographical location of the authors of the papers merits further investigation. The results from this initial examination of 43 articles suggest that European articles are less quantitative in nature. Figure 1 shows the mean number of mentions of basic statistics, more advanced statistics and of any numerical, graphical or statistical quantitative methods per article for authors in different locations. Note that "India sc" represents the Indian sub-continent.

It can be seen that Europe lags behind the other locations. However, note that 43 articles from one journal is just a tiny sample; the analysis is included because it indicates that further investigation may be fruitful.

COMMENTARY

Some of the statistics reported in the TQM journal could be improved upon, for example in one article chi-square analysis is used to assess the association between opinion scores and satisfaction, however, both variables are on 5 part Likert scales and many of the cells are likely to have very small expected frequency. The chi-square values have 16 degrees of freedom but even though the sample size is large at over 300, it is unlikely that all 25 cells have any values in them. In other articles, the arguments could have been strengthened by further statistical analysis, for example in one article, summary statistics are given for a survey but there is no comparison of values across demographic groups.

Graphical presentation is under-used. In several articles, the results could have been presented more clearly (and impressively) if graphics had been employed, for example when large numbers of means and standard deviations are tabulated, a bar chart would give a good visual summary. The excellent and versatile graphical displays common in websites such as SCImago (2007) are not yet replicated in scientific journals.

Many of the surveys and data collection exercises result in copious quantities of data; it could be helpful for such studies to be analysed using data mining methods. For example, decision tree analysis will quickly identify important variables, some of which were not thought to be strongly related to the study. In Figure 2, the satisfaction scores from 105 employees were collected together with their training record, time in current job role, age, gender and other characteristics.

Decision tree analysis with satisfaction as the target identifies training as the variable that best separates employees with higher satisfaction from those with lower satisfaction; the time in current job role is identified as the variable which then further separates out those with less training and shows that in this dataset higher satisfaction arises with those who have been longer in the job role. Each node in the decision tree shows the summary statistics for the sub-group as well as a bar chart. The bar charts show that the data are positively skewed with some employees being much more satisfied than others.

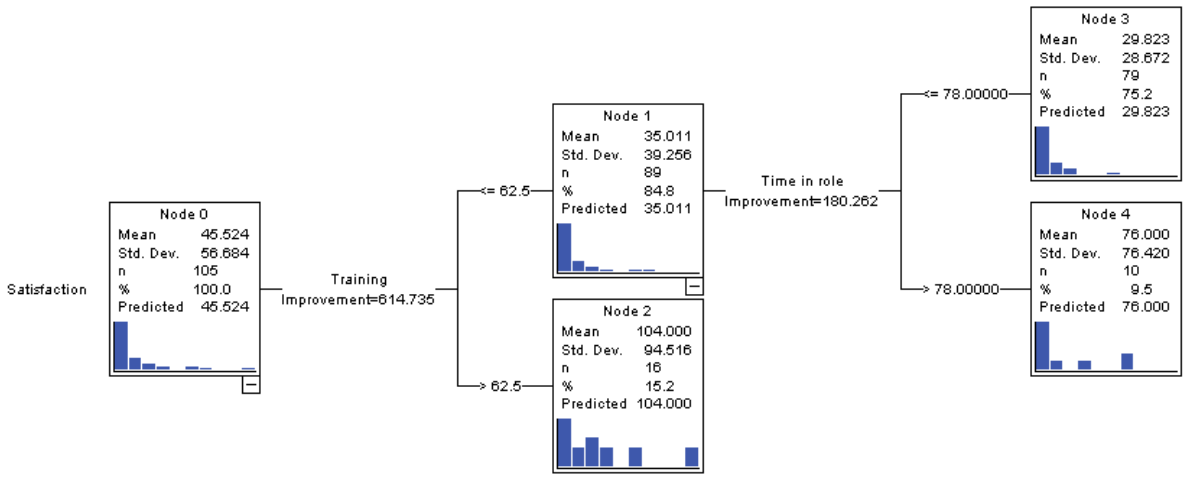

Figure 2 Decision tree analysis of survey data

The decision tree is accompanied by a gain table showing which nodes have the higher satisfaction. In Figure 2 it can be seen that n=16 employees have the higher mean satisfaction score of 104 and these employees have had more training. Many statistical software packages offer decision tree analysis; Figure 2 is from SPSS but the free software "R" also has algorithms that can be used (Kenett & Salini, 2011).

Another type of analysis that does not often appear in management journals is Bayesian analysis. This is a very natural way of analysing data that builds on evidence to propose a model of reality in contrast to the rather convoluted traditional statistical approach using null hypotheses and p values (for a typically animated discussion of Bayes vs traditional statistical analysis, see for example, http://www.stats.org.uk/statistical-inference). Bayesian analysis has the overhead of requiring statistical software to carry out the analysis rather than the ease of hand calculation offered by t tests and correlation. However, "R" provides relevant algorithms and any situation where modelling is appropriate can be analysed in a Bayesian way. Similarly Bayesian Belief Networks can complement structural equation modelling, partial least squares and canonical correlation in a more flexible and user friendly way (Kenett & Salini, 2011).

A revolution in statistical analysis in medical journals took place around 1980 and most medical journals are now very strict about including numerical evidence to support research with correct statistical analyses. Checklists are available, although interestingly these also have their problems and have been blamed for wrongful discrimination of articles (da Costa, Cevallos, Altman, Rutjes & Egger, 2011).

DISCUSSION

The quantitative methods needed to read the articles in the TQM journal, taken as an example of a typical business and management research journal, fall into three groups; the very simple numerical summaries using tables and % with basic graphics; hypothesis testing with p values and confidence intervals and the rather more complex multivariate analysis.

The extent of quantitative methods knowledge in the UK is unsatisfactorily low. It has been shown that the numeracy of UK students lags behind that of other European countries and the US. Indeed the motivation for this research was the requirement for a course to equip postgraduate students for the higher numeracy expected when they complete the second half of their studies in the Netherlands. A further example of the lower numeracy in the UK is shown by the type of articles published by UK professors; for example many UK professors of operations management, supply chain, inventory and quality management prefer to publish qualitative analysis with only about 10% publishing quantitative analysis in contrast to about 35% of US professors. A better balance between qualitative and quantitative analysis is evident in management science.

Clearly more journals need to be reviewed and analysed before any sound conclusions can be made. The TQM journal contains very basic statistics with few examples of factor analysis, structural equation modelling and multivariate analysis. One possible explanation is that researchers who have carried out more sophisticated statistical analyses tend to publish in higher ranking journals, like the Journal of Quality Technology. Another explanation is that some journals find it difficult to find reviewers who are competent in statistics to review heavily statistical papers and so the turn-around time for getting an article published is unacceptably long which necessarily encourages authors to keep the analysis simpler.

The review of the TQM journal in 2013 shows that the quantitative methods lecturer needs to teach basic numeracy, tables and % as well as descriptive statistics and graphical presentations. Traditional t-tests, F, chi-square, correlation and regression also feature regularly and need to be covered. Deeper statistical thinking resulting in design of experiments and multivariate analysis are less likely to be encountered. Data mining and Bayesian analysis are unlikely to be required at the moment although they may start to appear as numeracy improves. The shortcomings in articles and a comparison with the prevalence of statistics in US journals and medical journals should help students appreciate the importance of statistical analysis.

The research shows that around half of the surveyed articles include quantitative methods and yet many of the articles seek to test a hypothesis or report on a survey. Possible reasons include difficulty of publishing statistical articles, lack of pull by readers, lack of push by editors and lack of confidence by authors. The next step is to explore whether research progress is impaired by a lack of

quantitative methods and whether deeper statistical reporting would lead to faster development of ideas, less repetition of similar research and a more scientific approach to building knowledge.

REFERENCES

da Costa, B. R., Cevallos, M., Altman, D. G., Rutjes, A. W. S., & Egger, M. (2011). Uses and misuses of the STROBE statement: Bibliographic study. *BMJ Open 2011, 1*(1), e000048. doi:10.1136/bmjopen-2010-000048

Kenett, R.S., & Salini, S. (2011). *Modern analysis of customer satisfaction surveys: With applications using R.* Chichester, UK: John Wiley and Sons.

SCImago (2007). *SJR — SCImago Journal & Country Rank*. Retrieved 4th February, 2014 from http://www.scimagojr.com/journalrank.php?area=1400

Vorderman, C. (2011). *A world-class mathematics education for all our young people*. Retrieved 4th February, 2014 from http://www.tsm-resources.com/pdf/VordermanMathsReport.pdf

Wijlaars, L (2012). Is Nate Silver a witch? *Significance Magazine* (webexclusive), Nov 12, 2012. Retrieved 4th February, 2014 from http://www.significancemagazine.org/details/webexclusive/3612501/Is-Nate-Silver-a-witch.html