The pervasiveness and implications of statistical misconceptions among academics with a special interest in business research methods

Frank Bezzina[1] **and Mark NK Saunders**[2]
[1] Department of Management, Faculty of Economics, Management and Accountancy, University of Malta, Msida *frank.bezzina@um.edu.mt*
[2]Surrey Business School, Faculty of Business, Economics and Law, University of Surrey, Guildford, UK mark.saunders@surrey.ac.uk

Abstract: Statistics play a very important role in business research, particularly in studies that choose to use quantitative or mixed methods. Alongside statistical analysis, aspects related to research design (such as sampling, reliability and validity issues) require a good grounding in statistical concepts reinforced by careful practice to avoid potential mistakes arising from statistical misconceptions. Although quite a considerable number of published studies have focused on students' faulty thinking regarding statistical concepts, little research explores the extent to which these are also held by academics who are their instructors. This empirical study addresses this by answering the following questions: First, are statistical misconceptions pervasive among academics with a special interest in business research methods? If so, second, is there an association between the pervasiveness of statistical misconceptions and the preferred research tradition (qualitative, quantitative, mixed methods)?

Data were collected via a web questionnaire from a purposive sample of academics with an expressed interest in business research methods. The questionnaire comprised 30 categorical statements (agree, disagree, don't know) focusing on statistical misconceptions (and conceptions) relating to descriptive statistics, design strategies, inferential statistics and regression, and five demographic questions. We targeted a critical case purposive sample of 679 potential respondents. Although 166 consented to take part, only 80 completed the questionnaire and their responses form the basis of the statistical analysis, a response rate of 11.8 %. The study provides empirical evidence of both an absence of knowledge and a high pervasiveness of faulty notions that have infected the thinking of academics relating to both research design and the use of statistics. This is particularly so for academics who prefer quantitative methods, those preferring qualitative methods being more likely to admit that they do not know. The study argues that such lack of knowledge and misconceptions reduce the true utility of statistics in research. Recommendations are offered regarding the teaching of statistics within business research methods.

Keywords: research methods, misconceptions, conceptions, statistics, academics, research practice

This is a pre-publication version of
Bezzina F and Saunders M. "The Pervasiveness and Implications of Statistical Misconceptions Among Academics with a Special Interest in Business Research Methods." *The Electronic Journal of Business Research Methods Volume 12 Issue 2 2014 (pp 29-42),* available online at www.ejbrm.com

Introduction

Statistical misconceptions are argued to hinder meaningful learning, impede research progress and interfere with decision making (Huck, 2009). For students, such misconceptions may be generated by poor understanding reinforced by statements uttered or written by one's mentors (Huck, 2009). The study seeks to determine whether academics with a special interest in business research methods hold mainstream statistical misconceptions, thereby extending a recent study that investigated the prevalence of research methods mis/conceptions with the same target group (Bezzina & Saunders, 2013). To date, limited research has examined the pervasiveness of statistical misconceptions among academics; the studies we reviewed focused on either identifying students' statistical misconceptions (e.g., Bezzina, 2004; Huck, Cross & Clark, 1986; Mevareck, 1983) or statistical flaws made by authors in published articles, reports and textbooks (e.g. Huck, 2009; Lance, 2011; von Hippel, 2005). Consequently, this research enables academics to determine whether faulty thinking has infected academics' notions about mainstream statistical concepts and considers the impact of these on their students. In addition, in the light of the findings that emerge, this paper provides some important suggestions for the teaching of business research methods, particularly on what the state of practice should be.

The nature of misconceptions and the role of academics

Misconceptions are "views or opinions that are incorrect due to faulty thinking or misunderstanding" (Bezzina & Saunders, 2013, p. 41)", representing deviations from widely accepted norms and conventions. In some cases, the practices themselves are not intrinsically faulty but rather, it is the reasoning why or rationalisation used to justify the practices that is questionable (Lance & Vandenberg, 2009).

Misconceptions arise from prior learning or from interacting with the social/physical world and interfere with learning concepts (Smith, diSessa & Roschelle, 1993). Some are grounded in human intuition that leads to faulty thinking, while others are generated by inconsistencies in textbooks and oral presentations in classrooms (Huck, 2009). Garfield (1995, p.32) highlights that misconceptions are often so strong and resilient that "they are slow to change even when students are confronted with evidence that their beliefs are incorrect". Similarly, Mevareck (1983) argues that when statistical misconceptions become deeply engrained in the underlying knowledge base of the individual, mere exposure to more advanced courses is not sufficient to overcome them. However, Brown and Clement (1989) note that successful instructional confrontation can replace faulty misconceptions with new expert knowledge in a short period of time while Smith, diSessa and Roschelle (1993) advise that the emphasis should be on knowledge refinement and reorganisation rather than replacement. Given that faulty thinking is such a pervasive phenomenon, it is important that academics as instructors are aware of their own misconceptions and the impact of these upon their students (Bezzina & Saunders, 2013).

Statistical data analysis and statistical misconceptions

Statistical data analysis is the process by which data are transformed with the aim of extracting useful information and facilitating conclusions. Each statistical technique has underlying conceptual and statistical assumptions that must be met if the results are to be valid (Gel, Miao & Gastwirth, 2005). Various structured-model building approaches and step-by-step guides are available to facilitate this process of data analysis. The scope behind them is to provide researchers with "a broader base of model development, estimation and interpretation" (Hair et al., 1998. p. 25) not a rigid set of procedures to follow. Structured approaches do not come without criticism. Conflicting viewpoints arise on various aspects such as the required sample size, the statistical model to analyse the data, and the quality of the input data.

The statistical mis/conceptions addressed in this study are grouped under following headings: descriptive statistics, design strategies, inferential statistics and regression. Sentences presented in italics represent actual statements used in the research.

Descriptive Statistics

"A crucial human skill is to be selective about the data we choose to analyse and, where possible, to summarise the information as briefly and usefully as possible" (Graham, 1994, p.64). *A concise way of summarising a data set is to use an appropriate measure of central tendency* (a value that indicates where the centre of the distribution lies) *accompanied by a measure of spread* (a statistic that determines how clustered or scattered the data are). The type of measure chosen ultimately depends on the scale of measurement being used and the shape of underlying distribution (Graham, 1994). A common reported misconception in textbooks and published research reports is that *if a set of scores forms a positively skewed distribution, the mean is greater than the median, which is greater than the mode; and similarly, if a distribution of scores is negatively skewed, the mean is less than the median which is less than the mode*.

This rule is imperfect and most commonly fails in discrete distributions where the areas to the left and right of the median are not equal (von Hippel, 2005). Applying this misconception could allow researchers to make wrong judgements on the distributional shape by assessing lack of symmetry of a distribution via measures of central tendency rather than by means of a numerical index of skewness (Huck, 2009). Another misconception related to the shape of the distribution is that *standard scores such as z-scores are normally distributed*. This incorrect generalisation occurs where researchers are unaware that no finite distribution is exactly normal (Huck, Cross & Clark, 1986), and can result in inaccuracies when z-scores are converted to percentiles (Huck, 2009).

In summarising bivariate relationships, the correlation coefficient is generally used (e.g., Pearson's *r,* Spearman's *ρ*, and correlation ratio (eta). However, checks need to be made to see if the data are appropriate (e.g. whether or not the relationship is linear and whether outliers are present). A widely reported misconception is that *a single outlier will not greatly influence the value of Pearson's r, especially when N is large*. However, a single outlier can distort the nature and strength of *r* even when *N* is large. Consequently, Huck (2009) highlights the importance of conducting a visual or statistical check to see if any outliers are present. Another misconception is that *correlation never implies causation*. Huck (2009, p. 48) provides evidence that when a correlational study involves a manipulated variable and there are no plausible threats to internal validity, then "the correlation coefficient, *r*, speaks to the issue of cause and effect". He adds (2009, p.45) that data can be analysed in different ways and still give the same results; thus the warning 'correlation ≠ cause' "functions to keep the logical and mathematical equivalence of certain statistical procedures hidden from view".

Design Strategies

Sampling

An important step in planning of a statistical investigation is sample selection. This requires careful thinking (Lenth, 2000). A small sample is likely to produce a statistic of inadequate precision and makes the statistical test insensitive due to 'low' statistical power. Although an increase in sample size leads to an increase in precision, a very large sample makes the statistical test overtly sensitive (i.e. the identification of an effect relatively easy) due to 'too much' statistical power. Thus, the researcher must strike a balance between the level of statistical error and resulting power (Hair et al., 1998).

Sampling methods are generally classified into probability methods (utilizing some form of random selection) and non-probability methods (Saunders, Lewis & Thornhill, 2012). In this study, we focus on the random aspect of probability sampling, where each element of the population has a known, but possibly non-equal chance of being included in the sample. Within probability sampling, the *sample size determines precision, the selection process determines accuracy*. The following are the questionnaire items related to sampling:

1. *A random sample is a miniature replica of the population*. Statistical representativeness is generally achieved through random sampling (Thomas, 2004). However, a representative sample does yield a miniature repica (or exact replica) of the population. This is because the characteristics of a random sample are not error-free estimates of the population necessitating the specification of confidence intervals (Huck, 2009; Krzywinski & Altman, 2013). *Every sample* (even if generated in a random fashion) *possesses sampling error, provided the population is not totally homogeneous or the sample size is equal to the population size*. Huck (2009) argues that persons holding this misconception make various inferential mistakes.
2. Similarly, *a sample of individuals drawn from a finite population deserves to be called a random sample so long as (i) everyone in the population has an equal chance of receiving an invitation to participate in the study and (ii) random replacements are found for any of the initial invitees who decline to be involved.* This statement is also a misconception because those who choose not to participate are often different from those who participate. The probability of a person

responding depends on factors such as age, level of education, interest in the topic being studied and free-time available. If replacements are made, those who are not willing or able to participate are replaced by willing and able respondents. Hence, only a subsection of the population is actually represented and "any sample-to-population inferences will be distorted" (Huck, 2009, p. 129).

3. The statement *larger populations call for larger samples sizes and hence the ratio of sample size to population needs to be considered when determining sample size* is also a misconception. By definition, the precision of a sample depends on the sampling error and the larger sample, the smaller the sampling error, the greater the precision. However, the standard error formula shows that when N is much larger than n, the ratio of n to N does not influence the standard error to great extent. The precision of sample size is more influenced by n. Those who hold this misconception would wrongly dismiss the findings of a study if they believe that the sample was too small when compared to the size of the population (Huck, 2009).
4. Finally, the statement *a large sample does not guarantee validity* is correct. The common misconception is that the size of the sample guarantees validity. However, there is more strength (lack of bias) in fewer but well-chosen numbers (van Belle, 2008).

Reliability and Validity

Errors play a key role in degrading the quality of measurements. Two key issues related to the quality of measurements are reliability and validity (Murphy & Davidshofer, 2004). Reliability refers to the extent to which measures are repeatable and consistent. Validity is the degree to which measures accurately represent what they are supposed to conceptually measure. This study addresses the following mis/conceptions related to reliability and validity:

1. The statement *statistical indices of reliability and validity document important attributes of an instrument (e.g. test or questionnaire)* is incorrect. These indices of reliability and validity document important properties about the scores obtained from the instrument for a particular sample. If a person thinks that these are attributes of the test, then "a researcher may end up selecting what seems to be a good test for his or her study when in fact the selected test produces low-quality data" (Huck, 2009, p. 68).
2. The statement *a high value of Cronbach's alpha indicates that a measuring instrument's items are highly interrelated, thus justifying the claim that the instrument is uni-dimensional in what it measures*, is also flawed. Cronbach's alpha is a measure of internal consistency. Consequently, a high value of Cronbach's alpha does not indicate that the variables used are uni-dimensional (Hair et al., 1998). Even multidimensional instruments often yield high values of Cronbach's alpha. The resultant problem from this misconception is that the total score will not be interpreted correctly. As Huck (2009, p. 78) notes "high or low scores may be attributed to one thing when they are actually the result of something else". Rather, other specific techniques such as multitrait-multimethod matrix (MTMM), factor analysis (EFA or CFA), structural equation modelling (SEM) and related statistical procedures (see Westen & Rosenthal, 2003) need to be used to determine the dimensionality.
3. *Different procedures for estimating inter-rater reliability yield approximately the same reliability coefficients and so it does not make much difference which procedure is used* is flawed because various factors can affect inter-rater reliability. These include (i) whether or not chance agreement is considered, (ii) whether or not a dichotomy is imposed on the continuum, and (iii) whether or not the raters are considered a random sample from a larger population. The implication of this misconception is that a person using a particular procedure might think that raters are in close agreement with each other when in fact this is not the case when a different and more appropriate perspective is used (Huck, 2009).
4. A common misconception is that *if Pearson's r is used to determine the predictive validity, range restriction will cause r to underestimate the strength of the relationship between the predictor and criterion variables.* However, it is possible to 'correct' for range restriction when the data are linear and homoscedastic. Data collected in real validity studies are not usually very symmetric,

and correlations that are corrected for restriction are more likely to exaggerate rather than underestimate r_N. Consequently, persons holding this misconception are likely to use a formula to "correct" for range restriction which inflates Pearson's *r*, thus making them think that predictive validity is higher than what the original coefficient suggested (Huck, 2009).

Handling missing data

Although there are misconceptions concerning the need for high response rates, Newman (2009) provides evidence that low response rates (e.g., below 20%) need not invalidate study results but systematic (non-random) non-response will generally lead to bias in parameter estimates. Since almost any research has the potential for missing data, van Belle (2008) highlights that *in a research study it is important to plan for missing data and to develop strategies to account for them* prior to the initiation of the study. Furthermore, when the reasons for missing data are not identified, it is not possible to make statistical adjustments. However, s*ensitivity analyses are* purposely *designed to explore a reasonable range of explanations in order to test the robustness of the results*. Various creative statistical approaches have been developed to deal with missing data (see Cole, 2008).

Testing of statistical assumptions

Statistical methods rely on a variety of assumptions about the nature of underlying data. When the assumptions are not met, the results are often not valid (Gel, Miao & Gastwirth., 2005). This is crucial as those who are not aware of the related assumptions for a particular test may erroneously assume results are significant. A *violation of the statistical assumptions affects the significance level of a test* as well as the power of the test (Box & Tiao, 1964).

Inferential statistics

Inferential procedures are used to derive conclusions about a population. Both estimation and hypothesis testing are concerned with a parameter θ (theta) and may be considered as two sides of a coin. In estimation, *a statistic is an estimator of the true population parameter* θ if its intention is to be close to the unknown value of θ. Optimal estimators are derived according to criteria such as unbiasedness, equivariance and minimaxity (see Lehmann & Casella, 1998, for more details). *A confidence interval* is constructed to *give[s] an estimated range of values around the statistic that are believed to contain with a certain probability (e.g., 95%) the unknown population parameter* (Field, 2009). Hypothesis testing is a procedure that involves (i) setting up a null and alternative hypotheses, (ii) defining the test procedure including the levels of significance and power, (iii) calculating the test statistics and the p-value and (iv) making a decision on whether to retain or reject the null hypothesis. In the process, researchers are required to consider two types of statistical error. *The Type I error (α)*[alpha] *refers to the probability that one mistakenly rejects a true null hypothesis (i.e. a "false positive")*. The Type II error (β) [beta] refers to the probability that one mistakenly retains a false null hypothesis. The statistical power of a test is the probability of not making a Type II error and represents the odds that you will observe a treatment effect when it occurs. As power is increased, the chance of finding an effect if it's there increases; but this also increases the chance of making a Type 1 error. Since researchers aim for high power (e.g., 0.80) and low alpha (e.g., 0.05), and these do not add up to 1, there is an in-built tension here (Trochim, 2000). As alpha decreases, power decreases as well. So in determining power, the researcher must consider three factors simultaneously – alpha, sample size and effect size (Hair et al., 1998). The p-value is the probability of obtaining sample data that deviates as much or even more than the actual data observed, provided the null hypothesis is true (Huck, 2009). When the p-value is less than or equal to the probability of a Type I error, the null hypothesis is rejected and a statistically significant finding is reported. A statistically significant finding is not necessarily practically significant. *Practical significance* or effect size is the magnitude of the effect of interest in the population and *is focused on the study's possible impact on the work of practitioners or other researchers* (Hair et al., 1998). Thus, while it is incorrect to attach adjectives blindly (e.g., strong or weak) on the basis of $p< 0.05$, Cortina and Landis (2009) argue that it is even incorrect to attach adjectives blindly on the basis of Cohen's *d*. They (2009, p. 306) add that one is likely to choose the appropriate language for effect sizes when "one takes into account sample size…considers the measures involved, the nature of manipulation and the nature of the phenomenon in question".

The following is the list of misconceptions related to inferential statistics used in our study:

1. *The p-value is the probability that the null hypothesis H_0 is true* is clearly a misconception. The p-value is a random variable that varies from sample to sample and it is not the same as alpha (Good & Hardin, 2009). It is a conditional probability and hence should not be confused with alpha (Hubbard, 2004). According to Huck (2009), such faulty thinking produces errors when a null hypothesis is evaluated and in making everyday decisions based on probabilities.
2. The statement *when the whole population is used, no inferential statistics are required since the statistical summary of the data represents a parameter rather than a statistic* is flawed because inferential statistics do not require the population to have finite size. According to Fisher (1922), the goal of inferential statistics is to construct "a hypothetical infinite population" and the actual data collected is regarded as constituting a random sample (see Hacking (1979) and Seidenfeld (1979) for some interesting discussions). Thus, the true population of interest extends from the present into the future or into geographical areas not represented in the study and so persons holding this misconception wrongly assume that when data is collected from all N members of a population, a statistical summary of the data (e.g., a measure of central tendency or a percentage) produces a parameter not a statistic (Huck, 2009).
3. Similarly the statement *statistically significant results signify strong relationships between variables or big differences between comparison groups* is wrong. Effect size is concerned with the actual magnitude of the effect of interest and not statistical significance. Results which are statistically significant might not be practically meaningful while results which are not statistically significant might have a noteworthy effect size. Ellis (2010) warns that failure to distinguish between statistical and practical significance leads to Type 1 and Type 2 errors, wasted resources and potentially misleads future research on the topic.
4. Likewise, the statement *a non-directional alternative hypothesis always leads to a two-tailed test whereas a directional alternative hypothesis always brings about a one-tailed test* is faulty. With certain statistical tests, the nature of the alternative hypothesis depends on the sampling distribution of the test statistic. For instance, in the Chi-squared (χ^2) test and the ANOVA *F*-test, non-directional alternative hypotheses do not lead to two tails (split critical regions), as in the case of the *t*-test. Huck (2009) argued that persons holding this misconception will not be able to interpret properly the data-based p-value associated with the particular test used.

Regression

Regression is one of the most widely used statistical techniques and this is not limited to business and management research. It is a used to predict the likelihood or magnitude of the outcome of interest and to explore relationships and assess contributions. Various types of regression models exist (e.g., linear regression, non-linear regression, multiple regression, logistic regression, etc.). In this study, we address the following four misconceptions concerning multiple regression analysis.

1. The statement *when multiple regression is used to predict scores on a criterion variable, the worth of a particular variable is indicated by the variable's beta weight (i.e., its standardized beta coefficient)* is faulty. *In multiple regression, any statistical relationship between two variables may be altered by additional variables (*Meyers, Gamst & Guarino, 2005). When a new predictor is introduced in the model, variables can take a new level of importance within the expanded model, depending on the predictors included in the model and the degree of overlap between variables (Tolmie, Muijs & McAteer, 2011). The implication of this is that an estimated beta coefficient is not the true value of a given predictor variable. As Huck (2009) explained, double the dose of chilli powder in a recipe and the impact of the other ingredients such as onions and

beans (which previously played a prominant role) is significantly reduced. Hence it is important that when researchers interpret the beta weight, they do so relative to the specific model that produced it.

2. Likewise the statement *in multiple regression, an independent variable that is uncorrelated with the dependent variable ought to be left out of the model because its inclusion won't help to make the coefficient of determination (R^2) larger* is incorrect. Sometimes some variables which are uncorrelated with the dependent variable help to reduce the error variance in the other predictors; their inclusion better explains the variability in the criterion variable. Researchers holding this misconception are likely to eliminate such "suppressor" variables and hence they would end up with a model that falls short of its potential (Huck, 2009).
3. Finally the statement *regression analysis is superior to correlational analysis* is misconceived. This statement runs counter to the assertion that there is no univerally superior research design (Bryman, 2012) and that the research question is more important than either the method or the paradigm that underlines the method (Shavelson & Towne, 2004; Teddlie & Tashakkori, 2010). All statistical techniques have their strengths and weakness; some are simple while others are complex, but often very specific for certain purposes. Each statistical technique is a tool not an aim and hence the statistical technique chosen ultimately depends on (or is dictatated by) the research question being investigated, not vice-versa.

Research Paradigms

In the social sciences, research is very often divided into the qualitative camp and the quantitative camp. There has been an on-going debate on the distinction between the two. There are those who claimed that the distinction between the two is by no means clear (Bryman & Bell, 2011; Lincoln & Guba, 1985) while others argued that quantitative and qualitative traditions are so different in their epistemological and ontological assumptions as to be incompatible (Hammersley, 1992; Robson, 2011). According to Eby Hurst and Butts. (2009), the proponents of qualitative research make strong claims that their approach has greater ecological validity, that it provides richer and more descriptive accounts of real-world events and has a greater ability to uncover processes and mechanisms in natural settings, while the proponents of quantitative research emphasise their approach is advantageous due to strengths in the precision of measurement, experimental control and generalizability. Alongside the qualitative versus quantitative debate, there has been growing recognition of mixed-methods, which combine the qualitative and quantitative traditions (Bryman, 2006). In mixed methods, both deductive and inductive techniques may be selected and integrated to answer the research question or solve the problem be it theory testing or theory generation (Teddlie & Tashakkori, 2010).

Method

Research Questions

This study considers the following two research questions empirically:

1. Are statistical misconceptions pervasive among academics with a special interest in business research methods?
2. If so, is there an association between the pervasiveness of statistical misconceptions and the preferred research method (qualitative, quantitative, mixed methods)?

Procedure

The target population consisted of academics who are members of the Research Methodology Special Interest Groups (RM SIGs) of either the British Academy of Management (BAM) or the European Academy of Management (EURAM) (540 people), or have attended the European Conference on Research Methodology (ECRM) at least once in the past three years (139), an estimated total of 679 potential respondents after accounting for multiple list membership. A questionnaire was created using the Survey Monkey online tool. The front page provided respondents with information regarding the research, requested their consent, and assured them of anonymity. The main questionnaire consisted of 30 randomly ordered categorical statements representing statistical mis/conceptions. Respondents were requested to tick one from ‘agree’, ‘disagree’ or ‘don’t know’, the latter being included to avoid forcing the respondents to provide a response when they did not have such knowledge. The statements are presented in Table 1, the majority being adapted from Huck (2009) while the remainder were adapted from Box and Tiao (1964), Field (2009), Good and Hardin (2009), Hair et al. (2008), van Belle (2008) and von Hippel (2005). The final section requested demographic information about the respondents. Respondents were able to amend their responses until the questionnaire was submitted, while the software restricted one respondent per work station to prevent multiple completions. The e-mail with weblink targeted 679 potential respondents. 166 questionnaires were returned (24.4%), but 86 respondents although consenting to take part, reported that they ‘don’t do quants’ or the questionnaire was too ‘complex’, ‘confusing’ or ‘tricky’. This resulted in 80 complete returns (a response rate of 11.8%) that formed the basis of the statistical analysis. The questionnaire took approximately 10 minutes to complete. The preferred research method of the respondents was qualitative (47.5%), followed by mixed methods (27.5%) and quantitative (25.0%). The single largest groups were male (51.2%), those in possession of a doctoral degree (70.0%), those based in the United Kingdom (53.8%), and those involved in research methods as project or dissertation supervisors for taught Master’s degree programmes (47.5%). Since the respondents were principally academics with a documented interest in research methodology and methods, they can be considered to be a purposive sample comprising critical cases. It seems likely that if misconceptions are prevalent with this sample, other academics are also likely to hold them (Patton, 2002).

In the analysis, we generated frequency tables and computed the proportion of respondents that hold the misconception (p) together with the standard error of sample proportion ($SE(p)$). In computing p and $SE(p)$, the ‘don’t know’ responses were not considered to represent misconceptions, but highlight absence of statistical knowledge. To test the null hypothesis that the response (agree, disagree, don’t know) was independent of the preferred research method (qualitative, quantitative, mixed methods), the Chi-squared (χ^2) test was used. Due to the presence of ‘cells with expected counts less than 5’, the assumptions of the asymptotic method could not be met. So, we used the exact significance since “the exact calculation always produces a reliable result, regardless of size, distribution, sparseness, or balance of data (Mehta & Patel, 2010, p. 3). As a measure of effect size, we used Cramer’s *V*.

Findings

Table 1 provides a summary of the responses for each of the 30 items addressed in this study. It is clear that statistical misconceptions are pervasive among academics with a special interest in business research methods. In fact, the proportion of respondents that hold particular statistical misconceptions reached a 76.0% ($SE(p)$ = 0.05) for the statement ‘*Statistical indices of reliability and validity document important attributes of an instrument (e.g. test or questionnaire)*’.

Table 1: Academics’ mis/conceptions regarding statistical thinking

Survey items pertaining to statistical mis/conceptions	**A**	**D**	**DK**	**p SE(p)**
Descriptive Statistics				
A concise way of summarising a data set is to use an appropriate measure of central tendency accompanied by a measure of spread (Q11)	52	<u>7</u>	21	0.09 0.03
<u>If a set of scores forms a positively skewed distribution, the mean is greater than the median which is greater than the mode. On the other hand if a distribution of scores is negatively skewed, the mean is less than the</u>	<u>32</u>	12	36	0.40 0.05

median which is less than the mode (Q25)				
Standard scores such as z-scores are normally distributed (Q2)	35	17	28	0.44 0.06
A single outlier will not greatly influence the value of Pearson's r, especially when N is large (Q26)	33	16	31	0.41 0.05
Correlation never implies causation (Q9)	49	24	7	0.61 0.05
Design Strategies				
Sampling				
Every sample possesses sampling error provided the population is not totally homogeneous or the sample size is equal to the population size (Q18)	58	11	11	0.14 0.04
A random sample is a miniature replica of the population (Q1)	44	34	2	0.55 0.06
Larger populations call for larger samples sizes and hence the ratio of sample size to population needs to be considered when determining sample size (Q3)	41	35	4	0.51 0.06
A sample of individuals drawn from a finite population deserves to be called a random sample so long as (i) everyone in the population has an equal chance of receiving an invitation to participate in the study and (ii) random replacements are found for any of the initial invitees who decline to be involved (Q14)	56	15	9	0.70 0.05
A large sample does not guarantee validity (Q16)	74	3	3	0.04 0.02
Sample size determines precision not accuracy. The selection process determines accuracy (Q21)	52	12	16	0.15 0.04
Reliability and Validity				
A high value of Cronbach's alpha indicates that a measuring instrument's items are highly interrelated, thus justifying the claim that the instrument is uni-dimensional in what is measures (Q5)	26	15	39	0.33 0.05
Different procedures for estimating inter-rater reliability yield approximately the same reliability coefficients and so it does not make much difference which procedure is used (Q10)	8	32	40	0.10 0.03
Statistical indices of reliability and validity document important attributes of an instrument (e.g. test or questionnaire) (Q27)	61	4	15	0.76 0.05
If Pearson's r is used to determine the predictive validity, range restriction will cause r to underestimate the strength of the relationship between the predictor and criterion variables (Q29)	13	4	63	0.16 0.04
Missing Data				
It is important to plan for missing data and to develop strategies to account for them (Q19)	69	10	1	0.13 0.04
Sensitivity analyses are designed to explore a reasonable range of explanations in order to test the robustness of the results (Q30)	43	6	31	0.08 0.03
Statistical Assumptions				
A violation of the statistical assumptions affects the significance level of a test (Q4)	42	18	20	0.23 0.05
Inferential Statistics				
A statistic is an estimate of a true population parameter (Q15)	41	17	21	0.21 0.05
A confidence interval is a statement about the unknown population	40	8	32	0.10

parameter (Q17)				0.03
A Type I error (α) represents the probability that one mistakenly rejects a true null hypothesis (i.e. "a false positive") (Q28)	37	7	36	0.09 0.03
Practical significance is focused on the study's possible impact on the work of practitioners or other researchers (Q7)	47	13	20	0.16 0.04
The p-value is the probability that the null hypothesis H_0 is true (Q6)	32	27	21	0.40 0.05
When the whole population is used, no inferential statistics are required since the statistical summary of the data represents a parameter rather than a statistic (Q12)	33	22	25	0.41 0.05
Statistically significant results signify strong relationships between variables or big differences between comparison groups (Q13)	42	28	10	0.53 0.06
A non-directional alternative hypothesis always leads to a two-tailed test whereas a directional alternative hypothesis always brings about a one-tailed test (Q24)	28	13	39	0.35 0.05
Regression				
In multiple regression, any statistical relationship between two variables may be altered by additional variables (Q20)	55	15	10	0.19 0.04
When multiple regression is used to predict scores on a criterion variable, the worth of a particular variable is indicated by the variable's beta weight (i.e., its standardized beta coefficient) (Q23)	27	6	47	0.34 0.05
In multiple regression, an independent variable that is uncorrelated with the dependent variable ought to be left out of the model because its inclusion won't help to make the coefficient of determination (R^2) larger (Q22)	18	19	43	0.23 0.05
Regression analysis is superior to correlational analysis (Q8)	21	48	11	0.26 0.05

Note: A = Agree, D = Disagree, DK = Don't Know, p = pervasiveness of misconception as %, SE (p) = standard error of sample proportion; underlined scores represent faulty thinking; underlined statements represent misconceptions.

In investigating whether the responses varied as a function of the preferred research method of respondents, we found that a significant association occurred in only 15 out of the 30 items presented. Three of these statements represented statistical conceptions, namely '*a concise way of summarising a quantitative data set is to use an appropriate measure of central tendency together with a measure of dispersion (spread)*' [$\chi^2(4) = 9.97$, p = 0.037, V = 0.25], '*the confidence interval is a statement about the unknown parameter*' [$\chi^2(4) = 11.11$, p = 0.023, V = 0.26] and '*a Type 1 error represents the probability that a a true null hypothesis is rejected (i.e. "a false positive")*' [$\chi^2(4) = 25.58$, p < 0.001, V = 0.40] . As one would expect, the respondents who prefer quantitative research were the most knowledgeable about these statistical conceptions, followed by those who prefer mixed methods and qualitative methods respectively. The other 11 statements represented statistical misconceptions but here different patterns emerged:

a) For the statement '*different procedures for estimating inter-rater reliability yield approximately the same reliability* coefficients. *Therefore it does not make much difference which procedure is used*' [$\chi^2(4) = 15.62$, p = 0.003, V = 0.31], respondents who prefer quantitative research were more likely to disagree with this faulty statement; those who prefer qualitative and mixed methods were more likely to admit that they don't know.
b) For the two statements *'larger populations call for larger sample sizes and hence the ratio of sample size to population needs to be considered when determining sample size'* [$\chi^2(4) = 9.33$, exact sig. = 0.044, V = 0.24] and '*statistically significant results signify strong relationships between variables or big differences between comparison groups*' [$\chi^2(4) = 12.79$, p = 0.011, V =

0.28], those who prefer qualitative and mixed methods were more likely to hold these misconceptions while quantitative researchers were more likely to disagree.

c) For the statement 'Statistical indices of reliability and validity document important attributes of an instrument (e.g. test or questionnaire' [$\chi^2(4) = 10.79$, $p = 0.024$, $V = 0.26$], those who prefer quantitative research were the most likely to hold this misconception, followed by those who prefer mixed methods and qualitative research respectively.

d) For the remaining eight statements - '*standard scores such as z-scores are normally distributed*' [$\chi^2(4) = 20.81$, $p < 0.001$, $V = 0.36$], '*a high value of Cronbach's alpha indicates that a measuring instrument's items are highly interrelated, thus justifying the claim that the instrument is uni-dimensional in what is measures*' [$\chi^2(4) = 20.95$, $p < 0.001$, $V = 0.36$], '*the p-value is the probability that the null hypothesis H_0 is true* [$\chi^2(4) = 18.97$, $p = 0.001$, $V = 0.34$], '*in multiple regression, an independent variable that is uncorrelated with the dependent variable ought to be left out of the model because its inclusion won't help to make the correlation of determination (R^2) larger*' [$\chi^2(4) = 25.69$, $p < 0.001$, $V = 0.40$], '*when multiple regression is used to predict scores on a criterion variable, the worth of a particular predictor is indicated by the variable's estimated beta weight (i.e. its standardized regression coefficient*' [$\chi^2(4) = 21.82$, $p < 0.001$, $V = 0.37$], '*a non-directional alternative hypothesis always leads to a two-tailed test whereas a directional alternative hypothesis always brings about a one-tailed test'* [$\chi^2(4) = 17.46$, $p = 0.001$, $V = 0.33$], '*if a set of scores forms a positively skewed distribution, the mean is greater than the median which is greater than the mode…*' [$\chi^2(4) = 27.76$, $p < 0.001$, $V = 0.42$] and '*a single outlier will not greatly influence the value of Pearson's r, especially when N is large*' [$\chi^2(4) = 18.71$, $p = 0.001$, $V = 0.34$] - the respondents who were most likely to hold the misconceptions were those who prefer quantitative research methods, followed by those who prefer mixed methods and qualitative methods respectively. This unexpected result might be explained by the fact those who prefer qualitative research methods were more likely to admit they 'don't' know', with those who prefer mixed methods doing so at a lesser extent.

Discussion

The findings of this study suggest both a lack of knowledge and a high pervasiveness of statistical misconceptions among academics with a special interest in business research methods. However, we do not want to convey the message that the misconceptions we have reported here are pandemic to the field of business and management as that would, unfairly, discredit the work of competent researchers.

When academics cannot separate fact from fiction regarding mainstream statistical concepts, it could be hindering them in making appropriate methodological choices (Good & Hardin, 2009; Huck, 2009; Lance, 2011), not to mention the impact on their student's research efforts (Bezzina & Saunders, 2013). Additionally, academics who are not so conversant with statistical concepts (evidenced by those who opted for "don't know" or withdrew from the survey) might prefer to take a qualitative stance in their research study rather than incorporate in it statistical thinking. The consequence could be that rather than answering the question that they think is the important question, the research question fits the convenient design (Shavelson & Towne, 2004). We believe that this issue warrants attention in the teaching of business research methods.

A second major finding in this study is that in half of statements addressed, the pervasiveness of the misconceptions was not associated with the preferred research method. However, where a significant association was found, in most cases quantitative researchers were more likely to endorse the misconception. This could be attributed to the fact that a number of 'statistical rules of thumb' endorsed by these academics are flawed and, unlike qualitative researchers, they are less likely to be aware of their own lack of knowledge; represented by 'don't know' responses. We hope that this research will help such academics to identify misconceptions and to understand the impact of these on their students. Today, various books and interactive Internet activities are available to help those interested to 'undo' misconceptions, although the strategies suggested for addressing such statistical and methodological misconceptions might themselves require evaluation in future studies.

There are some limitations to our findings that should be noted. First, we used a critical case purposive sample. Consequently it could be argued that this sample is suited more to the logical than the statistical generalisations we have made. Second, the findings of this study are based on a relatively small sample, despite follow-ups to potential respondents restating the web-link and re-emphasizing the deadline. This, combined with the high withdrawal rate might have biased to some extent parameter estimates. Third, the concepts addressed in this study are not exhaustive. Fourth, although we asked respondents to highlight their preferred research method, we believe that the choice of method should be dependent upon the question being answered (Saunders, Lewis & Thornhill, 2012).

Concluding comments

The teaching of statistics is often seen as an initiation into rules and procedures which might be seen as attractive and powerful by instructors, yet meaningless by pupils (Bezzina, 2004). Easy practices tend to take the short route by by-passing the detailed study necessary to get it right (Lenth, 2000), As Good and Harding (2009, p. xi) argued, the availability of statistical software packages and high-speed computers "will no more make one a statistician than a scalpel will turn one into a neurosurgeon". Allowing these tools to do our thinking will obscure the true value of statistics when applied correctly in research.

To develop a thorough understanding of the statistical foundations requires careful practice sustained by sound rationale and justification that goes beyond simply applying rules and procedures. To enable this, statistics need to be taught by instructors who, through their expertise are fully aware of and can explain prevalent misconceptions. Where instructors are not statistical experts it is important that, as was the case in for many of our 'qualitative' respondents, as well as at least some of those who withdrew, they recognise their lack of knowledge. We believe that students are more likely to benefit in quantitative research methods classes and courses if they are given the opportunity (i) to get involved in the struggle with the statistical concepts, (ii) to get involved in dialogue and (iii) to focus on formulating reasonable solutions that are timely, accurate, flexible, economical, reliable and easy to understand and use, rather than just applying procedures.

Acknowledgement

Both authors contributed equally to this paper.

References

Bezzina, F. (2004) "Pupils' Understanding of Probability & Statistics (14-15+): Difficulties and Insights for Instruction", *Journal of Maltese Education Research*, Vol. 2, No. 1, pp. 53-67.

Bezzina, F. & Saunders, M.N.K. (2013) "The Prevalence of Research Methodology Mis/conceptions among Business and Management Academics", in, Mesquita, A. & Ramos, I. (eds) *Proceedings of the 12th European Conference on Research Methodology for Business and Management Studies*, Reading, ACPI, pp. 40-47.

Box, G.E.P. & Tiao, G.C. (1964) "A note on criterion robustness and inference robustness", *Biometrica*, Vol. 51, pp. 169-173.

Brown, D.E. & Clement, J. (1989) "Overcoming misconceptions via analogical reasoning; abstract transfer versus explanatory model construction", *Instructional Science*, Vol. 18, pp. 237-261.

Bryman, A. (1988) *Quality and Quantity in Social Research*, Unwin Hyman, London.

Bryman, A. (2012) *Social Research Methods* (4th edn) Oxford University Press, Oxford.

Bryman, A. & Bell, E. (2011) *Business Research Methods* (3rd edn), Oxford University Press, Oxford.

Cole, J. C. (2008) "How to deal with missing data", in Osborne, J.W. (ed.), *Best practices in quantitative methods*, Sage, Thousand Oaks, CA, pp. 214–238.

Cortina, J.M. & Landis, R.S. (2009) "When Small Effect Sizes Tell a Big Story, and When Large Effect Sizes Don't", in Lance C.E. & Vandenberg, R.J. (eds) *Statistical and Methodological Myths and Urban Legends*, Routledge, London, pp. 287-308.

Eby, L.T., Hurst, C.S. & Butts, M.M. (2009) "Qualitative Research: The Redheaded Stepchild in Organisational and Social Science Research?", in, Lance C.E. & Vandenberg, R.J. (eds) *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organisational and Social Sciences*. Routledge, London, pp. 219-246.

Ellis, P.D. (2010) *The Essential Guide to Effect Sizes: Statistical power, Meta-Analysis and the interpretation of research results*, Cambridge University Press, Cambridge.

Field, A. (2009) *Discovering Statistics using SPSS* (3rd edn), SAGE, London.

Fisher, R.A. (1922) "On the mathematical foundation of theoertical statistics", *Philosophical Transactions of the Royal Society of London*. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 222, pp. 309-368.

Garfield, J.B. (1995) "How students learn statistics", *International Statistics Review*, Vol. 63, pp. 25-34.

Gel, Y., Miao, W. & Gastwirth, J.L. (2005) "The importance of checking the assumptions in underlying statistical analysis: Graphing methods for assessing normality", *Jurimetrics*, Vol. 46, No. 1, pp. 3-29.

Good, P.I. & Hardin, J.W. (2009) *Common Errors in Statistics (and how to avoid them)*, Wiley, Hoboken, NJ.

Graham, A. (1994) *Statistics: An Introduction*, Hodder & Stoughton, London.

Hacking, I. (1979) *Logic of Statistical inference*, Cambridge University Press, Cambridge.

Hair, J.F., Anderson, R.E., Tatham, R.L. & Black. W.C. (1998) *Multivariate Data Analysis* (5th edn), Prentice Hall, Upper Saddle River, NJ.

Hammersley, M. (1992) "The paradigm wars: Reports from the front." *British Journal of Sociology of Education*, Vol. 13, No. 1, pp. 131-143.

Hubbard, R. (2004) "Blurring the Distinctions Between p's and a's in Psychological Research", *Theory Psychology,* Vol. 14, No. 3, pp. 295-327.

Huck, S. W. (2009) *Statistical Misconceptions*. Psychology Press, London.

Huck, S.W., Cross, T.L. & Clark, S.B. (1986) "Overcoming misconceptions about z-scores", *Teaching Statistics*, Vol. 8, No. 2, pp. 38-40.

Krzywinski, M. & Altman, N. (2013) "Points of significance: Importance of being uncertain", *Nature Methods*, Vol. 10, pp. 809-810.

Lance, C.E. (2011), "More Statistical and Methodological Myths and Urban Legends", *Organisational Research Methods*, Vol.14, No. 2, pp. 279-286.

Lance, C.E. & Vandenberg, R. J. (2009), *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organisational and Social Sciences*, Routledge, London.

Lehmann, E.L. & Casella, G. (1998) *Theory of Point Estimation* (2nd edn), Springer-Verlag, New York.

Lenth, R. V. (2000) Two Sample-Size Practices That I Don't Recommend, available at: http://www.stat.uiowa.edu/~rlenth/Power/2badHabits.pdf (accessed November 28, 2013)

Lincoln, Y. S., & Guba, E. G. (1985) *Naturalistic Inquiry*, SAGE, Beverly Hills, CA.

Mehta, C.R. & Patel, N.R. (2010) *IBM SPSS Exact Tests*, SPSS Inc., Cambridge, MA.

Mevareck, Z.R. (1983), "A deep structure model of students' statistical misconceptions", *Educational Studies in Mathematics*, Vol. 14, pp. 415-429.

Meyers, L.S., Gamst, G. & Guarino, A.J. (2005) *Applied multivariate research: Design and interpretation*, Sage, Thousand Oaks, CA

Murphy, C.O. & Davidshofer, K.R. (2004) *Psychological Testing: Principles and Applications* (6th edn), Pearson/Prentice Hall, Upper Saddle River, NJ.

Newman, D.A. (2009) "Missing Data Techniques and Low Response Rates: The Role of Systematic Nonresponse Parameters", in, Lance C.E. & Vandenberg R.J. (eds) *Statistical and Methodological Myths and Urban Legends*, Routledge, London, pp. 7-36.

Patton, M. (2002) *Qualitative Evaluation and Research Methods*, SAGE, London.

Saunders, M., Lewis, P. & Thornhill, A. (2012) *Research Methods for Business Students* (6th edn), Pearson, Harlow.

Robson, C. (2011) *Real World Research: A Resource for Users of Social Research Methods in Applied Settings* (3rd edn), John Wiley, Chichester.

Seidenfeld, T. (1979) *Philosphical Problems of Statistical Inference: Learning from R.A. Fisher*, D. Reidel Publ., Dordrecht, Holland.

Shavelson, R.J., & Towne, L. (2004) What drives scientific research in education? Questions, not methods, should drive the enterprise, *American Psychological Society Observer*, Vol. 17, no. 4, April 2004, [online], http://www.psychologicalscience.org/index.php/uncategorized/what-drives-scientific-research-in-education.html

Smith, J. P., diSessa, A. A. & Roschelle, J. (1993) "Misconceptions Reconceived: A constructivist analysis of knowledge in transition", *Journal of the Learning Sciences*, Vol. 3, No. 2, pp. 115-163.

Teddlie, C. & Tashakkori, A. (2010) "Overview of contemporary issues in mixed methods research", in Tashakkori, A. & Teddlie, C. (eds), *The Sage Handbook of Mixed Methods in Social and Behavioural Research* (2nd edn), Sage, Thousand Oaks, CA, pp. 1-41.

Thomas, A.B. (2004) *Research Skills for Management Studies*, Routledge, London.

Tolmie, A., Muijs, D & McAteer, E. (2011) *Quantitative Methods in Education and Social Research using SPSS*, Open University Press/McGraw Hill, Maidenhead, Berkshire.

Trochim, W. (2000). *The Research Methods Knowledge Base* (2nd edn), Atomic Dog Publishing, Cincinnati, OH.

van Belle, G. (2008) *Statistical Rules of Thumb* (2nd edn), Wiley, Seattle, WA.

von Hippel, P.T. (2005) "Mean, Median and Skew: Correcting a Textbook Rule", *Journal of Statistics Education,* Vol. 13, No. 2, [online], http://www.amstat.org/publications/jse/v13n2/vonhippel.html

Westen, D. & Rosenthal, R (2003) "Quantifying construct validity: Two simple measures", *Journal of Personality and Social Psychology*, Vol. 84, No. 3, pp. 608-618.