# Militating against data fabrication and falsification: A protocol of trias politica for business research

**Anthony Stacey**
**Graduate School of Business Administration, University of the Witwatersrand, Johannesburg, South Africa**
Anthony.Stacey@wits.ac.za

**Abstract:** Data fabrication and falsification are clear breaches of research ethics, but have been shown to be insidious factors in various research disciplines. It would be naïve to believe that data fabrication and falsification do not affect the validity and reliability of business research. It behoves all users of such research to militate against these unethical practices and ensure that they do not go undetected. This paper briefly reviews the motivations for researchers, interviewers or surveyors to falsify or fabricate research data. This is followed by a discussion of techniques in the literature for detecting such unethical and fraudulent practices. Typically, these rely on the premise that falsification or fabrication of data results in anomalies in the dataset that cannot be attributed to sampling or methodology.

A number of business case studies are discussed involving subtle data anomalies that could be attributable to fabrication or falsification or data. It is demonstrated that tried and tested parametric or non-parametric statistical tests are often more than sufficient to identify these anomalies that characterise bogus data. However, data fabrication and falsification are not necessarily self-evident and it may therefore require an unconventional and innovative approach to determine the appropriate variables of analysis. Analysis of the phenomenon leads to the conclusion that data fabrication and falsification are most easily detected by carrying out analyses on apparently extraneous variables, as these would tend to be neglected by the errant interviewer or surveyor. This leads to proposing a generic approach to detecting bogus data and a corresponding protocol to militate against it.

A protocol is proposed that separates the essential research functions by adopting the *trias politica* principle, or separation of powers, analogous to the three branches of government: the legislature, the executive and the judiciary. The protocol requires the three functions of research design plus substantive analysis, data collection, and data verification to be separated. Suggestions for presenting data analyses and research findings that will ensure greater transparency, militate against data fabrication and falsification, improve reliability, and promote research integrity are included. The paper concludes with a specific recommendation to academics, consultants, reviewers, examiners, and other users of business research to hold researchers more accountable for their validity and reliability of their research outputs.

**Keywords:** Data falsification; Data fabrication; Bogus research data; Research protocol; Separation of responsibilities; Trias politica.

## 1 Background

In the context of research, data fabrication is the creation of bogus data that either supplements or substitutes for genuine research data, while data falsification is the deliberate altering of research data (Herndon, 2016) and is symptomatically no different from data fabrication. Data fabrication or falsification – sometimes euphemistically referred to as data "curbstoning" (e.g. Birnbaum, Borriello, Flaxman, DeRenzi, & Karlin, 2013; Clark & Moul, 2004; Kennickell, 2015) – can occur at various stages in the research process, and can be perpetrated by interviewers, enumerators, surveyors, researchers, respondents, or any others participants in the research. There can be no debate that data fabrication and falsification fall within the definition of research misconduct or questionable research practices (Banks, Rogelberg, Woznyj, Landis, & Rupp, 2016) and there can be no justification or rationale for tolerating them in business research.

It is worth reflecting on the potential impact of data fabrication or falsification on other researchers, and research users. The most direct consequence is that the users of the research (and potentially the researchers themselves) are likely to be misled to assume validity and reliability that the research does not merit. Depending on the extent and nature of the fabrication or falsification, results may be obfuscated, exaggerated, or entirely erroneous. Strategic and operational business decisions, as well as future academic research activities will be misinformed while stakeholders and decision makers remain obliviously vulnerable.



Reference this paper: Stacey A "Militating against data fabrication and falsification: A protocol of trias politica for business research" *The Electronic Journal of Business Research Methods Volume 14 Issue 2 2016*, (pp72-82) available online at www.ejbrm.com

It is acknowledged that measures to detect and militate against data fabrication and falsification are most effective when implemented *ex ante* or during the data collection process, for example through call-backs and face-to-face re-interviewing (Bredl, Storfinger, & Menold, 2011). However, such measures are not always practicable due to timing, logistical and outsourcing arrangements. The focus of this research is on *ex post* analysis to detect bogus data, and measures to militate against fabrication and falsification in the dataset under analysis. This research will not consider data fabrication in clinical research as the protocols and consequences are typically different from that of business research.

In the remainder of this article, prior research pertaining to data fabrication are reviewed, case studies illustrating data anomalies indicative of data fabrication or falsification are presented, specific statistical tests that could be used to identify data anomalies are highlighted, and a research protocol to militate against data fabrication and falsification in business research is recommended and discussed.

## 2 Literature review

It may be useful to begin by discussing the relationship between outliers, extreme values, and fabricated data. An outlier has been defined as data "that appears to deviate markedly from other members of the sample in which it occurs" (Grubbs, 1969, p. 1). Grubbs (1969) is specific that an outlying value may either be a legitimate, extreme value that has occurred as a result of the inherent variability of the variable in which case the value must be retained, or may be an anomaly or error in which case the value may be omitted from the dataset. It is noted that fabricated or falsified data may or may not appear as outliers or extreme values, and therefore procedures for detecting outliers (e.g. Grubbs, 1969) are not appropriate for detecting fabricated or falsified data.

### 2.1 Motivation for fabricating or falsifying data

Bredl et al. (2011) refer to a number of reasons why interviewers might be inclined to falsify survey data. For example, typically interviewers are not involved in the data processing, and do not have a vested interest in the quality of the data and the research output. Interviewers may not be knowledgeable in research design, sampling protocols and research ethics (American Association for Public Opinion Research, 2003) and may not appreciate their impact on the research. Typically, interviewers' performance and remuneration are evaluated not in terms of data quality, but rather by the number of completed interviews (Kennickell, 2002), thereby creating a perverse incentive to falsify data. Other possible reasons for fabricating data include avoiding asking sensitive questions which could provoke abusive responses, visiting risky interview locations, travelling time to locations, and lengthy interviews (Winker, Menold, Storfinger, Kemper, & Stukowski, 2013).

### 2.2 The first digit phenomenon

The counterintuitive observation that the first significant digits of numbers comprising large datasets often do not occur with equal probability is referred to as the first digit phenomenon. The phenomenon was first analysed by (Newcomb, 1881) and then independently by Benford (1938), after which it become commonly referred to as Benford's Law. Benford's Law is applicable to variables that are logarithmically distributed, and has been used to detect anomalies in data that is expected to follow logarithmic distributions (e.g. Diekmann, 2007; Judge & Schechter, 2009). The diagnostic potential of the first digit phenomenon was clearly recognised by Newcomb is his concluding observation that: "It is curious to remark that this law would enable us to decide whether a large collection of independent numerical results were composed of natural numbers or logarithms" (Newcomb, 1881, p. 40).

### 2.3 Metadata and missing data

Metadata (i.e. data about the data) can indicate potential data quality issues. Examples of survey metadata include date- and time-stamps, interview "hit rate", interviewer identifier, interview location, language, and medium of data collection – e.g. pencil and paper, computer-assisted personal interviewing (CAPI), computer-assisted telephonic interviewing (CATI), computer aided web interviewing (CAWI). Bredl et al. (2011) point out that if date and time stamps are available, then interview lengths and the completion rate of interviews can be analysed for anomalies.

Similarly, Schraepler and Wagner (2005) show that those who fabricate or falsify data tend to avoid skipping survey items on the incorrect assumption that genuine participants respond to all survey items. Therefore, analysing patterns of missing data may also reveal anomalies indicative of bogus data.

### 2.4 Dataset content

Anomalies in the dataset content are not only damaging to results of the data analysis, but may also be diagnostic of data fabrication or falsification. Schraepler and Wagner (2005) found that perpetrators can reasonably accurately replicate the central tendency of a legitimate sample of responses, but that they tend to underrepresent the variability of responses to specific survey questions.

Similarly more complex relationships within a dataset, such as correlations, are also difficult to fabricate and anomalies can then be identified by analysing interrelationships among variables using techniques such as factor analysis or regression analysis (Bredl et al., 2011). Other multivariate techniques, such as cluster analysis and discriminant analysis, have also been used by Bredl, Winker, and Kötschau (2012) to detect bogus data.

## 3 Case studies

Research into methods of *ex post* detection of data fabrication and falsification is ongoing (e.g. Simmons, Mercer, Schwarzer, & Kennedy, 2016). The following case studies, focusing on identifying anomalies in a dataset that provide *prima facie* evidence of data fabrication or falsification, are used for illustrative purposes and are by no means intended to be exhaustive or representative. The analysis itself does not constitute proof of dishonesty and further investigation would typically be required to confirm who may be responsible for such anomalies. Once fabrication or falsification is confirmed then such anomalous data would necessarily be filtered from the dataset for analysis.

### 3.1 Response distributions

A large independent study was carried out to determine perceptions of 30 organisations among the general public using a battery of 29 survey items. Participants were contacted telephonically by 20 interviewers who then coded and captured the responses using a 7 point Likert-type response format. Interviewers were randomly assigned participants, and each participant was randomly allocated up to 5 organisations to rate, provided they were sufficiently familiar with the organisations. There were some concerns regarding the face validity of the results of the initial analysis of the survey data. Therefore the survey data were subjected to further analysis.

It can be shown that a normal distribution can be used to approximate a binomial distribution, for a large number of binomial trials. The reciprocal implication of this is that if a single outcome is dependent on a large number of independent stimuli, then the probability distribution of that outcome variable will be approximately normal. Hence the normal distribution being suitable to analyse many variables including portfolio returns, employee performance, process variations, and inventory forecasts. This would also apply to attitudes and perceptions of survey participants towards the organisations in this study. Therefore, not only should the distributions of participants' responses to the 29 survey items be normal, but because the participants and organisation have been randomly assigned to the 20 interviewers, the distributions of the responses captured by each of the interviewers should also be normal.

An analysis to test the fit of normal distributions to the response data captured by each of the interviewers was carried out using the $\chi^2$ goodness of fit test (Stacey, 2005, 2012) the results of which a shown in Table 1. It is evident that in all instances, except Interviewers 12 and 17, at a 5% significance level there is no significant difference between the response data and a normal distribution, as expected. However, in the case of Interviewers 12 and 17, the data could not be fitted to a normal distribution, suggesting that the data captured by these interviewers does not correspond to the attitudes and perceptions of actual survey participants.

**Table 1:** Goodness of fit between response data and normal distributions

| Interviewer ID | Number of ratings | $\chi^2$ statistic | *p* value |
|---|---|---|---|
| **1** | 899 | 195.1 | 0.0746 |
| **2** | 470 | 142.4 | 0.9241 |
| **3** | 200 | 110.3 | 0.9998 |
| **4** | 510 | 122.5 | 0.9966 |
| **5** | 637 | 129.7 | 0.9872 |
| **6** | 1008 | 139.9 | 0.9437 |
| **7** | 215 | 101.1 | 1.0000 |
| **8** | 428 | 133.7 | 0.9761 |
| **9** | 867 | 134.6 | 0.9727 |
| **10** | 794 | 99.4 | 1.0000 |

| Interviewer ID | Number of ratings | $\chi^2$ statistic | *p* value |
|---|---|---|---|
| 11 | 847 | 107.0 | 0.9999 |
| 12 | 1156 | 266.8 | < 0.0001 |
| 13 | 99 | 112.3 | 0.9997 |
| 14 | 255 | 81.0 | 1.0000 |
| 15 | 928 | 184.0 | 0.1881 |
| 16 | 884 | 192.9 | 0.0910 |
| 17 | 175 | 626.0 | < 0.0001 |
| 18 | 116 | 124.2 | 0.9952 |
| 19 | 210 | 123.2 | 0.9961 |
| 20 | 119 | 110.8 | 0.9998 |

*Degrees of freedom = 168*

Figure 1 illustrates and compares the distributions of responses to a particular item for one interviewer whose response distributions fit normal distributions (on the left) and another whose response distributions are significantly different from normal distributions (on the right).

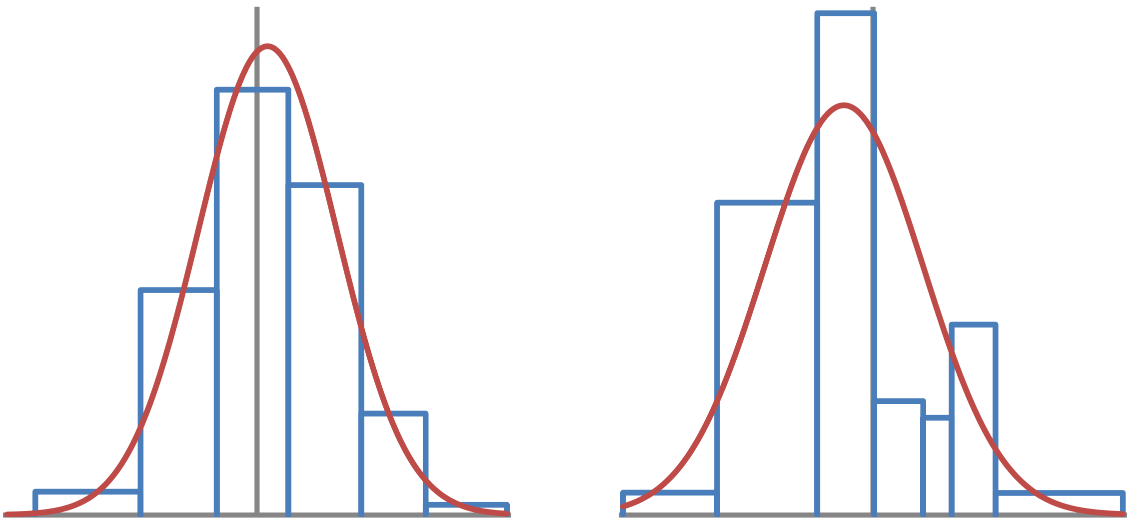

**Figure 1:** Comparison of response distributions fitting and significantly different from normal distributions

It is clear from Table 1 that only for interviewers 12 and 17 are the best fitting normal distributions significantly different from the response distributions. At a 5% significance level, there was not sufficient evidence to suggest that the response distributions for any of the remaining interviewers were anything other than normal. Because the participants and organisations have been randomly assigned to the 20 interviewers, this anomaly is evidence of possible fabrication of the data captured by interviewers 12 and 17.

## 3.2 Factor structure

In a context similar to that described in the previous example, two batteries of survey items are used to evaluate organisations' reputations (7 items) and the strength of their branding (8 items). A principal component analysis (PCA) has been carried out in order to determine the underlying structure of the dataset. Two components have been extracted on the basis of the Kaiser criterion; the eigenvalues and factor loadings have been charted in Figure 2.

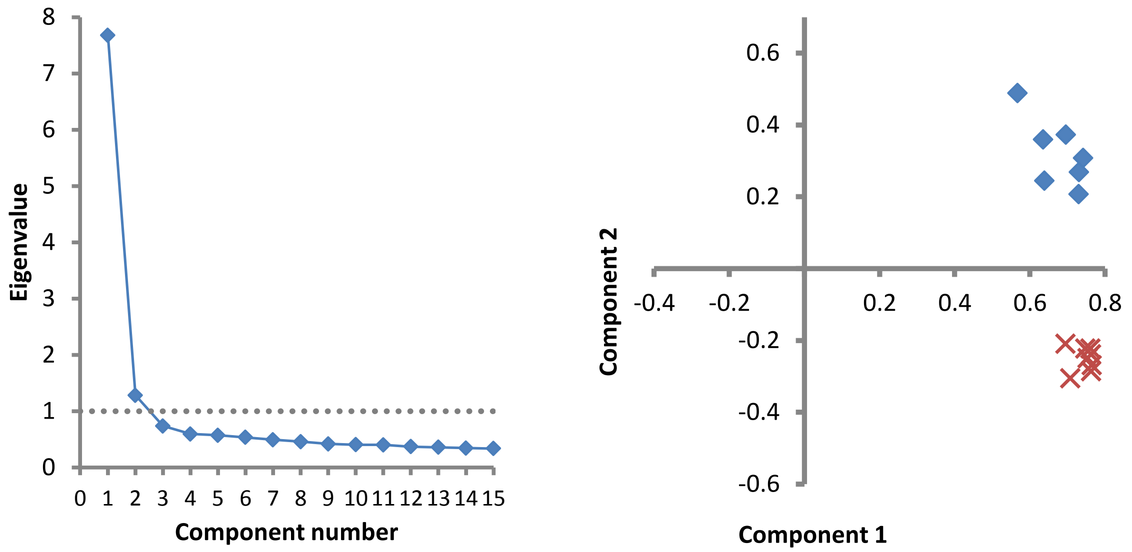


**Figure 2:** Plots of eigenvalues and principal component loadings for complete dataset

A well define underlying structure is evident in the data with the first dimension accounting for a large proportion of the overall variance. By contrast, a similar PCA has been carried out on the response data captured by one specific interviewer in order to compare the underlying structure of this sub-set of the data with that of the overall dataset. The eigenvalues and factor loadings have been charted in Figure 3. Two components have been extracted to facilitate the comparison.

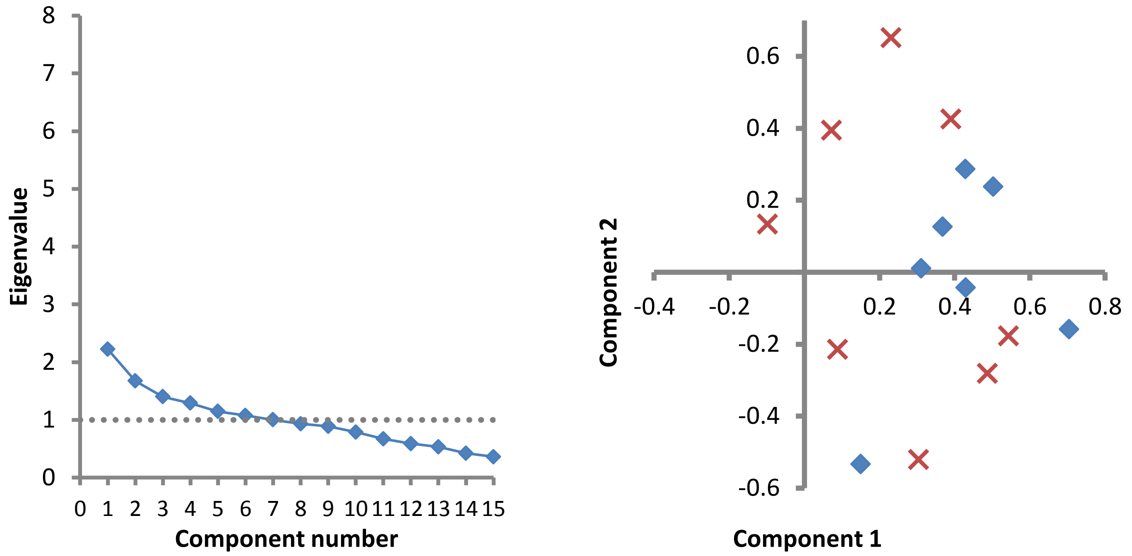


**Figure 3:** Plots of eigenvalues and principal component loadings for anomalous subset of data

Comparing Figure 2 and Figure 3, it is clear that the underlying structure of the data captured by the specific interviewer differs substantially from that of the overall dataset. Indeed, a parallel analysis (Hayton, Allen, & Scarpello, 2004; Horn, 1965) would show that the data captured by the specific interviewer differs very little from a dataset comprised of random numbers. This could be a strong indication of the specific interviewer fabricating the response data.

### 3.3 Distribution of response time intervals

The organisation in question operates a national chain of specialist retail outlets that operate daily between 08:00 and 23:00. In order to measure and track their customers' service experience, the organisation has installed electronic customer feedback devices at all of their outlets. Using these devices, customers are able to provide feedback at the point of sale or service provision by responding to a few critical questions using the customised keypad. Although there was no evidence of quality issues in the response data *per se*, the organisation's executive wanted confirmation of the validity of the results in order to monitor continuous improvement of the customers' service experience.

Analysis of the time-stamps corresponding to the responses, which were available in the raw data file from the electronic customer feedback devices, was particularly enlightening. The distribution of the responses received during the course of the operating hours of the retail outlets would show that the proportion of responses is highest during the daytime hours and declines during the evening through to when the outlet closes at 23:00. As such, there is nothing to cast doubt on the validity of the response data. However, if the assumption is made that customers provide feedback independently of one another, then this can be analysed as a Markov process. As such, the time interval between any two consecutive customers providing feedback will be a random variable with an exponential distribution. The distributions of the time intervals between responses at various outlets showed similar characteristics to an exponential distribution with the same mean; differences between the distributions of the time intervals between responses and the exponential distributions are not readily apparent when charted on linear scaled axes.

Figure 4 illustrates the distributions of the time intervals between responses for two outlets compared with an exponential distribution, using a logarithmic scale for the time interval rather than a linear scale.

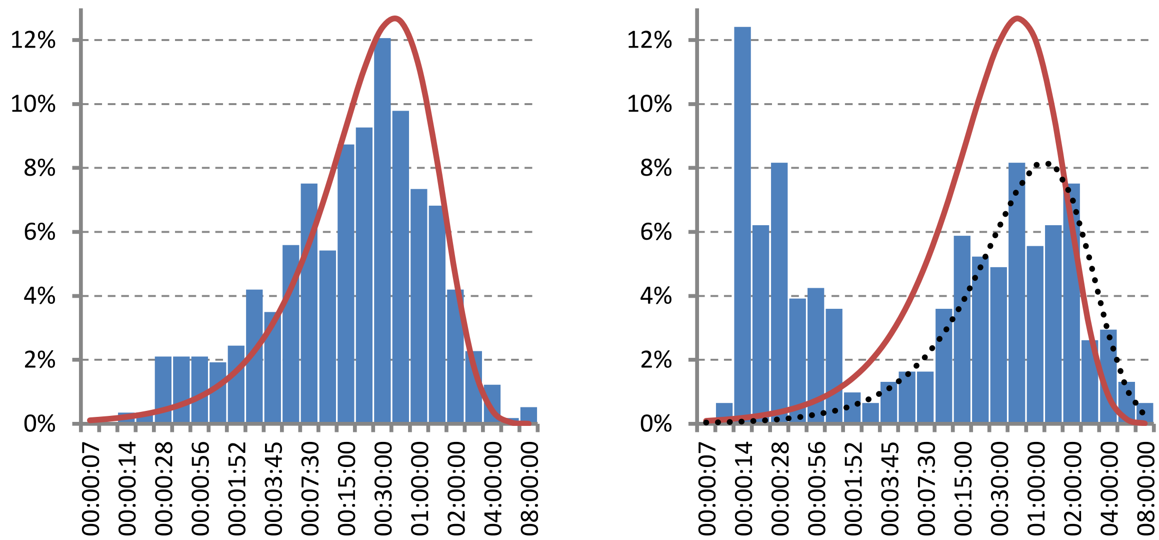


**Figure 4:** Distributions of the time intervals between responses (logarithmic scale)

It is evident that the distribution of time intervals between responses for the outlet charted on the left is a satisfactory approximation to an exponential distribution. Therefore, the assumption that customers provide feedback independently of one another appears reasonable. In clear contrast, the distribution of time intervals between responses for the outlet charted on the right is substantially different from an exponential distribution. Therefore, the assumption that customers provide feedback independently of one another is clearly unreasonable. In this instance, 39.2% of responses were received within 1 minute 20 seconds of each other, whereas if the responses were indeed independent, it would have been expected that only 3.5% of the time intervals would have been less than this time.

Clearly the responses for which the time interval distribution is charted on the right of Figure 4 were not independent of one another. While there are arguments for some degree of interdependence of responses during the data gathering process, the magnitude of the data anomaly is evidence of data fabrication.

It is interesting to observe in Figure 4 that, in the anomalous dataset, if the responses with time intervals are less than 1 minute 20 seconds are ignored, then the distribution of the time intervals between responses closely approximates the exponential distribution as expected. Interestingly, analysis of the anomalous responses with time intervals of less than 1 minute 20 seconds showed a distinct negative bias. This suggests that dissatisfied customers may have been falsifying data by repeating their responses as a way of emphasising and reiterating their dissatisfaction.

### 3.4 Character counts of open ended responses

In this case, a telephonic survey included six open-ended questions; interviewers were required to capture participants' responses as close to verbatim as reasonably possible. This textual data was then subjected to a thematic content analysis. An additional factor in the data gathering was that although the survey was compiled in English, some of the participants were not first language English speakers, and interviews may have been carried out partially in a vernacular language, requiring some translation by the interviewer.

As participants were randomly assigned to the eight telephonic interviewers, it would be expected that there would be no significant difference in the amount or the content of the qualitative data captured by the interviewers. A reasonable quantitative hypothesis would therefore be that there is no difference in the average string lengths of open ended responses captured by the interviewers. To obviate the need to make any assumptions about the distributions of string lengths of open ended responses, nonparametric one-way ANOVA tests have been carried out on the responses to the six open-ended questions, the results of which are tabulated in Table 2.

**Table 2:** Median character counts and results of Nonparametric One-Way ANOVA

| | Number of interviews | Question 1 | Question 2 | Question 3 | Question 4 | Question 5 | Question 6 |
|---|---|---|---|---|---|---|---|
| Interviewer A | 47 | 28 | 15 | 51 | 33 | 22 | 7 |
| Interviewer B | 48 | 70.5 | 59 | 133 | 129.5 | 110 | 92.5 |
| Interviewer C | 45 | 42 | 37 | 36 | 30 | 24 | 5 |
| Interviewer D | 18 | 17 | 15 | 99 | 60.5 | 47 | 8 |
| Interviewer E | 50 | 23.5 | 34 | 81.5 | 54 | 50 | 8 |
| Interviewer F | 25 | 74 | 61 | 64 | 58 | 50 | 14 |
| Interviewer G | 25 | 65 | 50 | 62 | 48 | 36 | 7 |
| Interviewer H | 42 | 81.5 | 64.5 | 80.5 | 56.5 | 33 | 4 |
| | $\chi^2$ value | 140.03 | 124.05 | 98.90 | 82.25 | 91.96 | 119.08 |
| | $p$ value | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 | <0.0001 |

It is clearly evident that Interviewers A and C have consistently captured significantly shorter responses than, for example, Interviewer B. As respondents were randomly assigned, it may be inferred that this significant difference is attributable to the specific interviewers. This may indicate data falsification, or at least some variability of data quality.

## 4 Non-parametric statistics and bogus data

The underlying premise of applying inferential statistics to detect bogus data is that within any population there exists variability of demographics, perceptions, attitudes, opinions, etc. that are difficult for any single individual to emulate reliably in both magnitude and distribution (Mosimann, Wiseman, & Edelman, 1995; Schraepler & Wagner, 2005). Therefore, by hypothesising that the data *are* random, test statistics can be defined which have well defined confidence intervals for the case of random data. Then, should any given test statistics fall outside the predetermined confidence interval, provided the research design has been consistent with random sampling, it may be inferred that the data are *not* random and may be attributable to data fabrication or falsification. Well-established statistical tests are available to test such an hypothesis, some of which are identified below.

### 4.1 Wald–Wolfowitz runs test

Where a dataset comprises individual participants' responses to a battery of survey questions or items, it is reasonable to hypothesise that consecutive records will be mutually independent of one another. Conversely, if consecutive records of the dataset are not independent of one another, a plausible explanation is that the data has not been captured from independent participants, and may have been fabricated.

The Wald–Wolfowitz runs test (Wald & Wolfowitz, 1943) can be applied to any dichotomous variable and tests the hypothesis that elements in a sequence are mutually independent of one another. For example, if males and females are being randomly sampled from a given population, then the expected number of "runs" (consecutive participants of the same gender) in a given sample can be calculated with a corresponding confidence interval, and the sample statistic follows an approximation to the normal distribution. If the sample statistics (the actual number of runs) fall outside the confidence interval, then it can be inferred that consecutive records in the dataset are not mutually independent.

The attractiveness of the Wald–Wolfowitz runs test in the context of potentially bogus data is that it is distribution free and can be applied to categorical variables, numerical variables and missing data.

### 4.2 Siegel–Tukey test

It has been mentioned that it is difficult for an individual interviewer to emulate the distribution of a population variable (Mosimann et al., 1995; Schraepler & Wagner, 2005). The value of the Siegel–Tukey test (Siegel & Tukey, 1960) is that it is distribution free, and can be used on both ordinal level and numeric variables. It tests whether or not two groups of data have the same dispersion or variability of values. By testing the dispersion of data captured by each interviewer against the remainder of the sample data, the Siegel–Tukey test may be used to identify the lower variance of specific answers discussed by Schraepler and Wagner (2005).

### 4.3 Chi-squared goodness of fit test

If survey respondents have been selected at random from a given sample frame, then the distribution of demographic characteristics of the sample should closely match that of the sample frame. Where an interviewer has fabricated responses, it is unlikely that they will be able to replicate the demographics of the fabricated responses reliably.

Using the chi-squared goodness of fit test it would be hypothesised that the distribution of the demographics of each interviewer's sub-sample would not be significantly different from that of the sample frame or population. Should the result of the test be statistically significant then it would be regarded as an anomaly and a possible indication of fabrication or falsification of the data. The discriminance of the test would be considerably enhanced by evaluating the bivariate goodness of fit (e.g. age category and gender) rather than the conventional univariate goodness of fit, as this would be more difficult for a perpetrator to fabricate.

## 5 Principles of detection of bogus data

### 5.1 Prerequisites and assumptions

A prerequisite for detecting bogus data is that there is a degree of consistency in the genuine data in order that anomalies due to fabrication or falsification will manifest. Such consistency would exist provided the sample is sufficiently representative of the research population, and therefore it is assumed that the data is derived from a representative sample. Should this not be the case, fabricated or falsified data may not manifest as an anomaly and the malpractice may go undetected.

If data equivalence is not ensured in the research design, then anomalies may arise that are due to some form of systematic bias in the data collection process rather than due to deliberate misconduct, and therefore it is assumed that issues that may affect equivalence (e.g. language, culture, and data collection medium) have been adequately resolved. If issues of equivalence remain unresolved, then anomalies may be misdiagnosed.

It may seem trivial to state the assumption that there exists innate variability in any population. However, it is worth stating this assumption because detection of bogus data is frequently dependent on deviations from the distributions or parameters that characterise the variability of the population.

### 5.2 Analysis of detection of bogus data

Some consistent themes are evident in the detection of fabricated or falsified data. Unless the perpetrator is unusually naïve or clumsy, data that has been falsified or fabricated will appear at face value to display the same characteristics as legitimate, reliable data. However, genuine datasets have a variety of characteristics and properties representing the underlying data, not all of which necessarily pertain explicitly to the particular research study. Falsified and fabricated data generally does not manifest all the underlying characteristics and properties identically; arguably if it could, then the data can be regarded as reliable and the malpractice becomes moot.

Data fabrication and falsification may therefore be detected by analysing ancillary data, including metadata, rather than the variables that pertain directly to the research. The analysis of the response time intervals and the character counts of open ended responses are examples from preceding case studies. Similarly, data fabrication and falsification may be detected by carrying out extraneous analyses that are unrelated to the central topic of the research. Examples from preceding case studies are the fitting of distributions to the response data and analysing the factor structure of the responses.

Analyses for the purpose of detecting bogus data follow the scientific method, and are conceptually similar to and comparable with statistical hypothesis testing. In effect, the null hypothesis is that the research data are genuine, legitimate and valid, and the alternative hypothesis is that the data has been fabricated or falsified. The data are not rejected as bogus unless anomalies manifest during the analyses which may be the result of malpractice. Of course, there remains a finite probability (referred to in inferential statistics as the significance level) that anomalous data is valid and not indicative of any misconduct.

### 5.3 Conflicts of interest

A generally accepted definition of a conflict of interest in research is a situation "in which financial or other personal considerations may compromise, or have the appearance of compromising a researcher's professional judgment in conducting or reporting research" (e.g. University of California, 2016). An application of this definition is that a conflict of interest may arise when the reliability and validity of the research may not be the highest priority for a stakeholder or participant in the research. Therefore data fabrication and falsification would typically be directly associated with some degree of conflict of interest.

None of the stakeholders in research are necessarily insusceptible to conflicts of interest. For example, the "publish or perish" imperative for academic researchers (Herndon, 2016) is a prime example in which the priority of obtaining novel and meaningful (i.e. publishable) outcomes may supersede research integrity. Professional researchers may have a commercial imperative to fulfil client expectations that may transcend sound research ethics, while the perverse motivation of those involved with data collection to fabricate response data has been addressed earlier.

The need to address the improper or undue exercise of power in government was recognised long ago in ancient Greece by the philosopher Aristotle. However, the explication of separation of powers is attributed to Montesquieu in his 1748 work *The Spirit of Laws* (Casper, 1989) as he described what we now call the legislature, the executive and the judiciary, and is referred to as the *trias politica*. Simplistically, the principle of *trias politica* is that there is a balance of power between the three branches of government, preventing any undue exercise of power, which is regarded as a fundamental principle of good governance. In essence, *trias politica* structurally militates against conflicts of interest.

## 6 A research protocol for militating against data fabrication and falsification

The impact of fabricated or falsified research data can potentially be mitigated by detecting the bogus data, excluding it from the dataset and, where practicable, replacing it with reliable data. A more robust approach would be to adopt a research protocol that militates against fabrication and falsification.

Two principles underpin the research protocol that would militate against data fabrication and falsification. The first is that the ethical obligations of researchers should include disclosure of diagnostic analyses carried out with the sole purpose of detecting or demonstrating the nonexistence of fabricated or falsified data. This recommendation is based on the premise that researchers cannot ethically abdicate or outsource responsible for data quality. Indeed, such disclosure would be explicit evidence in support of the validity and reliability of their research.

The second foundation of the proposed protocol is the principle of *trias politica*; that is, the separation of the power and responsibility for research design plus substantive analysis, data collection, and data verification. Using the three branches of government as analogies, it is suggested that there are parallels between the role of research design plus substantive analysis and the role of the legislature in government, between data collection and the executive, and between data verification and the judiciary. The legislative branch of government is responsible for determining policy and writing the laws that will give effect to the policy. The analogous role of research design plus substantive analysis is to operationalise the research objectives by designing an appropriate research instrument and carrying out the substantive analysis of the data so as to resolve the research problem. The executive branch of government is involved with implementing the laws that have been written by the legislature. The parallel in research is that those involved with data collection are giving effect to the research design of those who compiled the research instrument. Finally, the judicial branch of government neither makes nor enforces the law, but rather ensures the appropriate interpretation and application of the law and may have some influence on the legislative process. In a similar sense, the data verification role is concerned primarily with the validity and reliability of the data that has been collected, although there may be some contribution to the design of the research instrument.

The proposed *trias politica* research protocol requires that the research design plus substantive analysis, the data collection and data verification are treated as independent (i.e. separate and discrete) research activities. The practical implication of this is that the three activities may be carried out by different parties, or at the very least, be addressed as separate activities when conducting and reporting on research.

## 7 Practicability of the trias politica research protocol

Adoption of the *trias politica* research protocol would have an important and substantial impact on the way business research is carried out.

In the case of large, funded, contract research, it would be appropriate for the client to appoint separate and independent contractors to carry out the research design plus substantive analysis, the data collection and the data verification. The research design should facilitate data verification by avoiding nonprobability sampling and interview quotas, including survey items with predictable response distributions (e.g. Benford's distribution), and incorporating ancillary variables and metadata (e.g. date, time, place, length of interview, interviewer, medium, and language) into the research dataset. Research reports should include disclosure of the results of analyses of ancillary variables and metadata designed to detect or demonstrate the absence of bogus data.

In research and consulting organisations, it would be appropriate to ensure that ethical "screens" or "walls" are employed between the departments or individuals that carry out the distinct functions of research design plus substantive analysis, data collection and data verification.

Academic research is often characterised by limited resources rendering strict separation of powers and responsibilities impracticable. A compromise solution would be to adopt a consortium approach, in which a cohort of researchers are each randomly and anonymously assigned the data collection and data verification functions of their research colleagues. Although unethical collaboration could not necessarily be ruled out, some degree of separation of responsibilities would be achieved.

Finally, ethical boundaries would be paradoxical in the case of individual academic research. However, it is suggested that the individual researcher would have the same obligation to design and disclose those measures that ensured the research data was authentic.

Resnik and Elmore (2016) note that peer reviewers seldom detect instances of research misconduct, including data fabrication and falsification. In order to hold researchers accountable it is recommended that all research users (examiners and moderators of academic research, reviewers of manuscripts and journal editors, clients of consultants and marketing researchers, professional bodies through implementation of codes of good practice, academic institutions, research students and their supervisors) raise awareness of vulnerability to data fabrication and falsification, and accept responsibility for enforcing more rigorous research protocols.

## References

American Association for Public Opinion Research. (2003). "Interviewer Falsification in Survey Research: Current Best Methods for Prevention, Detection and Repair of Its Effects", Retrieved 2015-12-28, from https://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/falsification.pdf

Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). "Editorial: Evidence on Questionable Research Practices: The Good, the Bad, and the Ugly", Journal of Business and Psychology, 31(3), p.323-338.

Benford, F. (1938). "The law of anomalous numbers", Proceedings of the American Philosophical Society, 78(4), p.551-572.

Birnbaum, B., Borriello, G., Flaxman, A. D., DeRenzi, B., & Karlin, A. R. (2013). "Using behavioral data to identify interviewer fabrication in surveys", Paper presented at the SIGCHI Conference on Human Factors in Computing Systems.

 ISSN 1477-7209

Bredl, S., Storfinger, N., & Menold, N. (2011). "A literature review of methods to detect fabricated survey data", Discussion Paper, Zentrum für internationale Entwicklungs-und Umweltforschung.

Bredl, S., Winker, P., & Kötschau, K. (2012). "A statistical approach to detect interviewer falsification of survey data", Survey Methodology*,* 38(1), p.1-10.

Casper, G. (1989). "An Essay in Separation of Powers: Some Early Versions and Practices", William and Mary Law Review*,* 30, p.211-262.

Clark, J. R., & Moul, D. A. (2004). "Census 2000 Testing, Experimentation, and Evaluation Program Topic Report No. 10, TR-10, Coverage Improvement in Census 2000 Enumeration", Washington, DC.

Diekmann, A. (2007). "Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data", Journal of Applied Statistics*,* 34(3), p.321-329.

Grubbs, F. E. (1969). "Procedures for detecting outlying observations in samples", Technometrics*,* 11(1), p.1-21.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). "Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis", Organisational Research Methods 7(2), p.191-205.

Herndon, N. C. (2016). "Research Fraud and the Publish or Perish World of Academia", Taylor & Francis.

Horn, J. L. (1965). "A rationale and test for the number of factors in factor analysis", Psychometrika*,* 30(2), p.179-185.

Judge, G., & Schechter, L. (2009). "Detecting problems in survey data using Benford's Law", Journal of Human Resources*,* 44(1), p.1-24.

Kennickell, A. B. (2002). "Interviewers and data quality: Evidence from the 2001 Survey of Consumer Finances", Paper presented at the August 2002 Joint Statistical Meetings, New York.

Kennickell, A. B. (2015). "Curbstoning and culture", Statistical Journal of the IAOS*,* 31(2), p.237-240.

Mosimann, J. E., Wiseman, C. V., & Edelman, R. E. (1995). "Data fabrication: Can people generate random digits?", Accountability in Research*,* 4(1), p.31-55.

Newcomb, S. (1881). "Note on the Frequency of Use of the Different Digits in Natural Numbers", American Journal of Mathematics*,* 4(1), p.39-40. doi: 10.2307/2369148

Resnik, D. B., & Elmore, S. A. (2016). "Ensuring the quality, fairness, and integrity of journal peer review: A possible role of editors", Science and Engineering Ethics*,* 22(1), p.169-188.

Schraepler, J. P., & Wagner, G. G. (2005). "Characteristics and impact of faked interviews in surveys–An analysis of genuine fakes in the raw data of SOEP", Allgemeines Statistisches Archiv*,* 89(1), p.7-20.

Siegel, S., & Tukey, J. W. (1960). "A nonparametric sum of ranks procedure for relative spread in unpaired samples", Journal of the American Statistical Association*,* 55(291), p.429-445.

Simmons, K., Mercer, A., Schwarzer, S., & Kennedy, C. (2016). "Evaluating a new proposal for detecting data falsification in surveys", Statistical Journal of the IAOS(Preprint), p.1-12.

Stacey, A. G. (2005). "Reliability and Validity of the Item Means and Standard Deviations of Ordinal Level Response Data", Management Dynamics*,* 14(3), p.2-25.

Stacey, A. G. (2012, 28-29 June 2012). "Parameter Estimation Using Asymptotic Analogy", Paper presented at the 11th European Conference on Research Methods, University of Bolton, United Kingdom.

University of California. (2016, 2016-03-10). "Conflict of Interest (COI)", Retrieved 2016-08-24, from http://coi.ucsf.edu/

Wald, A., & Wolfowitz, J. (1943). "An exact test for randomness in the non-parametric case based on serial correlation", The Annals of Mathematical Statistics*,* 14(4), p.378-388.

Winker, P., Menold, N., Storfinger, N., Kemper, C. J., & Stukowski, S. (2013). "A method for ex-post identification of falsifications in survey data", Interviewers' Deviations in Surveys–Impact, Reasons, Detection and Prevention, p.25-47.