American Journal of Applied Sciences

Original Research Paper

Biases from Poor Data Analyses

Tshepo Matsose and Solly Matshonisa Seeletse

Department of Statistics and Operations Research,
Sefako Makgatho Health Sciences University, PO Box 107, MEDUNSA, 0204, Gauteng Province, South Africa

Article history
Received: 10-07-2016
Revised: 29-09-2016
Accepted: 04-10-2016

Corresponding Author:
Solly Matshonisa Seeletse
Department of Statistics and Operations Research, Sefako Makgatho Health Sciences University, PO Box 107, MEDUNSA, 0204, Gauteng Province, South Africa
Email: solly.seeletse@smu.ac.za

Abstract: Non-statisticians with little knowledge in basic descriptive statistics tend to think that statistics field is limited to the content to which they are exposed. Many of them believe that a statistical package can augment the little Statistics knowledge they have. They often have a tendency to perform their own data analyses and do not even bounce it against Statistics experts for quality check. Many studies were concluded from data analyses performed by analysts who lack insight into statistical methods. Hence, results in some of their researches have flaws and distorted truths. The paper explains the defects in data analyses and research results that can be caused by influences in the data. Flawed research results may be caused when the data were not scanned for variations and other inconsistencies present in the data. Properly trained statisticians who also understand theories and methods of dealing with outliers can perform these analyses more effectively. However, many researchers fail to seek their advices. This study shows the extent of falsifications that contaminated data can produce and the massive loss to the factualness contained in the data.

Keywords: Data Variations, Information Falsification, Statistical Falsehood

Introduction

Statistics is an applied mathematics grounded on Mathematics concepts (Galbraith and Stone, 2011; Stolz, 2002). Statistical methods enable easy understanding and explanation of facts. There are many circumstances in which Statistics methods misinform and deceive the naïve observer into trusting distortions. Some cases are deliberate information spin for the personal advantages of the perpetrator while other cases occur due to analysts' incompetence. These cases constitute exploitation of Statistics in which a statistical argument is used to lie. Steele (2005) asserts that some Stat cases have misuses that are accidental. Any detectible errors in analysis constitute the minimum error in the analysis. There may be undetectable or hidden errors adding to the minimum.

Myths exist in Statistics practice and the false Statistics trap can damage the pursuit for honest knowledge. For instance, in the health sciences, correcting a myth may take decades while costing lives. Seife (2011) observed many fabricated journal articles published with unsound statistical methods. Misuses can occur easily. Ercan *et al*. (2007) also discuss numerous misusages of Statistics in medical research in which mistakes in applying statistical methods were not noticed before analyses. Consequently, false results were reported. The point is, many journal articles carry misleading conclusions which cannot assist to improve practice.

Professional scientists, including mathematicians and professional statisticians, can be fooled by some simple statistical methods, even if they are careful to check everything. Scientists have been known to fool themselves with Statistics due to lack of knowledge of probability theory and lack of standardization of their tests. Several authors (Asher, 1998; Best, 2002; Maier, 1999) have confirmed their experiences that politicians tend to use Statistics for support rather than for information. Thus, both illiteracy in statistical techniques and the intention to mislead can influence deficiencies in data analyses.

Data validity is a known idea, but practice to ensure it does not always happen. Many studies were completed with data that had not been pre-cleaned. Variations in data, such as outliers, could distort the results. However, a small proportion of analysts know about robust methods and cleaning data of outliers. The researches (mainly on business, health and medicine and social sciences) that were completed without consideration of variations in data or robust methods in data analyses could be flawed. This paper exposes the extent of falsification possible when data analyses do not include data pre-cleaning for influential elements.

Stimulants of this Paper

Research students of Sefako Makgatho Health Sciences University (SMU) wanted help with data interpretations from their own analysis in large numbers. Many of them had used incorrect methods and a lot more depended on conventional statistical methods that had not taken care of influential observations. From Statistics expertise viewpoint, many of the results were obviously very flawed. Upon realising that other studies from the health science disciplines could have been presented with statistical mistakes, the authors of this paper visited the SMU library to read some of the dissertations completed from the health sciences of SMU at the master's and doctoral degree levels. The observations of this effort triggered communication regarding inappropriate methods of data analysis. It justified the inclusion of data cleaning, outliers and robust statistics, among others, in the curricula of courses in basic Statistics at SMU. It also inspired an illustration with a formal academic paper. This paper was therefore motivated by the observations experienced from stored dissertations in the SMU library. The values indicating the measures required for that study should be analysed to show that little minimum input of outliers can provide massive falsification of meaning in the results.

Purpose of this Paper

The aim of this paper is to enlighten that data analysis should not be done by inadequate data analysts. Where simple analyses are done by non-experts, they should be quality assured by proficient, qualified and fully trained data analysts such as statisticians. The objectives were firstly, to expose outliers in a proper dataset collected for a study in SMU and then secondly, to illustrate the various magnitudes of fact distortions.

Statistical Modelling

Modelling refers to a practice of developing an archetypal description of a system by using concepts and language (Sokolowski and Banks, 2009). A statistical model is developed in order for statistical and quantitative methods to describe a phenomenon. It is usually presented as mathematical equations connecting random variables (Freedman, 2009). Statistical modelling enables statistical tests and estimation in order to ultimately make statistical inferences. Konishi and Kitagawa (2008) explain that a statistical model has three purposes, namely; prediction, extraction of information and description of stochastic structures. If a model is not an accurate depiction of the system, any prediction or results drawn from the model may distort the results required from a study. Some analysts develop models and never get to verify that they are accurately representing the system under study. The developed models should be tested for accuracy.

Statistics Teaching Non-Experts

The advice that one may give to a headache sufferer to drink a tablet does not qualify the advisor as a medical expert. However, there is a common tendency of experts of other fields to believe they can be statistical experts for having done a basic course in Statistics (Nikoletseas, 2014). There is more to Statistics than the basic concepts taught. The mathematics essentials taught in Statistics for statisticians are not taught with basic Stat courses. These mathematics concepts are crucial for data analyses. The distance measures for example, are basic supports for residual analyses which are useful in measuring error, or bias. When Statistics courses are taught by non-experts in the subject, there are often gaps undetected in the content. This gap is inherited in the statistical analysis performed by these researches (Lekganyane, 2015). The fact is, even if a student can know almost 100% of the knowledge in Statistics in the basic courses designed for applications in other fields, the knowledge acquired is so minimal and cannot make the student a statistician.

Outliers

There could be some values in a dataset that lie too different from the bulk. An outlier is an observation that is far removed from the rest of the observations (Maddala, 1992). These are the kind of datasets which naïve data analysts analyse without testing their influence on the results. The fundamental concern is to ensure that outliers do not prejudice the results of the analyses. Unchallenged interpretation of statistics derived from data containing outliers may misinform the audience or users (Liu *et al*., 2004).

Statistical methods do exist to test if some data in a dataset are outliers. However, some outliers can be easily identified without having to employ sophisticated statistical methods while others are concealed. A conservative habit is that conventional courses in statistical methods do not teach about outliers. They also tend to involve small simulated datasets for illustrations. The fact is that teaching methods in Statistics are based on the assumption that the data used are clean without polluters. On the other hand, when outliers exist in the data, they can be detrimental to analysis results. It is therefore crucial that every data analysis exercise ensures that outliers are treated to avoid their influence on the results. The outliers identified in any dataset logically signify the minimum error available in the data. Other outliers and influences may be undetected, or even undetectable. Common effective statistical methods to address outliers are robust analyses and removal of outliers. These methods follow below.

Robust Statistical Methods

The descriptive methods taught in basic Statistics are easily swayed by outliers and influential methods. However, nearly all of them have counterparts in robust analysis. The problem of robust methods is their failure to maintain the efficiencies of conventional methods. Robust statistical methods are statistical methods that are resistant to influences posed by outliers and other influential observations (Jaulin, 2010). When outliers appear in large datasets, they should not automatically be discarded. Robust methods can assist in outlier identification, which can be difficult, but still ensuring that outliers do not distort the results (Chambers *et al.*, 2004; Dawson, 2011). Data analysis should apply robust methods. Some efficient robust statistical methods developed along common traditional ones are the least absolute deviation, least trimmed squares, S-estimation and M-estimators (McKean, 2004; Rousseeuw and Leroy, 2003; Strutz, 2010). Advancing robust methods includes MM-estimation which pools the robustness of S-estimation with the efficiency of M-estimation (Hampel *et al.*, 2005).

Removing Outliers

Some data analyses benefit best by removal of outliers. However, before considering to remove outliers from the data, an attempt should be made to understand why they appeared in the first place and whether it is likely that similar values will continue to appear (Steele, 2005; Tufte, 1997). Outliers can contain valuable facts about the process under investigation, or the data gathering and recording process. They can also be bad data points. Qualitative judgment may be used to decide on removal and retention of outliers. An error attributable to an outlier in the study should be deleted. Other outliers may be kept in the data. Outliers honestly obtained and giving new insight into the phenomenon being measured should be kept unless analysed separately. Outlier removal can lead to omitting information signaling a new discovery. Thus, outliers should be analyzed separately when removed for data analysis.

Detecting Outliers

Small and large outliers can be identified using lower and upper bounds. Define Q_1 and Q_3 as the first and third quartiles of a dataset. Let *k* be a barrier constant normally chosen to be either 1.5 or 3. According to Dovoedo (2011), Tukey's boxplots boundaries' method defines outliers as observations outside the interval with lower and upper boundaries:

$$L = Q_1 - k\left(Q_3 - Q_1\right) \tag{1}$$

$$U = Q_3 + k\left(Q_3 - Q_1\right) \tag{2}$$

Data Analysis Issues

Analyzing data entails a process of inspecting, cleaning, transforming and modelling raw data in order to discern useful information, suggest conclusions and support decision-making (O'Neil and Schutt, 2014). It covers organisation, cleaning, exploration, analysis and interpretation. The process converts data into information useful for decision-making. It has many aspects and approaches, including diverse techniques under many names and various domains. Hair (2008) explains that effective data analysis entails extracting relevant facts from the analysed data to answer the research questions, support a conclusion and/or test a hypothesis. Such facts should be undeniable such that other analysts can independently confirm them. However, effective data analysis does not always take place. There are many published studies that have been flawed in one or other data analysis aspect. Sources of ineffectiveness include poor statistical modelling, deficient data analysis, failure to clean the research data and the presence of outliers in the data (Becker and Gather, 2001).

Many analysts reach flawed conclusions due to outlier influence. Zwane (2015) showed that outliers often signify lies alongside facts. If not restrained, outliers can have limitless deceptions. Jaffe and Spirer (1987) warn that some outliers are consciously included to sway data analysis results for selfish reasons. Furthermore, barriers to effective analysis may exist if outliers, influential observations and leverage points are in the data. Paulos (1988) warn against doing statistical data analysis when not mathematically literate. Many cannot split fact from opinion. Therefore, cognitive biases and innumeracy are challenges to sound data analysis. According to Cousineau and Chartier (2010), novices in Statistics can neither detect nor treat outliers.

Statistical Illustration

Study Context

A study was undertaken for an Occupational Therapy (OT) project by Occupational Therapy IV research students of SMU. It was undertaken at the Dr George Mukhari Academic Hospital in Gauteng Province, South

Africa. It measured some traits of nurses working in the paediatric ward regarding some OT issues. The example presents only one random variable. Instances with many variables may include multicollinearity, which escalates data complexities. Members of the SMU's Department of Statistics and Operations Research were asked for help during data collection planning and analysis.

Data Collection

A structured questionnaire was used for data collection. Data capturing benefited from using a spreadsheet Table 1.

Data Organisation

This dataset is clear and simple. Three outliers are easily identified. However, the OT researchers involved could not identify the outliers. It took an effort to make them understand this. Table 2 below and face-value inspection are used to identify the outliers.

Even from Table 2, outliers are easily detectible. An analysis is undertaken to exhibit the extent of impact imposed on analyses. This demonstrates that distortions can occur even at very large proportions if analyses are based on the data with outliers.

Outlier Identification

The recognizable outliers are 26, 30 and 31. It may be difficult to identify outliers. Three current outliers are easy to identify. However, proper outlier detection methods are vital to ensure that concealed outliers are also recognized. Hence, outlier identification Equation 1 and 2 are required. The cumulative frequencies are needed:

From Table 3, $Q_1 = 2$ and $Q_3 = 4$. From equation (1) with $k = 3$, the lower boundary is:

$$L = Q_1 - k(Q_3 - Q_1) = 2 - 3(4 - 2) = -4$$

From Equation 2, the lower boundary is:

$$U = Q_3 + k(Q_3 - Q_1) = 2 + 3(4 - 2) = 4$$

Values below –4 and above +8 are therefore outliers. The initially identified outliers 26, 30 and 31 are also confirmed. Thus, removing these outliers is justified. This paper measures the effect of outliers as difference made on the results with the minimum input of outliers against the results when outliers are removed from the data. The illustration demonstrates the impact of this minimum input.

Compared to Table 2 totals, the total of values in Table 4 shows a massive difference. The next table reveals the difference in the descriptive analysis of data with outliers against the analysis of the same data with outliers removed. The impact of the fact distortion that is due to the presence of outliers is obtained from the difference of the two results. This difference is the impact of outliers.

Table 5 generated 14 descriptive measures. Three measures not affected by outliers are the median, mode and minimum. Therefore, outliers can distort 11 of 14 (79%) of the facts in the descriptive measures. Furthermore, most of these descriptive measures have been affected by huge percentages. Only mean and sample size were affected by less than 75% each. These influences include extents of 89% and 95% of influences. Also, the value affected the most by outliers is the impact of over 310%.

Clearly, if the outliers are not addressed, the final results will be hugely affected. Hence, severe outliers should be removed before data analysis. In addition, the average distortion by outliers amounts to just over 91% with a standard deviation of about 77%. The next display demonstrates that data shapes with outliers are also distorted.

Appraisal of Outlier Impact

Graphical and Structural Falsification

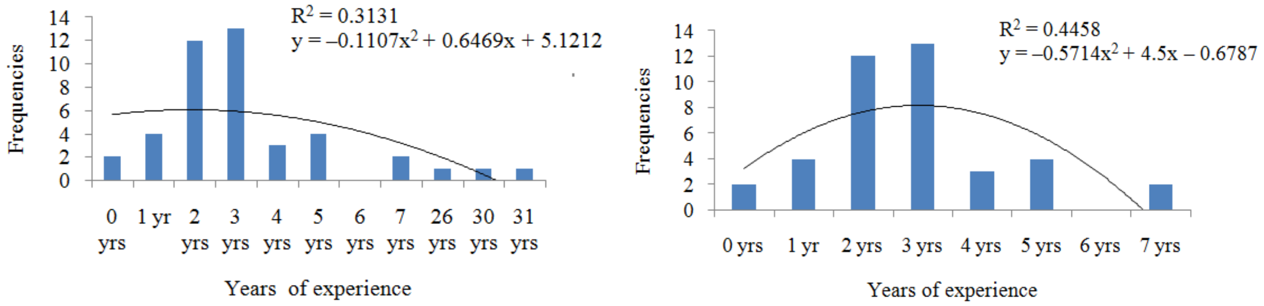
The data with outliers (Fig. 1a) show less accuracy ($R^2 = 0.3131$) as compared with data without outliers ($R^2 = 0.4458$ in Fig. 1b). Also, the coefficients of the quadratic equation differ widely between the one with outliers and the one without outliers. This difference justifies removal of the outliers as a way to clean the data of inconsistencies.

The trend line in data with outliers seems to indicate a hyperbola shape while a parabola is more suited when the outliers have been removed. This adds to the number of distortions from fact as caused by outliers. Structurally, the mathematical models for the two cases give parabola equations. However, these have completely different coefficients and intercept. This is another impact of the outliers.

The other outlier effect was on the quality measure. Outliers produced $R^2 = 0.3131$, which increased to $R^2 = 0.4458$ after outlier removal. Therefore, the quality reduction using the R-squared measure in this case was about 30%.

Least Possible Deception

This section focuses only on the distortion caused by the three outliers that were removed out of 43 values. Other potential distortions could come from the outliers left in the data for analysis. Hence, despite the massive amounts of distortions shown, the distortion presented is the minimum potential impact resulting from the least input of outliers.



(a)

(b)

Figure 1. Bar chart illustrating outlier effect, (a) Bar chart *with* outliers, (b) Bar chart *without* outliers

Table 1. Raw data

26	1	3	2	3	4	3	1	3	3	5	2	2
2	2	1	3	3	3	2	2	0	2	3	5	3
31	7	5	0	3	3	2	7	3	5	2	2	4
30	4	2	1									

Table 2. Frequency table

	Categories											Total
Values	0	1	2	3	4	5	6	7	26	30	31	115
Frequencies	2	4	12	13	3	4	0	2	1	1	1	43

Table 3. Cumulative frequency table

Values	0	1	2	3	4	5	6	7	26	30	31
Cum frequencies	2	6	18	31	34	38	30	40	41	42	43
Quartile location			Q_1	Q_2	Q_3						

Table 4. Frequency table without outliers

	Categories								Total
Values	0	1	2	3	4	5	6	7	28
Frequencies	2	4	12	13	3	4	0	2	40

Table 5. Descriptive statistics

Values *with* outliers		Values *without* outliers		Impact of outliers
Mean	4.6512	Mean	2.8250	39.3%
Standard Error	1.0581	Standard Error	0.2478	76.6%
Median	3	Median	3	0.0%
Mode	3	Mode	3	0.0%
Standard Deviation	6.9381	Standard Deviation	1.5671	77.4%
Sample Variance	48.1373	Sample Variance	2.4558	94.9%
Kurtosis	9.8720	Kurtosis	1.1198	88.7%
Skewness	3.2642	Skewness	0.8100	75.2%
Range	31	Range	7	77.4%
Minimum	0	Minimum	0	0.0%
Maximum	31	Maximum	7	77.4%
Sum	115	Sum	28	310.7%
Count	43	Count	40	7.0%
Confidence Level (95.0%)	2.1352	Confidence Level (95.0%)	0.5012	76.5%

From 14 descriptive measures, over six other distortions caused by outliers were revealed. These came from graph shape, regression equation (3 coefficients and sign of intercept) and quality of fit. These were 20 statistical components, 17 of which were distorted by outliers. This is a distortion of a massive 85%, caused by only about 10% false inputs (3 outliers of 43 values). This is consistent with the proposition that a tiny lie can cause maximum distortion.

Reflection

During the data analysis stage, the outliers were pointed out and explained to these OT researchers. The huge effect that the outliers caused on the initial statistics was demonstrated. These researchers were utterly shocked. They stated that the people who taught them Statistics never mentioned outliers, or data cleaning. They were not even warned about the possibility of results being influenced by data contaminations. Since many other studies from the health science disciplines could have been presented with Statistics mistakes, this initiative triggered communication regarding inappropriate methods of data analysis. It justifies the inclusion of data cleaning, outliers and robust statistics, among others, in the curricula of courses in basic Statistics at SMU. The values indicating the measures required for that study were analysed to show that one or few outliers can massively falsify facts.

Conclusion

The study demonstrated with actual data that results would have been distorted by outliers if there was no intervention of qualified statisticians. The OT thesis would have used statistical measures from analysis by occupational therapists and then assessed by an OT expert without realising the major information biases caused. Distortion caused by outliers would most likely have been missed without using robust methods which the non-statisticians do not know. This weakness in data analysis is valid with many articles in journals and dissertations in libraries. The deceptions contained in these documents augment to the inadequacies of studies to address real life problems in those fields.

In reiterating, this paper reveals that the smallest contamination in the data can cause massive amounts of distortions. The articles published in many journals could be products of distortions from the theses and dissertations kept in the various libraries. This could be one reason for limited progress in some practices of many disciplines despite findings and recommendations from the studies completed. Flawed results reached deliberately or due to ignorance are worthless in actual practice. Hence, effective practice fails to improve as many researches are guided by untruths.

Recommendations

The study recommends to the researchers in the social, business and medical sciences (among others), that prior to performing statistical analyses; analysts who are not statisticians should seek advice from Statistics experts.

Those analysts who know some Statistics and performing own analyses should also bounce the results of their analyses against real experts, not just experienced researchers.

Also, for practical usefulness of research, this study recommends that postgraduate research should be undertaken on the basis of its value in actual practice, not merely for submission of the dissertation and graduation.

Acknowledgement

The project materialized through the support of the Department of Statistics and Operations Research of SMU in which it took place. We also thank the reviewers who gave feedback that was used to improve this paper. We also thank the editor for advises provided.

Funding Information

The project was funded through the Department of Statistics and Operations Research Fund of SMU.

Author’s Contributions

Tshepo Matsose: Conducted the study mainly for the MSc degree in the department.

Solly Matshonisa Seeletse: Supervised the study and provided exercises for experimental purposes.

Ethics

There will be no harm of emotional, physical or other form to any living organism that can because by the work in this study.

References

Asher, H., 1998. Polling and the Public: What Every Citizen Should Know. 1st Edn., Cq Press, ISBN-10: 0871874024, pp: 168.

Becker, C. and U. Gather, 2001. The largest nonidentifiable outlier: A comparison of multivariate simultaneous outlier identification rules. Comput. Stat. Data Analysis, 36: 119-127. DOI: 10.1016/S0167-9473(00)00032-3

Best, J., 2002. Damned Lies and Statistics: Untangling Numbers from the Media, Politicians and Activists. 1st Edn., University of California Press, California, ISBN-10: 0520228650, pp: 190.

Chambers, R., A. Hentges and X. Zhao. 2004. Robust automatic methods for outlier and error detection. J. Royal Stat. Society: Series A, 167: 323-339. DOI: 10.1111/j.1467-985X.2004.00748.x

Cousineau, D. and S. Chartier, 2010. Outliers detection and treatment: A review. Int. J. Psychol. Res., 3: 58-67.

Dawson, R., 2011. How significant is a boxplot outlier. J. Stat. Educ., 19: 1-12.

Dovoedo, Y.H., 2011. Contributions to outlier detection methods: Some theory and applications. Phd Thesis, University of Alabama, Tuscaloosa, USA.

Ercan, I., B. Yazici, Y. Yang, G. Ozkaya and S. Cangur *et al*., 2007. Misusages of statistics in medical researches. Eur. J. General Med., 4: 128-134.

Freedman, D.A., 2009. Statistical Models. London: Cambridge University Press.

Galbraith, J. and M. Stone, 2011. The abuse of regression in the National Health Service allocation formulae: response to the Department of Health's 2007 'resource allocation research paper'. J. Royal Stat. Society, Series A, 174: 517-528. DOI: 10.1111/j.1467-985X.2010.00700.x

Hair, J., 2008. Marketing Research. 4th Edn., McGraw Hill, London.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, 2005. Robust Statistics: The Approach Based on Influence Functions. 1st Edn., Wiley, New York, ISBN-10: 0471735779, pp: 536.

Jaffe, A.J. and H.F. Spirer, 1987. Misused Statistics: Straight Talk for Twisted Numbers. 1st Edn., M. Dekker, New York, ISBN-10: 0824776313, pp: 237.

Jaulin, L., 2010. Probabilistic set-membership approach for robust regression. J. Stat. Theory Pract., 4: 155-167. DOI: 10.1080/15598608.2010.10411978

Konishi, S. and G. Kitagawa, 2008. Information Criteria and Statistical Modeling. 1st Edn., Springer Science & Business Media, New York, ISBN-10: 0387718869, pp: 273.

Lekganyane, M.M., 2015. Statistical analysis of outliers and leverage points for CD4 counts in Dr. George Mukhari Academic Hospital. MSc Thesis, University of Limpopo. South Africa.

Liu, H., S. Shah and W. Jiang, 2004. On-line outlier detection and data cleaning. Comput. Chem. Eng., 28: 1635-1647. DOI: 10.1016/j.compchemeng.2004.01.009

Maddala, G.S., 1992. 'Outliers': Introduction to econometrics. 2nd Edn., MacMillan, New York.

Maier, M.H., 1999. The Data Game: Controversies in Social Science Statistics. 3rd Edn., Routledge, ISBN-10: 1315501929, pp: 320.

McKean, J.W., 2004. Robust analysis of linear models. Stat. Sci., 19: 562-570.

Nikoletseas, M.M., 2014. Statistics: Concepts and Examples. 1st Edn., Michael Nikoletseas, ISBN-10: 1500815683, pp: 236.

O'Neil, C. and R. Schutt, 2014. Doing Data Science: Straight Talk from the Frontline. 1st Edn., O'Reilly Media, Inc., O'Reilly, ISBN-10: 144936389X, pp: 408.

Paulos, J.A., 1988. Innumeracy: Mathematical Illiteracy and its Consequences. Farrar, Straus and Giroux, New York, ISBN-10: 0809074478, pp: 135.

Rousseeuw, P.J. and A.M. Leroy, (2003). Robust Regression and Outlier Detection. 1st Edn., John Wiley and Sons, ISBN-10: 0471488550, pp: 329.

Seife, C., 2011. Proofiness: How you're Being Fooled by the Numbers. 1st Edn., Penguin Publishing Group, New York, ISBN-10: 1101443502, pp: 320.

Sokolowski, J.A. and C.M. Banks, 2009. Principles of Modeling and Simulation: A Multidisciplinary Approach. 1st Edn., Wiley, Hoboken, N.J., ISBN-10: 0470289430, pp: 280.

Steele, J.M., 2005. Darrell Huff and fifty years of how to lie with statistics. Stat. Sci., 20: 205-209.

Stolz, M., 2002. The history of applied mathematics and the history of society. Synthese, 133: 43-57. DOI: 10.1023/A:1020823608217

Strutz, T., 2010. Data fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond. 1st Edn., Vieweg+Teubner Verlag, Wiesbaden, ISBN-10: 3834810223, pp: 244.

Tufte, E., 1997. The Visual Display of Quantitative Information. Graphics Press, Cheshire, Connecticut.

Zwane, P., 2015. Critique of misuse of statistics in some statistical applications. MSc Thesis, Sefako Makgatho Health Sciences University, Gauteng Province, South Africa.