International Journal of Research and Review
www.ijrrjournal.com E-ISSN: 2349-9788; P-ISSN: 2454-2237

Review Article

The Quest for Statistical Significance: Ignorance, Bias and Malpractice of Research Practitioners

Joshua Abah Abah

Department of Science Education
University of Agriculture, Makurdi, Nigeria

ABSTRACT

There is a growing body of evidence on the prevalence of ignorance, biases and malpractice among researchers which questions the authenticity, validity and integrity of the knowledge been propagated in professional circles. The push for academic relevance and career advancement have driven some research practitioners into committing gross misconduct in the form of innocent ignorance, sloppiness, malicious intent and outright fraud. These, among other concerns around research data handling and reporting, form the basis for this in-depth review. This discourse also draws attention to the recent official statement on the correct use of the p-value and the need for professional intervention is ensuring that the outcomes of research are neither erroneous nor misleading. The expositions in this review express cogent implications for institutions, supervisors, mentors, and editors to promote high ethical standards and rigor in scientific investigations.

Keywords: Research, Research misconduct, Bias, P-value, Statistical significance, ANCOVA Assumptions.

INTRODUCTION

Research is an enterprise aimed at finding solutions and answers to existing problems. Research can be seen as an objective, systematic, controlled and critical activity planned and directed towards the discovery and development of dependable knowledge (Emaikwu, 2012). Literally "re-search" means to "search again". It connotes patient study and scientific investigation wherein the researcher takes another, more careful look at data to discover all that can be known about the subject of the study (Bodla, 2017). Broadly, research entails bringing together some content that is of interests, some ideas that give meaning to that content and some techniques or procedures by means of which those ideas and content can be studied (Deshmukh, n.d.). According to O'Donnell (2012), research can be defined as the creation of new knowledge and/or the use of existing knowledge is a new and creative way so as to generate new concepts, methodologies and understandings. This could include synthesis and analysis of previous research to the extent that it leads to new and creative outcomes. From all indications, research can be described as an organized mechanism for studying phenomenon and testing hypotheses.

Research is an indispensable tool for growth and development in all fields of human endeavour. It has been a means of breaking forth into new frontiers in medicine agriculture, banking, education, food security, sociology, literature, arts and the sciences. Outcomes of diverse researches across different disciplines constitute the fuel for the present scientific and technological advancement the world is witnessing. The world today, being a "global village" is driven by the quest to know more, to venture into the unknown and make human existence much better than ever. As a result this significance of research, it is gradually becoming a sub-discipline in itself, within every discipline. This implies that within every field of study, there is a prescribed way of doing research, broadly referred to as "Research methodology".

Research methodology consists of learning how to adopt several common approaches when doing research, and how to conceive a research design (Jonker & Pennink, 2010). Methodology is a systematic plan for thinking and acting in the conduct of research work. Emaikwu (2012) maintains that scientific research methods must be verifiable, cumulative, ethical, theoretical and empirical. How well a research project is planned and how well the steps in the plan are integrated can make the difference between success or failure. In this respect, a plan consists of two general areas, namely research concepts and context, and research logistics (Congdon & Dunham, 1999), which are coordinated within a given time frame, culminating in the writing of a research report. The research report is the output of the entire research process made visible to a targeted audience and/or the public. For academics and researchers in universities, research centres, science laboratories and other research generating agencies, the production of quality and relevant research reports is a measure of growth and a determination of career and institutional relevance. Research reports are often published in professional journals, institutional bulletins, associations' notices and government agencies gazettes. They can also be presented at workshops, seminars and conferences, where learned contributions, corrections and suggestions can be synthesized into the research process before publishing for public use. Such rigorous vetting is essential considering the fact that a published work is expected to be an addition to existing knowledge and a reference point for future studies.

In light of the ripple effect of research in the knowledge-generation circle, researchers and academic institutions place serious emphasis on research ethics. In the words of Norris (1997):

> Research demands skepticism, commitment and detachment. To understand the object or domain of inquiry takes an intense degree of commitment and concentration. To remain open minded, alert to foreclosure and to sources of error needs some measure of detachment. As with other forms of art, research requires detachment from oneself, a willingness to look at the self and the way it influences the quality of data and reports; in particular research demands a capacity to accept and use criticism and to be self-critical in a constructive manner (p.173).

Ethical conduct, in general refers to actions that one takes pride in according to his or her conscience and that lives up to his or her responsibility as a member of society. Kim (2009) asserts that research ethics is a special social norm that researchers are obliged to abide by as criterion of judgment for researchers not to operate against their professional integrity and to carry out socially responsible research activities. Ethical standards are set by professional associations, educational institutions, journal publishers and government regulatory agencies. It is likely that these organizations vary considerably in the attention they invest and the procedures they deploy to uphold research ethics (Johnson, Parker & Clements, 2001). Practices carried out by researchers outside these regulatory guidelines constitute research misconduct.

By definition, research misconduct entails fabrication, falsification or plagiarism in proposing, performing or reviewing research or in reporting research results (OSTP, 2002). Research misconduct may occur if the conduct represents a significant departure from accepted practices; has been committed intentionally, knowingly or recklessly and can be proven by a preponderance of evidence (Inzana, 2008). The ramification of research misconduct has been broaden to include other serious deviation from accepted guidelines of the scientific community for maintaining the integrity of research record and retaliation of any kind against a person who reported or provided information about suspected or alleged misconduct and who has not acted in bad faith (Fisehen, n.d.). Among the three "cardinal sins" of research conduct, only plagiarism seems to be in the public eye, with the other two (falsification and fabrication) completely reduced to bare whispers. Falsification is the changing or omission of research results (data) to support claims, hypotheses and other data. Falsification can include the manipulation of research instrumentation, materials, processes, images or representation in a manner that distorts the data or "read too much between the lines" (Schienke, 2017). On the other hand, fabrication is the construction or addition of data, observations or characterizations that never occurred in the gathering of data or running of experiments. According to Schienke (2017), fabrication can occur when "filling out" the rest of the experiment runs and where claims are made based on incomplete or assumed results.

Kim (2009) explains why academics hardly raise their voice when discussing research ethics:

> One of the biggest reasons for past negligence of research ethics is believed to be the public confidence in scientist or the confidence among researchers in the self-control system. As quantitative assessment of researchers becomes widespread and the commercial application of science and technology is growingly emphasized, we can no longer rely merely on the value-neutral and reasonable inclinations of scientists and the self-correcting system in science circles. Therefore, it is essential for us to contemplate what responsible conduct of research actually entails and fully establish research ethics as an integral part of our academic culture (p.1).

The pressure on academics to increase their number of publications in line with requirements for promotion and career growth has also contributed to this grave concern for research ethics. In the view of Mullane and Williams (2013), bias in research, where prejudice or selectivity introduces a deviation in outcome beyond chance, is a growing problem, probably amplified by "first to publish" and "publish or perish" drive and more recently, the monetization of science for personal gain. The matter is made worst by student researchers who often do not have the depth of experience and tenacity to match with the scope of some sensitive research areas. The practice of polishing some of these students' "shallow" findings for publications without rigorous checks by supervisors is in itself an assault on quality. The outcome of such practice is the proliferation of ignorance, personal biases and malpractice in the name of research. The current mess being made of statistical approaches and unsubstantiated significant results assembled by so called "research analysts" which are difficult to decipher constitute a major cause for worry among the few who are still interested in classical statistical methods.

The problem under consideration is a widespread one and not unique to any specific field of practice. This implies that the emphasis on integrity and quality that is intended in this work may not be very useful if restricted, for instance, to mathematics education. Thus, a multidisciplinary approach is adopted here, drawing on in-depth background in mathematical statistics and modern statistical computing. The role of statistical analysis in research is first presented. This is followed by discussions on ignorance, bias and malpractice among research practitioners. By "research practitioners" this discourse implies all stakeholders involved in the process of producing research reports, including the researcher, the supervisors (where applicable), the data analyst and vetting authorities. A final section of this essay focuses on the place of professional intervention in improving the integrity of research works.

The Role of Statistical Analysis in Research

In order to investigate phenomenon, researchers need to gather information about the phenomenon in a planned manner. Such investigations lead to the generation of research data. Data itself is the collected factual material commonly accepted in the scientific community as necessary to validate research outcomes. Research data is data that is collected, observed or created, for purposes of analysis to produce original research results (Boston University Libraries, n.d.). Research data is often obtained in raw form and require statistics to bring out its essence and interpretation. Emaikwu (2012) provides a robust background definition of statistics:

> Statistics is a branch of mathematics which deals with the collection, classification, analysis and interpretation of numerical data. It deals with quantitative analysis of numerical data so as to make wise decision. Statistics helps in arriving at empirically verifiable research and possible replication of such information by other researchers (p. 89).

Statistics can also be seen as a collection of methods for planning experiments, obtaining data and then organizing summarizing, presenting, analyzing, interpreting and drawing conclusions based on the data (Deshmukh, n. d.). Statistical analysis facilitates comparison, exposes relationships between phenomena and returns meaning to raw research data for inferential purpose. There exist a wide range of statistical tools for the analysis of research data depending on the design adopted for the research. Broadly, available tools can be classified as either parametric statistics or non-parametric statistics. Likewise, several descriptive statistical tools can be used to augment inference by presenting information in simple and understandable format. In fact statistics can be said to be the language of research. But that is not to say that a mere quantitative results can prove anything if the application of statistical methods is handled wrongly.

When one makes a statistical inference, namely, an inference which goes beyond the information contained in a set of data, one must always proceed with caution. In the view of Miller and Freund (1977), one must decide carefully how far one can go in generalizing from a given set of data, whether such generalizations are at all reasonable or justifiable, whether it might be wise to wait until there are more data and so forth. The roots of statistical inference are the appraisal of the risks and the consequences to which one might be exposed by making generalizations from sample data. This includes an appraisal of the probabilities of making wrong decisions, the chances of making incorrect predictions and the possibility of obtaining estimates which do not lie within permissible limits.

What is drivable from the history of statistical inference is the carefulness and nobility required of the statistician in the drawing up of conclusions based on research data. The weight of statistical conclusions drives the delicate job of the analyst who must deploy his expertise and use tools correctly without bias. According to Emaikwu (2012), the misuse of statistics will arise from the following situations:

i. Analysis without any definite purpose
ii. Carelessness in the collection and interpretation of data
iii. Misleading others for self-interest and cooking up of data
iv. Pressure on statisticians and bias and prejudice of the statisticians

v. Wrong definitions, inadequate data, wrong methods and in appropriate comparison.

It can thus be summarized that if a problem can be properly formulated and measurement data can be generated, whether it arises in physical, biological, and social sciences or any other discipline, statistical tools can be designed to provide a scientific solution (Chakrabarty, 2012). Thus, it is widely recognized that the proper use of statistics s a key element of scientific enquiry. According to Chakrabarty (2012), quality and integrity of data is the most important element in the success and utility of statistics.

Ignorance of Research Practitioners

Some of the commonly observed misuse of statistics by research practitioners arises out of shear ignorance and misunderstanding of statistical approaches and tools. This realization is being compounded by the misuse of modern statistical software packages by untrained "statistical analysts" who are better computer operators than the "label" they carry in the deployment of their exploitative merchandise. These so-called "analysts" feed off the ignorance of their clients and churn out incompatible statistics that cannot be rightly interpreted. This kind of misuse of statistics can be viewed as negligence or deficits of competence since it arise as a result of lack of depth on the part of the researchers on whom the responsibility for such research work lie. Inexperienced researchers generally tend to abuse statistics via bad samples, small samples, loaded questions, misleading graphs, pictographs, precise numbers, distorted percentages, partial pictures and distortions (Deshmukh, n. d.). With preordained intentions, it is easy to get any conclusions out of any given research data. Other common method ignorance that can seriously hamper the outcome of statistical analysis is given an extensive coverage in Podsakoff, Mackenzie, Lee and Podsakoff (2003).

The shortcomings arising out of the ignorance of research practitioners are mostly thrown up when numbers that are anecdotal and not generalizable are reported in cumulative form. In an effort at tracking such misuse in the public domain of information security research, Ryan and Jefferson (2003) reported that:

> What is lost in the stories of these various research efforts is the nuances and subtleties of the research methodologies used, the statistics applied and the data reported … in many cases, the research methodologies were not sound (in some cases, the results were specifically identified as being unscientific). The statistical analyses were in some cases inappropriate and in general only partial results reported in the press (as might be expected).

When statistical procedures which can produce very accurate results are often used in manners for which they are not intended they produce erroneous and misleading results. Graham (2001) identifies the concentration of misuse of statistics in null hypotheses significance testing (NHST), ignoring of assumptions, and handling of ANOVA interaction effects. For statistical procedures that depend heavily on specific assumptions about the distribution of the sample, ignorance displayed in departures from these assumptions can be misleading. The over dependence on distribution-dependent statistical methodologies is definitely increasing the tendency to misapply statistics in research. Such encumbrance can be avoided if research practitioners exhibit their freedom to choose statistical approaches they deeply understood. Nearly all classical general linear models (GLM) requires that the assumptions of normality of distribution, homogeneity of variance and random samples be met, but where it is difficult to test assumptions, non-parametric alternatives are conveniently available to drive substantial inference making.

Another indication of statistical ignorance among research practitioners is the tendency to mistaken correlation for causation (false causality). Correlation is just a linear association between two variables, meaning that as one variable rises or falls the other variables rises or falls as well. This association may be positive, in which case both variables consistently rise, or negative, in which case one variable consistently decreases as the other rises (Martz,2013). Even a correlation of +1 still does not imply causality, since the correlation coefficient only measures linear relationships. Martz (2013) observes that a meaningful non-linear relationship may exist even if the correlation coefficient is 0. Additionally, because the Pearson Correlation Coefficient can be very sensitive to outlying observation it can be highly susceptible to sample selection biases. It is also a misguided analysis to use correlation to measure agreement.

ANCOVA: Still a Delicate Instrument

Analysis of Covariance (ANCOVA), dubbed a delicate instrument by Janet Elashoff, is still delicate. Carefully handled, though, it is an excellent device for the analyst's tool kit (Owen & Froman, 1998). The professional usage of this powerful statistical procedure continues to litter the field of research methodology with various pitfalls that can deliver misleading results for the unwary analyst.

Analysis of Covariance (ANCOVA) is a combination of analysis of variance (ANOVA) and regression analysis or indeed, a more complex extension of both (Emaikwu, 2012). The ANCOVA procedure involves measuring one or more concomitant variables (also called covariates) in addition to the dependent variable (Kirk, 1982). The concomitant variable represents a source of variation that has not been controlled in the experiment and one that is believed to affect the dependent variable. ANCOVA serves two primary purposes: (a) to improve the power of a statistical analysis by reducing error variances and (b) to statistically equate comparison groups (Owen & Froman, 1998). Experimental error can be reduced if a portion of the error variance $\hat{\sigma}^2$ associated with the dependent variable is predictable from a previous knowledge of the concomitant variable. Kirk (1982) observes that removing this predictable portion from $\hat{\sigma}_E^2$ results in a smaller error variance, and, hence, a more powerful test of a false null hypothesis.

As robust as the ANCOVA Procedure is, ignorance of the developmental history and techniques of the analysis on the part of researchers and analysts is on the increase. Even amongst the standard descriptions of ANCOVA assumptions and tests are some ambiguous and subtly misleading accounts. In this respect, Rutherford (2001) observes that it is important to distinguish genuine statistical assumptions from the made to simplify ANCOVA interpretation to test the appropriate statistical assumptions and to employ pertinent techniques to assess the tenability of these assumptions. In addition to all ANOVA assumptions, traditional ANCOVA is based on three specific assumptions, namely:

i. The covariance is independent of the treatments
ii. In each treatment group the relationship between the covariance and the dependent variable is linear (the covariate and dependent variable are expressed at the first power only),
iii. The regression coefficients of the dependent variable on the covariate in each treatment group are homogenous. (Rutherford, 2001 p. 126).

To clarify, the first statistical assumption is that the covariate(s) is (are) uncorrelated with other independent variables. In an example provided by Owen and Froman (1998), in comparing lung vital capacity in smokers and non-smokers, one may ask if the selected confounding variable, age, correlated with the independent variable, smoking? If the correlation is non-zero, then removing the variance associated with age will also remove some of the variance associated with the grouping variable (smoking) in effect leaving less of the dependent variable's (lung vital capacity) variance to be accounted for the independent variable (smoking) (Owen & Froman, 1998). Evidently, analysis of covariance is not appropriate unless the effects eliminated by covariate adjustment are irrelevant to the objectives of the experiment or study (Kirk, 1982).

The second specific assumption of traditional ANCOVA is also known as the linearity assumption. In basic terms, this assumption states that the regression of the dependent variable on the covariate(s) in each of the experimental conditions is linear. Rutherford (2001) holds that the most obvious way to assess linearity of the separate groups' regressions is to plot the dependent variable against the covariate (or each covariate) for each experimental condition. Regression linearity can also be checked through a significant test for the reduction in errors due to the inclusion of non-linear components, applying a form of power transformation (e.g. quadratic, cubic) to the covariate before the ANCOVA analysis (Owen & Froman, 1998).

The third statistical assumption of traditional ANCOVA is the one mostly ignored or wrongly handled by research practitioners. If there is a positive relationship between covariate and the outcome (dependent variable) in one group, we assume that there is a positive relationship in all of the other groups too. If however the relationship between the outcome and covariate differs across the groups then the overall regression model is inaccurate. Field (2012) observes that the best way to think of this assumption of homogeneity of regression slopes is to imagine plotting a scatterplot for each experimental condition with the covariate on one axis and the outcome on the other. The regression lines for each of the scatter plots should look more or less the same. This feature, according to Rutherfield (2001), becomes more tenuous as the number of experimental conditions increases. The reason for the assumption is that all groups' dependent variable scores are adjusted based on a pooled regression slope, if the groups individual slopes differ sharply, then the pooling becomes a muddy average (Owen & Froman, 1998). Kirk (1982, pp 732-734) provides a demonstration of a statistical test for homogeneity of regression models. Likewise, Rutherford (2001, chapter 8) gives a comprehensive coverage of heterogeneous regression ANCOVA using more sophisticated GLMs.

Additional requirements for ANCOVA contain a provision for measuring the covariate without error, an often unmentioned assumption in statistics books. Owen and Froman (1998) mention that in the case of ANCOVA with random assignment, covariate measurement error does not bias the adjusted means, but it does produce less statistical power, which in turn increases the probability of Type II error. With a quasi-experimental design lacking random assignment, covariate measurement error creates bias in adjusted means. Quasi-experimental designs, common in educational and industrial research, usually employ intact groups because it is often impractical for administrative reasons to randomly assign treatments. With respect to the use of intact groups, Kirk (1982) gives this note of caution:

> Experiments of this type are always subject to interpretation difficulties that are not present when random assignment is used in forming the experimental groups. Even when analysis of covariance is skillfully used, we can never be certain that some variable that has been overlooked will not bias the evaluation of an experiment. This problem is absent in properly

> randomized experiments because the effects of all uncontrolled variables are distributed among the groups in such a way that they can be taken into account in the test of significance. The use of intact groups removes this safeguard (p. 718).

In line with this warning, Pedhazur (1994 in Owen & Froman 1998) affirms that unfortunately, applications of ANCOVA in quasi-experimental and non-experimental researches are by and large not valid. This is because the F-ratio in ANOVA/ANCOVA is predicated on the pre-condition that observations are random samples drawn from normally distributed populations. Random assignment is used to distribute the idiosyncratic characteristics of subjects over the treatment levels so that they will not selectively bias the outcome of the experiment (Kirk, 1982). Non-randomization leads to non-independence of errors which seriously affects both the level of significance and the power of the F-test.

As a rule, statistical packages encourage users to ignore assumptions and leap right in the main analysis. Owen and Froman (1998) note that inside ANOVA/ANCOVA programs, packages offer the Levine test for homogeneity of variance, but any other tests of assumptions must be arranged by the user.

The misapplication of ANCOVA often begins from the design phase of most research work, particularly in the identification of the concomitant variable. Many works in education that employs ANCOVA tend to, as a rule, use pretest scores of learning ability as a covariate without concern that other concomitant variables may have been overlooked, such as number hours spent in study by students in different intact classes, peculiar historical background of the subjects of the study and other intermittent factors. Many research practitioners are virtually unaware that effects eliminated by a covariance adjustment must be irrelevant to the objectives of the experiment. In addition to meeting original ANCOVA assumptions, the following conditions guide the selection of concomitant variables:

i. The experiment contains one or more extraneous sources of variation believed to affect the dependent variable and considered irrelevant to the objectives of the experiment.
ii. Experimental control of the extraneous sources of variations is either not possible or not feasible.
iii. It is possible to obtain a measure of the extraneous variations that does not include effects attributable to the treatment. (Kirk 1982, p 719).

To improve the quality of ANCOVA studies, Owen and Froman (1998) recommend that the method be limited primarily to randomized designs. When the analyst wants to use ANCOVA with an intact group or other non-random assignments, the correlation between the covariate(s) and the independent variable(s) should be reported. As the correlations are increasingly non-zero, then conclusions drawn about the independent variables are increasingly suspicious (Owen & Froman, 1998). Weaver (2002) reported a vital warning thus:

> ANCOVA can often accomplish the purpose of increasing power but its ability to remove bias is fraught with technical difficulties that have been frequently ignored. Many novices have viewed ANCOVA as the “messiah” of statistical methods it has been asked to “give signs” and “perform wonders” - to reveal the truth amidst a bewildering array of uncontrolled and poorly measured confounding variables. Some have mistakenly assumed that ANCOVA, in effect transforms quasi-experiments (i.e. studies in which subjects are not randomly assigned to treatments but taken as they occurred naturally) into randomized experiments. In reality ANCOVA is

unable to give the results of a quasi-experiment the same degree of credibility provided by randomized experiments (p 20).

In view of the technical frailty of ANCOVA, suitable alternatives can be deployed. For instance, it can be far more informative, following a violation of homogenous slope, to calculate Johnson Neyman regions of significance. This technique, according to Owen and Froman (1998) helps to map out where groups do and do not differ along various values of the covariate. Weaver (2002) recommends the Treatment x Blocks design as a robust alternative to ANCOVA. The Treatment x Blocks design does not have restrictive assumptions and, for this reason, is to be preferred for its relative freedom from statistical assumptions underlying the data analysis (Keppet, 1982 in Weaver, 2002). This later design is sensitive to any type of relationship between treatments and blocks-not just linear.

The P-Value Controversy

Of all the areas of misuse of statistical procedures, none has stir up more controversy than the issue of the p-value. The wrong use of p-values permeates even the highest level of research and has eaten so deep into the fabric of research methodology textbooks that many are unwilling to let go. This stubbornness among some research practitioners has forced the most revered American Statistical Association (ASA) to issue a statement on the guiding principles of the use of the p-value. The statement officially released on 8th March, 2016 is the first time that the 177-year old ASA has made explicit recommendations on such a foundational matter (Baker, 2016). Before stating these guidelines here, a clearer view of the historical origins of this controversy may be necessary and educative.

As a way of definition, the p-value is a measure of discrepancy of the fit of a model or “null hypothesis” H to data *y*, mathematically defined as $\Pr(T(y^{rep}) > T(y)$ Given H$)$, where y^{rep} represents a hypothetical replication under the null hypothesis and T is a test statistic (i.e. a summary of the data perhaps tailored to be sensitive to departures of interest from the model) (Gelmen, 2013). Informally, a p-value is the probability under a specified statistical model that a statistical summary of the data (e.g. the sample mean difference between two compared groups) would be equal to or more extreme than its observed value (Wasserstein & Lazar, 2016). The p-value answers the question: If the null hypothesis had been true, what would have been the probability of obtaining data that looked as or more inconsistent with it than the data we observed in our sample? So the smaller is the p-value, the greater is the doubt that our data sheds on the null hypothesis.

In referring to the roots of NHST, Hubbard and Bayarri (2003) assert that classical statistical testing is an anonymous hybrid of the competing and frequently contradictory approaches by R.A. Fisher on the one hand, and Jerzy Neyman and Egon Pearson on the other. The ignoble p-value controversy is a widespread failure to appreciate the incompatibility of Fisher’s evidential p-value with the Type 1 error rate, α, of Neyman-Pearson statistical orthodoxy. This misuse reflects the fundamental differences between Fisher’s ideas of significance testing and inductive inference, and Neyman-Pearson views of hypothesis testing and inductive behaviour (Hubbard & Bayarri, 2003). A trip back to the very beginning of the methods of statistical inference is what most applied researchers require.

Fisher’s views on significance testing, presented in his research papers and in various editions of his enormously influential texts, Statistical Methods for Research Workers (1925) and The Design of Experiments (1935), took root among applied researchers (Hubbard & Bayarri, 2003). At the heart of his conception of inductive inference is what Fisher called the *null hypothesis*, H_o. Fisher was convinced that it is possible to argue from consequences to causes, from observation to hypothesis. Fisher significance test is defined as a procedure for establishing the probability of an outcome, as well as more extreme ones, on a null hypothesis of no effect or relationship. Hubbard and Bayarri (2003) assert that the distinction between the probability of the observed data given the null and the probability of the observed and more extreme data given the null is crucial; not only it has contributed to the confusion between *p's* and *α's*, but also results in an exaggeration of the evidence against the null provided by the observed data. Fisher regarded p-values as constituting inductive evidence against the null hypothesis that a sample comes from a hypothetical infinite population with a known sampling distribution. The null hypothesis is said to be disproved or rejected if the sample estimate deviate from the mean of the sampling distribution by more than a specified criterion, the level of significance (α).

On the contrary, the Neyman-Pearson approach (developed as an attempt to improve on Fisher's approach) formulates two competing hypotheses, the null hypothesis (H_o) and the alternative hypothesis (H_A). This framework introduced the probabilities of committing two kinds of error based on considerations regarding the decision criterion, sample size and effect size. These errors are false rejection (Type I error) and false acceptance (Type II error) of the null hypothesis. Type I error was designated α (level of significance) while Type II error was called β. Hubbard and Bayarri (2003) report that in contradiction to Fisher's ideas about hypothetical infinite populations, Neyman-Pearson results are predicated on the assumption of repeated random sampling from a defined population with α being the long-run frequency of Type I errors. With respect to this distinction, associated p-value (significance probability) determined in a statistical test cannot be interpreted as a frequency-based Type I error rate and it is incorrect to take $p<\alpha$ as a measure of evidence against H_o. Accordingly a p-value for Fishers represented an "objective" way for researchers to assess the plausibility of the null hypothesis.

> … the felling induced by a test of significance has an objective basis in that the probability statement on which it is based is a fact communicable to and verifiable by other rational minds. The level of significance in such cases fulfills the conditions of a measure of the rational grounds for the disbelief (in the null hypothesis) it engenders (Fisher 1959, p.43 in Hubbard & Bayarri, 2003).

Consequently, the tag "$p< 0.05$" and researchers quest for publishable statistical significance is a psychological practice in itself. According to Ludwig (2005) such quest only psychologically makes research practitioners feel good and fuel the wrong belief that the observed results of an experiment or observational study are not factual and therefore cannot be discussed unless some type of statistical sanctification is invoked. By going back to the roots of p-values, it is obvious that researchers are not to rely on p-values to make their case since literally "Fisher considered the use of probability values to be more reliable than, say, eyeballing results" (Hubbard & Bayarri, 2003 p.4). This appears to be the thoughts re-echoed recently by the ASA's official statement on the use of p values. The much be-lated statement came as a response to apparent editorial biases against scientifically important works that get relegated on the basis of non-significant p-values. The pursuit of the arbitrary threshold ($p < 0.05$) has also led to data dredging and diverse forms of misconduct that emphasize the search for small p-values over other statistical and scientific reasoning. Such quests tend to ignore many other, more appropriate statistical tools like graphic analysis, regression trees, bioinformatics, data mining and exploratory data analysis (Ludwig, 2005).

The American Statistical Association six principles, many of which address misconceptions and misuse of the p-value are the following:

i. P-values can indicate how incompatible the data are with a specified statistical model,
ii. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
iii. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
iv. Proper inference requires full reporting and transparency.
v. A p-value or statistical significance, does not measure the size of an effect or the importance of a result.
vi. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis. (Wassertein & Lazar, 2016).

In further explanations provided by Yaddanapudi (2016) on Principle 5, for instance, it is obvious a p-value of 0.01 does not mean that the effect size is larger than with a p-value of 0.03. With a particular example, Yaddanapudi showed that the p-value would have been 0.000002 if the sample were to be increased from 200 to 1000. The conclusion of Wassertein and Lazar (2016) on ASA's statement is noteworthy:

> Good statistical practice, as an essential component of scientific practice, emphasizes principles of good study design and conduct, a variety of numerical and graphical summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting and proper logical and quantitative understanding of what data summaries mean. No single index should substitute for scientific reasoning (p.132).

In particular, even in a designed experiment, statistical tests and p-values give very little information because they can answer only the one very specific question (Ludwig, 2005).

Bias of Research Practitioners

Research is a procedural activity that is thought to be sanctimonious over ordinary observation or judgment. It is an investigation that is valid and present truth claims in the form of statements of fact, descriptions, accounts, propositions, generalizations, inferences, interpretations, judgments and arguments (Norris, 1997). Being a scientific endavour, research is traditionally conducted around the four norms of science (articulated by Robert Merton in 1973). These are universalism, communalism, disinterestedness and organized skepticism. MacCoun (1998) elaborates:

> Universalism stipulates that scientific accomplishment must be judged by improved criteria; the personal attributes of the investigator are irrelevant. Communalism requires scientific information to be publicly shared. Disinterestedness admonishes investigators to proceed objectively, putting a side personal biases and prejudices. Finally, organized skepticism requires the scientific community to hold new findings to strict levels of scrutiny through peer review, replication and the testing of rival hypotheses (p.120).

These normative pillars crudely constitute a culture of appraisal of research work by both scientists and non-scientists alike. But research, whether quantitative or qualitative, experimental or naturalistic, is a human activity subject to the same kinds of failings as other human activities (Norris, 1997). Seasonal research experts know that researchers are fallible and that bias can find its way into any research programme (Sarniak, 2015).

Bias is an expression of unfair influence on the wholeness of an activity. In research, bias occurs when there is systematic difference between the results from a study and the true state of affairs (Sabin, 2010). It is the tendency to be partial which happens when the researcher does something that favours or skews towards a certain direction, leading to research outcomes that is inaccurate and unreliable (Regoniel, 2013). The worry about subjectivity arises particularly because the data obtained in a research must "go through" the researcher's mind before it is put on paper (Rajendran, 2001). MacCoun (1998) reports that the very decision to study certain topics is sufficient to prompt some observers to infer that the investigator is biased. In this respect, it is always possible that the bias lies in the accuser rather than (or in addition to) the accuser.

The existence of bias in research tends to be observed by the sheer volume of data reported. Data is generally viewed as a key basis of competition, productivity, growth and innovation, irrespective of its conception, quality, reproducibility and usability (Mullane & Williams, 2013). Sabin (2010) notes that bias is often introduced when a study is being designed, but can be introduced at any stage. In view of this, it is preferable to design the study in order to avoid bias in the first place. Bias by design reflects in critical features of experimental planning ranging from the design of an experiment to support rather than refute a hypothesis, lack of consideration of the null hypothesis, failure to incorporate appropriate control and reference standards, and reliance on single data points (Mullane & William, 2013). Selection bias and information bias may also arise from measurement, misclassification, observation, regression, dilution and missing data, all of them being inadequacies that point to a hasty study design. But of all biases, personal, fraudulent bias of the researcher is the most dreaded.

Researchers are an inherently optimistic group who are constantly tempted by the tendency for over-statement and over simplification (Mullane & Williams, 2013). Many of those who conduct research fail to do good research because they want to do it at their convenience. For instance, instead of getting a random sample of respondents, a researcher may just interview anyone that gets in his way, thereby introducing a selection bias (Regoniel, 2013). Likewise, while the nature of one's research may be argumentative, favouring a preconceived position on the subject of investigation can bias the outcomes. Some researchers fall for the tendency to steer the results of their studies to the direction they want, sometimes "p-hacking" their data analysis to yield statistically significant results or indulging in selective reporting. According to Mullane and Williams (2013) the retrospective selection of data for publication can be influenced by prevailing wisdom promoting expectations, or, where the benefit of hind-sight at the conclusion of a study allows an uncomplicated sequence of events to be traced and promulgated, as the only conclusion possible.

Research practitioners who deliberately promulgate research findings out of their biases fail to acknowledge that research findings are rarely a direct determinant of policy decisions. Social scientists are sometimes strikingly naïve about the gaps between research findings and the inputs needed for sound policy formation (MacCoun, 1998). For instance, a research work that manipulates its way out to establish a significant outcome in favour of a non-contextual and inadequately available technology will not necessarily contribute to the expected wide adoption, since it failed to acknowledge the extant context and the possibility of implementation of such technology. The hypothetical researcher in this example commits a confirmation bias when he forms a hypothesis or belief and uses respondents' information to confirm that belief. He judges and weighs responses that confirm his hypothesis as relevant and reliable, while dismissing evidence that does not support the hypothesis. Confirmation bias is deeply seated in the natural tendencies people use to understand and filter information which often lead to focusing on one hypothesis at a time (Sarniak, 2015).

One of the personal biases that can dent the validity of a research work may stem from the cultural perspective of the researcher. Assumptions about motivations and influences that are based on one's cultural lens (on the spectrum of ethnocentricity or cultural relativity) create the culture bias (Sarniak, 2015). Broadly, while ethnocentrism is judging another culture solely by the values and standards of one's own culture, cultural relativism is the principle that an individual's beliefs and activities should be understood by others in terms of that individual's own culture. Sarniak (2015) suggested that although complete cultural relativism is never 100 percent achievable, researchers must move toward cultural relativism by showing unconditional positive regard and being cognizant of their own cultural assumptions. The data must bear the weight of any interpretation, so the researcher must constantly confront his or her own opinions and prejudices with the data (Rajendran, 2001). If the worth of a study is the degree to which it generates theory, description or understanding, then researchers must constantly view the threat of personal bias with a grave concern. Mullane and Williams (2013) express the expanding concerns regarding scientific integrity and transparency in the following terms:

> While research misconduct in terms of overt fraud and plagiarism is a topic with high public visibility, it remains relatively rare in research publications why data manipulation, data selection and other forms of bias are increasingly prevalent. Whether intentional, the result of inadequate training or due to lack of attention to quality controls, they foster an approach and attitudes that blurs the distinction between necessary scientific rigor and deception.

Malpractice of Research Practitioners

For centuries knowledge meant proven knowledge, proven either by the power of the intellect or by the evidence of the senses. Wisdom and intellectual integrity demands that one must desist from unproven utterances and minimize, even in thought, the gap between speculation and established knowledge (Lakatos, 1970). Inherently, there are certain important values shared by genuine researchers. The foremost of these values are integrity, accuracy, efficiency and objectivity. Integrity simply refers to the ability to deliver information as it is and respect promise made. Accuracy ensures the report of research results as they are and the assurance to avoid errors. Efficiency is ability to utilize resources wisely and avoid wasting them. Objectivity is the readiness to embrace facts as they are and refrain from biases. Misconduct or malpractice results from the gross departure from these and other shared values. Malpractice in this sense is the deliberate or repeated non-compliance with research requirements (Lepay, 2008).

Malpractice by research practitioners could be attributed to innocent ignorance, sloppiness and malicious intent (falsification or fraud). With respect to fraudulent practices by researchers, Simmons, Mercer, Schwarzer and Courtney (2016) maintain that:

> Concern about data falsification is as old as the profession f public opinion polling. However, the extent of data falsification is difficult to quantity and not well documented. As a result, the impact of falsification on statistical estimates is essentially unknown (p.1).

Falsification occurs when researchers go against the code of ethics on the maintenance and preservation of research data. This ethical standard requires research practitioners to record data, samples and other materials used or generated throughout the course of research and retain them for a given period of time (Kim, 2009). Any preconceived influence forced upon the interview process and data compilation amounts to malpractice. Filling out of missing data and partial coverage of study area are becoming prevalent. There are also unconfirmed tales of "assumed research"- study reports cooked up from the imagination of inadvertent authors. Meta-data are generated in a day and questions and hypotheses are succinctly handled to support the perspective of these rogue authors. The thought of the possibility of such sacrilege even abhors but there are people who condescend so low to this abysmal level of malpractice.

Another common form of malpractice could be seen in the cloning of results for unreachable sample units in experimental and quasi-experimental studies. After a robust presentation of research methodology at the proposal stage, some researchers fear that they may not be able to reach the planned sample units. For instance, the rigor of setting up the treatment for the experimental group across all aforementioned sample units has been a tempting factor for many educational researchers, particularly when the study entails the deployment of delicate technology and complex pedagogical sequence. For surveys, there have been documented reports of duplication of data sets (Simons et al, 2016). In the natural sciences, food technology and agriculture, the practice of posting samples to "specialized laboratories" for analysis (in the absence of the researcher) raises suspicions for the outcomes. Sometimes such "results-by-correspondence" arrive in non-interpreted formats leaving the researcher to whimsically infer any outcome of choice. There are undocumented instances involving research students who cannot explain the mechanisms of their laboratory and statistical analysis, obviously because they were not involved in those stages in the first place.

Research data may be termed "falsified" in the following ways:

i. Creating data that were never obtained
ii. Altering data that were obtained by substituting different data
iii. Recording or obtaining data from a specimen, sample or test whose origin is not accurately described or in a way that does not accurately reflect the data.
iv. Omitting data that were obtained and originally would be recorded. (Lepay, 2008)

Malpractice in research is a serious offence in many climes and should be eschewed by all well-meaning researchers. Instead of the usual institutional cover-up of professional misconduct of researchers, efforts must be geared towards prevention, retraining and possibly, open rebuke or reprimand for those found wanting.

The Place of Professional Intervention

It is obvious that relatively little attention is given, at least publicly, to the contrasting problem of data falsification and other malpractices by survey staff and researchers in general (Johnson, Parker & Clements, 2001). That the misuse of statistical procedures has continued for so long does not excuse its existence (Graham, 2001). It is the responsibility of every profession to ensure that the results of their research are neither erroneous nor misleading. The weight of the consequences of malpractice such as possible safety risk, jeopardizing of the reliability of published data, undermining of regulatory authorities, decreasing public confidence and the risk of putting people of questionable character in respectable positions they did not actually merit, must be projected at all times by professional bodies (Lepay, 2008). Professional associations must intervene by making clear their position on ethical misconduct. When self-scrutiny fails, the onus falls on institutional safeguards such as peer reviewing, research replication, meta-analysis and expert panels to mitigate the onslaught on professional misconduct in research (MacCoun, 1998). For the practice of survey research, John, Parker and Clement (2001) suggest:

> Expectations and consequences of falsification should be clear and acknowledged, and it should be clear to staff what the general procedures for monitoring staff performance include. Further, all staff responsible for the collection and/or processing of survey data are asked to sign a statement indicating their awareness and understanding of the policies relevant to data falsification. Careful supervision of interviewer and data coding staff is critical to the prevention of data falsification (p.277).

Faculties and professional bodies can deploy available detection methods to help in evaluating the performance of the costly prevention methods and to identify falsified results that slipped past prevention measures. Detection methods entail evaluation of key indicators, including para-data (interview length, time stamps, geocoding, timing of interviews), interviewer-related data (experience, daily workload, success rates) and interview-related data (characteristics of respondents, interview recordings, back-checking results) as well as analysis of the structure of responses (refusals, extreme values, coherence of responses, consistency in time series, duplicates) (Simons et al, 2016). Identification of falsified data is not the result of a single measure, but an assessment of the different aspects within the study-specific environment in which research practitioners carry out their work. Institutional review boards of educational institutions are expected to evaluate their students' methodological competence and investigate the applicability of all types of statistical analysis across all applications and on the basis of a cost-utility analysis (Graham, 2001). In practical terms, institutional review boards must set up mechanisms for cross-checking the authenticity of field data. Such mechanism might entail the on-site supervision of field work and administrative collaboration between faculties and authorities of partnering institutions from where research students obtain primary and secondary data. Attestation from partnering institutions on the extent of work done in their premises by research students will go a long way in raising quality assurance of graduate research works.

With the increase in bias, data manipulation and fraud, the role of the professional journal editor has become more challenging, both from a time perspective and with regards to avoiding peer review bias (Mullane & Williams, 2013). While keeping standards high, much of the process of producing quality research reports still depends on the integrity and ethics of authors and their institutions. Mullane and Williams (2013) assert that it is paramount that institutions, mentors and researchers promote high ethical standards, rigor in scientific thought and ongoing evaluations of transparency and performance that meet exacting guidelines. Institutions and the research community must ensure that allegations of research malpractice proven by a preponderance of the evidence (Inzana, 2008). According to Fischer (n. d.), common features of research policy and regulation with respect to handling misconduct issues include:

i. Discrete, separate phases of inquiry, investigation, adjudication and appeal
ii. Reliance on community based standards ("serious deviation" or "significant departure").
iii. Partnership with institutions
iv. Level of intent and standard of proof
v. Confidentiality for subjects and informants

vi. Fair, accurate, timely, fact-and document -based process (p.4).

The time to act is now. Voices from within the research community must rise, loud and clear, in unison and defense of our noble professions. People are encouraged to put aside their silence and secret whispers in order to push for the right things to be done at all times. It is ripe to correct the notion of not washing the dirty linens of our researchers in the public. Constructive criticism and provision of information on social responsibility in the practice of research should be the duty of all enlightened minds.

CONCLUSION

Growing concerns for the integrity of research work are to be taken more seriously now than ever, considering the ubiquity of statistical approaches and computational software that are easily abused in the quest for statistical significance. This pertinent review has attempted to draw attention to the ignorance and omissions of research practitioners in their misunderstanding and misapplication of statistical routines and tools. The influence of personal expectations for statistical outcomes and the crime of data falsification were also discussed in detail. Given the increasing tendency for misconduct in research reporting, the need for professional intervention was explored with the intention of early prevention, detection and further education.

REFERENCES

- Baker, M. (2016). Statisticians issue warning on p values. *Nature, 531*, 151.
- Bodla, B. S. (2017). *Introduction to research methodology. CP-206*. Retrieved on 27th May, 2017 from http://www.ddegjust.ac.in/studymaterial/mba/cp-206.pdf
- Boston University Libraries (n.d.). *What is research data*? Retrieved on 29th May, 2017 from http://www.bu.edu/datamanagement/background/whatisdata/
- Chakrabarty, K. C. (2012). *Uses and misuses of statistics*. Address delivered at the DST Centre for Interdisciplinary Mathematical Sciences, Faculty of Science, Banaras Hindu University, as part of the 150th Birth Anniversary Celebrations of Mahanama Pandit Madan Mohan Malviya, Varanasi, 20th March, 2012. pp 1-5.
- Congdon, J. D. & Dunham, A. E. (1999). Defining the beginning: The importance of research design. In K. L. Eckert, K. A. Bjorndal, F. A. Abreu-Grobois, & M. Donnelly (Eds.), *Research and Management Techniques for the Conservation of Sea Turtles*. pp 1-5. Washington, DC: IUCN/SSC Marine Turtle Specialist Group Publication.
- Deshmukh, S. G. (n.d.). *Research, statistics and misuse/abuse of statistics*. Retrieved on 27th May, 2017 from https://www.researchgate.net/profile/S_G_Deshmukh/publication/280090993_SGD-Research-abuses-of-statistics/links/55a7e95d08ae815a0421280a.pdf
- Emaikwu, S. O. (2012). *Fundamentals of research methods and statistics*. Makurdi: Selfers Academic Press Limited. pp 1-428.
- Field, A. (2012). Analysis of covariance (ANCOVA). Retrieved on 29th June, 2017 from http://discoveringstatistics.com
- Fischer, P. (n.d.). *New research misconduct policies*. Arlington, VA: National Science Foundation. Pp 1-15. Retrieved on 3rd June, 2017 from https://www.nsf.gov/oig_pdf/presentations/session.pdf
- Gelman, A. (2013). P values and statistical practice. *Epidemiology, 24(1)*, 69-72.
- Graham, J. M. (2001). *The ethical use of statistical analysis in psychological research*. Paper presented at the annual meeting of Division 17 (Counselling Psychology) of the American Psychological Association, Houston, Texas. pp 1-23.
- Hubbard, R. & Bayarri, M. J. (2003). *P values are not error probabilities*. Retrieved on 29th May, 2017 from http://ftp.stat.duke.edu/WorkingPapers/03-26.pdf
- Inzana, T. (2008). *Research misconduct: What it is and how to avoid*. A publication of Research Integrity Office, Virginia Tech. Retrieved on 3rd June, 2017 from

https://www.research.vt.edu/research-integrity-office/brochure/misconduct-brochure.pdf

- Johnson, T. P., Parker, V. & Clements, C. (2001). Detection and prevention of data falsification in survey research. *Survey Research, 32(3)*, 1-15.
- Jonker, J. & Pennink, B. W. (2010). *The essence of research methodology: A concise guide for Master and PhD students in management science*. Heidelberg: Springer
- Kim, M.-K. (2009). *The responsible conduct of KAIST research community, KAIST I&TM*, 1-5. Retrieved on 28th May, 2017 from http://ethics.kaist.ac.kr/irb/life/life.jsp
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioural sciences*. Pacific Groove, California: Brooks/Cole Publishing Company.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press. pp 91-196.
- Lepay, D. A. (2008). *Misconduct in research: Detecting falsification*. APEC GCP Inspection Workshop. Washington, DC: US Food and Drug Administration.
- Ludwig, D. A. (2005). Use and misuse of p-values in designed and observational studies: Guide for researchers and reviewers. *Aviation, Space, and Environmental Medicine, 76(7), Section 1*, 675-680.
- MacCoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology, 49*, 259-287.
- Martz, E. (2013). *No matter how strong, correlation still doesn't equal causation*. Retrieved on 29th May, 2017 from http://blog.minitab.com/blog/understanding-statistics/no-matter-how-strong-correlation-still-doesnt-imply-causation
- Miller, I. & Freund, J. E. (1977). *Probability and statistics for engineers*. Englewood Cliffs, NJ: Prentice-Hall Inc.
- Mullane, K. & Williams, M. (2013). Bias in research: The rule rather than the exception? Retrieved on 30th May, 2017 from https://www.elsevier.com/editors-update/story/publishing-ethics/bias-in-research-the-rule-rather-than-the-exception
- Norris, N. (1997). Error, bias and validity in qualitative research. *Educational Action Research, 5(1)*, 172-176.
- O'Donnell, J. (2012). *What is research?* Retrieved on 27th May, 2017 from https://theresearchwhisperer.wordpress.com/2012/09/18/what-is-research/
- Office of Science and Technology Policy - OSTP (2002). *Federal policy on research misconduct*. Retrieved on 28th May, 2017 from http://www.ostp.gov/html/0012017_3.html
- Owen, S. V. & Froman, R. D. (1998). Uses and abuses of the analysis of covariance. *Research in Nursing & Health, 21*, 557-562.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88(5)*, 879-903. DOI: 10.1037/0021.9010.88.5.879
- Rajendran, N. S. (2001). Dealing with biases in qualitative research: A balancing act for researchers. A paper presented at the Qualitative Research Convention 2001: Navigating Challenges held 25-26 October, 2001 at the University of Malaya, Kuala Lumpu. pp 1-15.
- Regoniel, P. A. (2013). How to reduce researcher bias in social research. Retrieved on 25th May, 2017 from http://simplyeducate.me/2013/08/12/how-to-reduce-researcher-bias-in-social-research/
- Rutherford, A. (2001). *Introducing ANOVA and ANCOVA: A GLM approach*. London: SAGE Publication Ltd.
- Ryan, J. J. C. H. & Jefferson, T. I. (2003). The use, misuse, and abuse of statistics in information security research. *Proceedings of 2003 ASEM National Conference*, St. Loius, MO. Retrieved on 3rd June, 2017 from http://attrition.org/archive/misc/use_abuse_stats_infosec_research.pdf
- Sabin, C. A. (2010). What is bias and how can it affect the outcomes from research? Retrieved on 3rd June, 2017 from http://www.ukcab.net/wp-content/uploads/2010/06/csabin-bias.pdf
- Sarniak, B. (2015). *9 types of research bias and how to avoid them*. Retrieved on 31st May, 2017 from

http://www.imoderate.com/blog/9-types-of-research-bias-and-how-to-avoid-them/

- Schienke, E. W. (2017). BIOET 533: *Ethical dimensions of renewable energy and sustainability systems.* Retrieved on 27th May, 2017 from https://www.e-education.pse.edu/bioet533/node/654
- Simmons, K., Mercer, A., Schwarzer, S. & Courtney, K. (2016). *Evaluating a new proposal for detecting data falsification in surveys*. Pew Research Centre. Retrieved on 2nd June, 2017 from http://www.websm.org/uploadi/editor/doc/1457270637Simmons_2016_Evaluating_a_New_Proposal_for_Detecting_Data_Falsification.pdf
- Wasserstein, R. L. & Lazer, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70(2)*, 129-133. DOI: 10.1080/00031305.2016.1154108
- Weaver, B. (2002). Analysis of covariance. Retrieved on 3rd February, 2018 from http://www.angelfire.com/wv/bwhomedir/notes/ancova.pdf
- Yaddanapudi, L. N. (2016). The American statistical association statement on p-values explained. *Journal of Anaesthesiology and Clinical Pharmacology, 32(4)*, 421-423.

How to cite this article: Abah JA. The quest for statistical significance: ignorance, bias and malpractice of research practitioners. International Journal of Research and Review. 2018; 5(3):112-129.

******