International Journal of Multidisciplinary Approach and Studies
ISSN NO:: 2348 – 537X

An Exploratory Data Analysis on Social Media and Youth Online Political Participation in Nigeria and Malaysia

Aishat Adebisi Abdulrauf*, Norsiah binti Abdul Hamid (PhD) ** & Mohd. Sobhi bin Ishak (PhD) ***

**Doctoral Candidate, School of Multimedia Technology and Communication, Universiti Utara Malaysia.*

***Senior Lecturer, School of Multimedia Technology and Communication, Universiti Utara Malaysia.*

****Senior Lecturer, School of Multimedia Technology and Communication, Universiti Utara Malaysia.*

ABSTRACT

This article examined the exploratory data analysis on social media and youth online political participation in Nigeria and Malaysia. Specifically the data screening procedures and preliminary analysis of the data collected were explored. A total of 369 Nigerian undergraduates of Ahmadu Bello University, Zaria and Malaysian undergraduates of Universiti Utara Malaysia completed a self-administered questionnaire on a 7-point semantic differential scale with bipolar adjectives as anchors. The data collected was analyzed using Statistical Packages for Social Sciences (SPSS) application software version 18. To fulfil the requirement for a multivariate analysis, response bias, missing values, outliers, normality and multi-collinearity test were conducted and the results indicated that the data for the study were fit and could be used for further multivariate analysis.

KEYWORDS: *Exploratory Data Analysis (EDA), Data Screening, Preliminary Analysis, Online Political Participation.*

INTRODUCTION

Ideally the first step to a proper data analysis whether for a simple or complex problem should be a detailed examination of the data. This process is known as Exploratory Data Analysis (EDA). EDA involves computing various statistics and graphs to determine whether or not a data is fit for further analysis. Hence, the primary purpose of EDA is to examine and get to know your data. As a result, Leech, Barrett and Morgan (2005) noted that it is important to input data in SPSS and conduct an EDA before carrying out further inferential analysis. This is done by checking data set for errors before analysis as some errors can completely distort analysis (Pallant, 2011). Hence, effort should be made to take a careful look at the data before further analysis is carried out.

Specifically, the main reasons for conducting EDA are; to know the extent to which statistical assumptions the researcher plans to use are met (Hair, Money, Samouel& Page, 2007)and to detect problems in data such as outliers, non-normal distributions, missing values, problems with coding or generally errors in inputting data. Other reasons could be to get basic demographic information about a study and also to determine the relationships between variables (Leech, Barrett & Morgan, 2005; Pallant, 2011).

Furthermore, the main advantage of EDA is that the more a researcher knows about their data, the better they can be used to develop, test and confirm theories. Thus EDA is important because it enables researchers known as much as possible about a variable or variables before carrying out further analysis on them to test theories of relationships in the humanities.
Additionally, the two basic principles on which EDA is based are skepticism and openness, as a result researchers must be aware that even in widely used statistical techniques may have unreasonable hidden assumptions about the nature of the data at hand, while at the same time being open to possibilities that they do not expect to find in their data.

Consequently, it could be said that EDA comprises of data screening and preliminary analysis. Specifically data screening involves checking for errors and finding and correcting the errors in the data file. Once data are screened, preliminary analysis can begin (Pallant, 2011). As a result, the importance of data screening in a multivariate analysis cannot be overstated (Adebambo, Hasbullah & Norani, 2014).

Unfortunately, most researchers do not undertake this activity as they analyze their data straight away without cleaning or screening. Their data are not explored to see if any assumptions of selected test are violated or not. Hence, not taking advantage of an activity which could give greater insight into their data.

Accordingly, the following preliminary analysis were performed for this study: response bias, missing value analysis, assessment of outliers, normality test and multicollinearity test (Hair, Hult, Ringle &Sarst edt, 2014). Foundational to the data screening, all the 369 returned and usable questionnaires were coded and entered into the SPSS. As there were no negatively worded items in the questionnaire, there was no need for reverse coding.

METHODS

Participants and Procedures

Determining an appropriate sample size is important in a survey research in order to minimize the total cost of sampling error. Sampling error can be avoided if the power of statistical test is considered. The power of statistical test is the probability that a null hypothesis will be rejected when it is in fact false (Faul, Erdfelder, Lange & Buchner, 2007).
Consequently, determining an appropriate sample size for a study cannot be overstated. Power analysis is one of the statistical procedure for determining the minimum sample size for a research (Bruin, 2006) based on the part of the model with the highest number of predictors (Hensler, Ringle, &Sarsledt, 2012). Thus, to know the required minimum sample size for this study, a *priori* power analysis was carried out using G*Power 3.1.7 software (Cohen, 1988; Faul, Erdfelder, Buchner & Lang, 2009). This method of sample size determination is used because Hair et al. (2014) asserts that the *priori* power analysis is the best method to determine the minimum sample size for a PLS study.

Accordingly, with the following parameters: power (1-β err prob; 0.95), an alfa significance level (α err prob; 0.05), medium effect size f^2 (0.15) and three main predictor variables (access to political information on Facebook and Twitter, political interest, policy satisfaction), a minimum sample of 119 would be required to test a regression based model (Faul et al., 2007; 2009). Although the result of the priori power analysis indicated the minimum of 119 respondents would be needed for this study, to avoid the issue of low response rate, it became necessary to contemplate other means to increase the sample size for this study.

In agreement, Kotrlik and Higgins (2001), stated that sample size determination procedures indicate the minimum sample size requirement for a study. As a result, since the response rates in most studies are typically below 100% due to reasons such as lost questionnaires or uncooperative participants, scholars (Fink, 1995; Salkind, 1997, Keyton, 2015) recommend that researchers oversample because by increasing the sample size the response rate might also increase (Keyton, 2015).

Consequently Kotrlik and Higgins (2001) contend that if a researcher chooses to oversample one of the recommended methods used to carry out this technique is to take the sample in two steps:first is to use the results of the first step to decide how many additional responses may be needed for the second step. Consequently, based on this approach, the minumum sample size of 119 suggested for this study was increased by 100% as recommended by Gregg (2008). Hence, based on the 100% increase another 119 (119×100÷100=119) was added to the already existing 119to make up a sample size of 238 (119+119=238) for this study.This increase was necessitated by the rule of thumb which states that researchers should select as large a sample as possible from the population (Creswell, 2012) to reduce sampling error especially in terms of probability sampling (Keyton, 2015).

However as this is a comparative study, each country was disproportionately allocated the sample size of 238. Therefore, 238 respondents were allocated to Nigeria and another 238 to Malaysia. Each country was allocated equal sample size to facilitate the notion of equivalence in cross-national comparative (Castletrione, 2015). Moreover, similar comparative studies (Casteltrione, 2014; Valeriani&Vaccari, 2015) also allocated equal sample sizes to each country for their study.

Accordingly, the demographic profile of respondents indicated that in terms of country, 54.5% of respondents were Nigerians while 45.5% were Malaysians, while the gender distribution of respondents was 46.6% for male and 53.4% for female. In terms of age, 26.3% of respondents were between the age ranges of 15-19 years, 61% were between 20-24 years, 10% were between 25-27 years,and while the age ranges of 30-35 years and 36- 40 years both accounted for 1.4% each.

Furthermore, in terms of ethnicity, ofthe 369 respondents, 23.6% were Hausa/Fulani, 11.7% were Yoruba, 7.6% were Igbo, and 11.7% were from other minority ethnic groups in Nigeria. While Malaysian ethnicity comprised of 26.6% Malays, 9.5% Chinese, 6.5% Indians and 3% representing other minority ethnic groups in Malaysia. Also the religion of respondents ranged from Nigerian respondents who are 27.4% Muslims, 26.6% Christians and 0.5% traditionalists, to Malaysian respondents who were 28.2% Muslims, 8.4% Buddhists, 5.4% Hindus, 3.3% Christians, and 0.3% had no religion.

RESULTS AND DISCUSSIONS

A total of 476 questionnaires were distributed to undergraduate students of Ahmadu Bello University Zaria and Universiti Utara Malaysia. Rigorous administration procedure was used to try to achieve as high a response rate as possible (Salant&Dillman, 1994). This high response rate from participant is sought so that the researcher can have confidence in generalizing the result from the sample to the population under study (Creswell, 2012). This was partly achieved by having students distribute the questionnaires, a technique commonly practiced by communication researchers (Keyton, 2015). Hence students were given instructions on the type of participant to seek that suits the study purpose. This technique was particularly useful in this study because the respondents were also students.

Response Rate

Out of the 476 questionnaires distributed, 383 questionnaires were returned indicating a response rate of 80%. This high return rate can be attributed to several reminders sent to respondents through phone calls (Salim, Smith &Bammer, 2002) and text messages. As can be seen in Table 1.1, of the 383 questionnaires returned, 14 were unusable because a significant part of the questionnaires were not completed by respondents (Keyton, 2015). Precisely, the unusable questionnaires had 15% uncompleted items in the overall questionnaire or 5% uncompleted items from a single construct, infact in some cases there were no response at all for multiple constructs, as a result they were dropped from the study (Hair et al., 2014).

Consequently, 369 usable questionnaires were left. This accounted for 77.5% valid response rate. Hence, based on the suggestion of Creswell (2012) that a response rate of 50% or above is adequate for surveys, the number of valid response were used for further analysis.

Table 1.1 ***Response Rate of the Questionnaires***

Response	Frequency/Rate
Number of distributed questionnaires	476
Returned questionnaires	383
Returned and usable questionnaires	369
Returned and excluded questionnaires	14
Questionnaires not returned	93
Response rate	80%
Valid response rate	77.5%

Response Bias

Although, the researchers tried to avoid the issue of non-response bias by adding to the sample size (Keyton, 2015) making the final number of sample 476, yet, the non-response bias test was carried for the study.

Response bias is simply when the responses of participants do not accurately reflect the views of the sample and the population (Creswell, 2012), hence non-response bias is the bias that happens when those who answer the questionnaire differ in meaningful ways from those who did not. This might affect the generalizability of the results to the population of the study.
Hence, as recommended by Malhotra, Hall, Shaw and Oppenheim (2006), late respondents were used in place of non-respondents in order to estimate the non-response bias rate, because the late responders to the questionnaires might not have responded if there was no rigorous follow-up procedure by the researcher. Hence, questionnaires returned within 60 days were treated as early responses while those returned after 60 days were regarded as late responses. Accordingly, 217 (59%) respondents who answered the questionnaires within 60

days were classified as early responses,while 152 (41%) who responded to the questionnaires after 60 days were categorized as late responses.

Specifically, an independent samples T-test was carried out to identify any possible non-response bias on the main study variables of Access to Political Information on Facebook and Twitter (APIFT), Political Interest (PI), Policy Satisfaction (PS) and Online Political Participation on Facebook and Twitter (OPPFT). Table 1.2 presents the results of the independent sample T-test obtained.

Table 1.2
Results for IndependentSamples T-test for Non- Response Bias

						Levene's Test for Equality of Variance	
Variables	**Group**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**	**F**	**Sig.**
APIFT	Early Response	217	3.1382	1.41937	.09635	.152	.697
	Late Response	152	3.0263	1.41301	.11461		
PI	Early Response	217	3.4217	1.30028	.08827	.267	.604
	Late Response	152	3.3298	1.37786	.11176		
PS	Early Response	217	2.8539	1.22099	.08289	.003	.957
	Late Response	152	2.7923	1.27137	.10312		
OPPFT	Early Response	217	2.5598	1.34601	.09137	1.416	.235
	Late Response	152	2.5187	1.23560	.10022		

As shown in Table 1.2, the results of the independent samples T-test indicates that the equal variance significance values for each of the four main study variables were greater than the 0.05 significance level of Levene's test for equality of variance (Pallant, 2010). This means that the assumption of equal the variance between early and late respondents has not been violated. Hence non-response bias was not an issue in the study.

Missing Value Analysis

The SPSS original data set for this study contained 25,092 data points, of which 88 were randomly missed representing 0.35%. Specifically, Access to Political Information on Facebook and Twitter (APIFT) and Political Interest (PI) each had 6 missing values, while Policy Satisfaction (PS) and Political Knowledge (PK) had 7 and 8 missing values respectively. However, Online Political Participation on Facebook and Twitter (OPPFT) had the highest number with 61 missing values. These can be seen in Table 1.3.

Hence, in as much as there is no rule on the acceptable number of missing values in a data set for making a valid statistical inference, scholars have generally agreed that missing rate of 5% or less is non-significant (Tabachnick& Fidel, 2007). Thus, the 0.35% of missing value in this study is within acceptable range.

However, before the missing values treatment was carried out, the researcher ensured that there were less than 5% values missing per indicator for all the remaining questionnaires (Hair et al., 2014). First questionnaires with more than 15% overall missing value for an observation were excluded from the analysis for this study. However, even some questionnaires that did not have up to 15% over missing value were excluded because respondents did not answer a high proportion of responses for a single constructs , in fact some didn't even respond to a single question in multiple constructs, hence such cases were removed (Hair et al, 2014). Consequently, Median of nearby points was used to replace missing data for the study.

Table 1.3 Total and Percentage of Missing Values

Latent Variables	**Number of Missing Values**
Access to Political Information on Facebook and Twitter (APIFT)	6
Political Interest (PI)	6
Policy Satisfaction (PS)	7
Political Knowledge (PK)	8
Online Political Participation on Facebook and Twitter (OPPFT)	61
Total	**88 out of 25,092 data points**
Percentage	**0.35%**

Note: Percentage of missing value is gotten by dividing the total number of randomly missing values for the entire data set by total number of data point multiplied by 100

Assessment of Outliers

Outliers are data points that deviate obviously from others (Cousineau&Chartier, 2010). Hence the decision researchers make about how to define, identify and handle outliers have important implications as they could significantly alter results in a study (Aguinis, Gottfredson&Joo, 2013).

The general view of outliers is that they are problematic and must be 'fixed' which is not necessarily appropriate in many research context (Aguinis, Gottfredson&Joo, 2013).

Unfortunately, the issue of how to define, identify and handle outliers has been a problematic one. For example in a meta-analysis of outlier studies in different research contexts Aguinis, Gottfredson, and Joo, (2013) noted that a sizeable number of works on outliers provide vague and inconsistent recommendations on how to define, identify and handle outliers, specifically they noted that Tabachnick and Fidell (2007) which is one of the most popular works in outlier studies also followed in this line.

Notwithstanding, outlier diagnostics was conducted for this study. To spot observations which appear to be outside the SPSS value labels due to wrong data entry, first frequency tables were tabulated for all the variables in this study using the minimum and maximum statistics. From the analysis of frequency statistics, no value was found outside the expected range.

Furthermore, Mahalanobis Distance (D2) was used to detect multivariate outliers (Osborne &Overbay, 2004, Pallant, 2011). Mahalanobis Distance is the distance of a case from the centroid of the remaining cases where the centroid is the point created at the intersection of all the means of all the variables (Stevens, 1984; Tabachnick&Fidell, 2007). A large Mahalanobis distance may indicate that the corresponding observation is an outlier (Aguinis, Gottfredson &Joo, 2013;Hair, Wolfinbarger, Ortinau& Bush, 2008).

To know outliers it is important to know the critical chi-square value using the number of individual variables as the degree of freedom (Pallant, 2011). Hence with the exclusion of demographic and other categorical variables the degree of freedom for this study became 53 (54-1=53). Consequently, based on the 53 items for this study, the recommended threshold of chi-square was 70.99 (p=0.05). Accordingly, after rearranging the Mahalanobis value on the SPSS in descending order, it was discovered that 53 Mahalanobis values exceeded this threshold.

However, the researcher made a decision not to delete the outliers. This is based on the premise that Cortina (2002) advised researchers should be careful in deleting outliers because it may increase their chances of finding what they want to find which is dangerous. Hence, once outliers have been identified, the researcher must decide what to do. If there are only a few outliers they could just simply be removed from the data set, however, if the outliers are much the researcher must decide whether or not to delete those (Hair et al, 2014). If outliers are legitimate, researchers' should use reasoned argument and thoughtful consideration in making decisions. Legitimate outliers can be kept and still not violate assumptions because when outliers are legitimate the data is most likely to be representative of the population as a whole hence outliers should be retained (Osborne &Overbay, 2004).

Similarly, Analytical Methods Committee (1989) noted that in as much as outliers are discordant to the rest of the data, they should not be seen as errors and should not be deleted just as Stevens (1984) also does not recommend outright deletion of outliers. Furthermore, Aguinis, Gottfredson, and Joo, (2013) noted that one of the outlier handling technique as recommended by scholars is to acknowledge its presence but still keep the outlier values prior to analysis.

Additionally, Burke (2001) warns that no value should be removed from a data set on statistical grounds alone. For example outlier test may indicate that there is an error in as study on the basis of certain assumptions but this is not to say that the point is wrong, because despite the extreme value in a data set the suspect value(s) could actually be the correct piece of information. Hence outliers should only be deleted when there is a technical reason to do so.

Hence if there are grounds for believing that the data is normally distributed (histogram normal probability plot in the case of this study) then outliers can be checked just to know their position (Burke, 2001). Moreover, a non-parametric analysis method like PLS-SEM does not assume that data are normally distributed, hence researchers can get results that are robust in the presence of outliers (Aguinis, Gottfredson&Joo, 2013). This explains why Osborne and Overbay (2004) advise that to deal with issues of outliers, non-parametric analysis can be used as they have few if any distributional assumptions.

Moreover as a rule of thumb, if more than 20% of data are identified as outliers the quality of data collected could be questions (Burke, 2001), hence since the number of outliers in this study are not up to 20% of data, the data for this study can be used for further analysis. Consequently, this study intends to utilized PLS-SEM which is a non-parametric analysis technique, and the outliers do not affect the normality of data (see Figures 1.1 and 1.2), the outliers for this study will not be deleted. Therefore, the data set for this study remained 369.

Normality Test

Even though PLS-SEM as a non-parametric statistical method does not require data to be normally distributed before analysis can be carried out, scholars have advised that researchers should perform a normality test on their data (Hair, Sarstedt, Ringle& Mena, 2012) as highly skewed or kurtosis data can inflate the bootstrapped standard error estimates which in turn could underestimate the statistical significance of the path coefficients (Ringle, Starstedt& Straub, 2012).

Against this backdrop, this study utilized a graphical approach to check for normality of the data collected (Tabachnick &Fidell, 2007). The graphical method was chosen because Field (2009) suggested that a study sample larger than 200 should look at the shape of the distribution graphically rather than look at the value of skewness and kurtosis statistics. Hair et al. (2004) also noted the importance of examining the skewness and kurtosis of a data distribution. As a result, with a sample size of 369, which is clearly larger than 200, using the graphical method to test for normality of data for this study is justified.

Accordingly, a histogram and normal probability plot were used to confirm that the assumptions of normality were met in this study. Figure 1.1 shows that the data collected for this study follows a normal pattern since all the bars on the histogram were close to a normal curve. Hence, the bell-shaped curve in Figure 1.1 indicates a normal distribution (Hair et al., 2014). Thus, even though PLS can work with non-normal data this study did not violate normality assumptions.

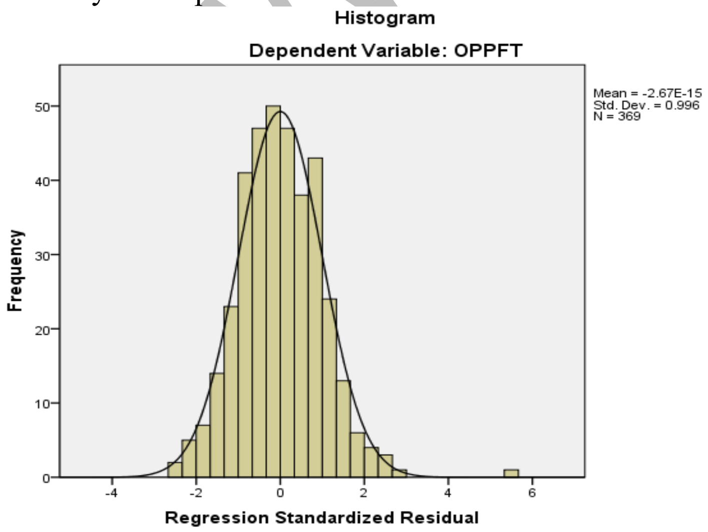

Figure 1.1: Histogram and Normal Probability Plot

Figure 1.1 shows scores appear to be normally distributed. This is also supported by the normal probability plots where the observed value for each score is plotted against the expected value from the normal distribution. The reasonably straight line seen in Figure 1.2 indicates a normal distribution (Pallant, 2011).

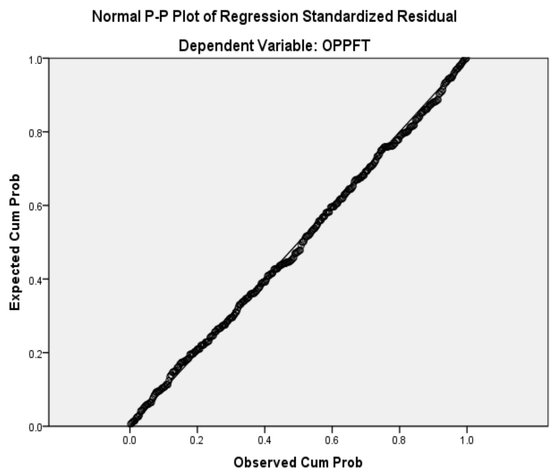

***Figure 1.2*: Q-Plot**

Multi-collinearity Test

Multi-collinearity is when one or more exogenous latent constructs become highly correlated. The presence of multicollinearity among the exogenous latent constructs can substantially alter the estimates of regression coefficients and their statistical significance test (Hair, Black, Babin, Anderson & Tatham, 2006). Specifically, multicollinearity increases the standard errors of the coefficients, which in turn render the coefficients statistically non-significant (Tabachnick&Fidell, 2007). Therefore, in this study, two methods (correlation matrix of exogenous latent constructs and Variance Inflated Factor) were used to detect multicollinearity.

First, the correlation matrix of the exogenous latent constructs was examined. A correlation coefficient of 0.90 and above indicates multicollinearity between exogenous latent constructs. Table 1.4 shows the correlation matrix of all the exogenous latent constructs.

Table 1.4 Correlation Matrix of the Exogenous Latent Constructs

	APIFT	PI	PS	OPPFT
APIFT	1			
PI	.557**	1		
PS	.429**	.565**	1	
OPPFT	.598**	.650**	.591**	1

Note: **correlation is significant at the 0.01 level (2-tailed)

As presented in Table 1.4, the correlations between the exogenous latent constructs were sufficiently below the recommended threshold values of 0.90 or more, indicating that the exogenos latent constructs for this study were independent and not highly correlated.

Secondly, the Variance Inflated Factor (VIF) tolerance value and condition index were used to detect multicollinearity of exogenous factors for this study. According to Hair, Ringle and

Sarstedt (2011), multicollinearity is an issue if VIF value is higher than 5 , tolerance value is less than 0.20 and condition index is greater than 30. Table 1.5 shows VIF values, tolerance values and condition indices for the exogenous latent constructs for this study.

Table 1.5: *Tolerance and Variance Inflation Factors (VIF)*

Dependent variable	**Independent variable**	**Collinearity Statistics**		**Condition Index**
		Tolerance	**VIF**	1.000
APIFT	PI	.681	1.469	
	PS	.681	1.469	5.925
PI	PS	.816	1.225	
	APIFT	.816	1.225	6.872
PS	APIFT	.690	1.449	
	PI	.690	1.449	9.113

Table 1.5 shows that multicollinearity did not exist among the exogenous latent constructs as all the VIF values were less than 5, tolerance values exceeded .20 and condition indices are below 30 as suggested by Hair et al. (2011). Therefore multicollinearity is not an issue in this study.

CONCLUSION

The importance of initial analysis before undertaking further advanced PLS-SEM analysis cannot be overstated as it could lead to inflated estimated standard error. Yet, studies have been carried out without considering data screening and preliminary analysis. As a result, this study was conducted to highlight an important part of multivariate analysis which includes assessment of missing values, outliers, normality and multicollinearity. Evidently, these analysis provide better insight into data characteristics of a particular study as well as help in meeting the assumptions of multivariate analysis.

ACKNOWLEDGEMENT

This article is a result of a research which was funded by Universiti Utara Malaysia, Malaysia under the Postgraduate IncentiveResearch Grant Scheme (S/O Code: 15697).

REFERENCES

i. Academic Affairs Department (2015) Statistics of Universiti Utara Malaysia Undergraduates based on Ethnicity. Academic Affairs Department, Universiti Utara Malaysia.

ii. Adebambo, H.O, Hasbullah, A. &Norani, N. (2014). Influence of top management commitment, stakeholder pressure and public concern on sustainable environmental management practices in Malaysia: Data Screening and Preliminary Analysis. *The*

International journal of Business Management, 2(11) pp.189-196 retrieved January 10 2015 from www.theijbm.com

iii. Aguinis, H., Gottfredson, R. K., &Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. 1-32, *Organizational Research Methods*, doi: 10.1177/1094428112470848.

iv. Bruin, J. (2006). Newtest: command to compute new test. *UCLA: Academic Technology Services, Statistical Consulting Group*.

v. Burke, S. (2001). Missing values, outliers, robust statistics & non-parametric methods. *LC-GC Europe Online Supplement, Statistics & Data Analysis*,*2*(0), 19-24.

vi. Casteltrione, I (2012). Facebook and political participation in Italy and the UK: Addressing gaps in political participation research, 4th ECPR Graduate Conference, Jacobs University, Brenen, 4-6th July, 2012.

vii. Casteltrione, I. (2014). Facebook and Political Information in Italy and the UK: An Antidote against Political Fragmentation and Polarization? *Online Journal of Communication and Media Technologies*, *4*(1), 27.

viii. Casteltrione, I (2015). The internet, social networking websites and political participation research: Assumption and contradictory evidence. *First Monday,* 20(3) doi:org/10.5210/fm.v20i3.5462

ix. Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

x. Cousineau, D. &Chartier, S. (2010). Outlier's detection and treatment: A review. International Journal of Psychological Research 3(1). Retrieved November 13, 2015 from http://mvint.unbned.ed.co

xi. Creswell, J. W. (2012). *Educational research. Planning, conducting and evaluating quantitative and qualitative research*. (4th ed.). Boston, MA. United States of America: Pearson Education, Inc.

xii. Faul, F., Erdfelder, E., Buchner, A., & Lang, A-G (2009). Statistical power analysis using G*Power 3.1: Tests for correlation and regression analysis. Behavior Research Methods. 41. Doi:10.3758/brm.41.41149.

xiii. Faul, F., Erdfelder, E., Lang, A-G & Buchner, A. (2007). G*power 3: A flexible statistical power analysis programme for the social, behavioral and biomedical sciences. Behavior Research Methods. 39. Doi: 10.3758/bf03193146.

xiv. Field, A. (2009). Discovering statistics using SPSS (3rded). London: SagePublications.

xv. Fink, A. (1995). The survey handbook (Vol. 1). *Thousand Oaks, CA*.

xvi. Gregg, M. B. (2008). *Field epidemiology*. Oxford University Press, USA.

xvii. Hair. J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). Multivariate data analysis (7thed.). Upper Saddle River, New Jersey: Prentice Hall.

xviii. Hair. J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2006). Multivariate data analysis (6thed.). Upper Saddle River, New Jersey: Pearson/Prentice Hall.

xix. Hair, J.F., Hult, G.T.M, Ringle, C.M. &Sarstedt, M. (2011). Partial least squares structural equation modelling: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19 (2), 139-151. doi:10.2753/mtp1069-6679190202.

xx. Hair, J.F., Hult, G.T.M, Ringle, C.M. &Sarstedt, M. (2014). *A Primer on partial least squares structural equation modelling (PLS-SEM)*. London: Sage Publications.

xxi. Hair, J. F., Money, A. H., Samouel, P., & Page, M. (2007). Research methods for business. *Education+ Training*, *49*(4), 336-337.

xxii. Hair, J.F., Sarstedt, M., Pieper, T.M., &Ringle, C.M. (2012). The use of partial least squares structural equation modelling in strategic management research: A review of past practices and recommendations for future application. *Long Range Planning,* 45, 320-340. Retrieved from http://www.elsiever.com/locate/lrp.

xxiii. Hair, J.F., Sarstedt, M., Ringle, C.M., &Menae, J.A. (2010). An assessment of the use of partial least squares structural equation modelling in marketing research. Academic Marketing Science, 40, 414-433. doi:10.1007/s1174-011-0261-6.

xxiv. Hair, J. F., Wolfinbarger, M. F., Ortinau, D. J., & Bush, R. P. (2008).*Essentials of marketing research (2nd ed)*. McGraw-Hill/Higher Education.

xxv. Hensler, J., Ringle, C.M. &Sarsledt, M. (2012). Using partial least squares path modelling in advertising research: Basic concepts and recent issue in Okazaki, S (Eds.) *Handbook of Research on International Advertising*. Massachusetts: Edward Elgar Publishing, Inc.

xxvi. Keyton, J. (2015). *Communication research: Asking questions and finding answers* (4thed.).New York: McGraw Hill Higher Education.

xxvii. Kotrlik, J. W.., & Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information technology, learning, and performance journal*, *19*(1), 43.

xxviii. Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation*. Psychology Press.

xxix. Malhotra, N., Hall, J., Shaw, M., & Oppenheim, P. (2006), Marketing Research. An applied orientation (3rded.). French Forest: Prentice Hall.

xxx. Management Information System (2015). Undergraduate Statistics. Management Information System (MIS), Ahmadu Bello University, Zaria.

xxxi. Osborne, J. W., &Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, *9*(6), 1-12. Retrieved January 11, 2016 from http://PAPEonline.net/getvn.asp?v=9&n=6

xxxii. Pallant, J. (2010). SPSS survival manual: A step by step guide to data analysis using SPSS (4thed). New York: Open University Press.

xxxiii. Ringle, C. M., Sarstedt, M., & Straub, D. (2012). A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly (MISQ)*, *36*(1).

xxxiv. Salant, P. & Dillman, D.A. (1994). How to conduct your own survey. New York: Wiley

xxxv. Salim, S.M., Smith, W.T., &Bammer, G. (2002). Telephone reminders are a cost effective way to improve responses in post health surveys. Journal of Epidemiology and Community Health. 56 (2) retrieved November 29, 2015 from http://www/jstor.org/sTable/25569634

xxxvi. Salkind, N. J. (1997). Exploring research (3 rd.). *Bartlett, JE, Kotrlik, JW, and Higgins, CC (2001). "Organizational Research: Determining Appropriate sample Size in Survey Research". Information Technology, Learning, and Performance Journal*, *19*(1), 43-50.

xxxvii. Sheppard, J. (2012). *Examining internet use and civic voluntarism in Australia and Canada.* Paper Presented at Canadian Political Science Association Annual Conference, University of Alberta, Edmonton.

xxxviii. Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, *95*(2), 334.

xxxix. Tabachnick, B.G. &Fidell, L.S. (2007). Using Multivariate Statistics (5th ed). Boston, Massachusettes: Alllyn and Bacon/Pearson Education.

xl. Valeriani, A. &Vaccari, C. (2015). Accidental exposure to politics on social media as online participation equalizer in Germany, Italy and UK. *New Media and Society.*doi: 10.1177/1461444815616223

xli. Whiteley. P. F. (2005). Citizenship education. *The political science perspective research report.* RR 631. National Foundation for Education Research.

xlii. Xenos, M., Vromen, A., & Loader, B. D. (2014). The great equalizer□ ? Patterns of social media use and youth political engagement in three advanced democracies, 17(2), 37–41. doi:10.1080/1369118X.2013.871318.