International Journal of Multidisciplinary Approach and Studies

ISSN NO:: 2348 – 537X

An Assessment of an Exploratory Data Analysis on Consumer's Motivations and Online Newspapers Consumption in Pakistan

Muhammad Ramzan Pahore*, Norsiah binti Abdul Hamid & Awan binti Ismail*****

**PHD candidate, School of Multimedia Technology and Communication, Universiti Utara Malaysia and University of Sindh, Jamshoro, Pakistan*
***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 scrutinized the exploratory data analysis on Consumers' motivations and Online Newspapers Consumption. Most importantly, the data screening processes and preliminary analysis of the data composed were explored. A total of 446 University students from three big universities of Sindh, Pakistan complete a self-administered questionnaire on 7-point Likert type scale. SPSS software version 23was usedto analyses collected data. For multivariate analysis all the required were completed, as response bias, missing values, outliers, normality and multi-collinearity and the results shows that the data for the study were suitable and can be used for additional multivariate analysis.

KEYWORDS: *Exploratory Data Analysis (EDA), Data Screening, Preliminary Analysis, Consumers' Motivations, Online Newspapers Consumption*

INTRODUCTION

Detailed examination of the date in considered the first step into a proper data analysis either problem is simple or complex. This detailed examination process is called as Exploratory Data Analysis. JohnW. Tukey has founded the term exploratory data analysis or EDA describing the act of looking at data to see what it seems to say (p-33, Morgenthaler, 2009). EDA contains computing many statistics and graphs to know whether data is fit for any further analysis. The main motive behind, doing EDA is to know and examine about your data. Hence, for carrying out further inferential analysis, it is mandatory to key in data in SPSS and conduct an Exploratory Data Analysis. According to Palland (2010) errors in the data may mislead the results therefore; EDA was carried out for checking data set for errors. Hence, this detailed process is only be made to take a careful look at the data before analysis is carried out.

Most importantly, the prime reasons for doing EDA are, according to Hair, Money, Samouel and Page (2007), to know at what extent the statistical assumptions are met that researchers have designed for the study also, to identify problems in the data for example; outliers, non-normal distributions, missing values, problems in coding or errors in keying in data into SPSS. Further, Pallant (2010) pointed out other reasons to determine relationships between modeled variables and to collect basic demographic information about the study.

Further, more a researcher knows about the data, the better they can use it to develop, test, and confirm a theory that is the most important advantage of doing EDA. Furthermore, the primary aim of EDA is to look at the data and to think about the data from many points of view (Morgenthaler, 2009) and it maximize the value of data. This allows researcher to know about the variables in his study before he carried out final analysis on them to test theories of the relationships. Moreover, the two basic rules on which EDA is standing are skepticism and openness. As a result, researchers should be beware that there are unreasonable hidden assumptions in the widely used statistical techniques at one hand, while at the same time being open to possibilities that researchers do not assume to find in data on other hand.

Hence, commonly it is said, EDA involves data screening and preliminary analysis. So, data screening comprises of error checking and modifying errors in the data file. According to Pallant (2010) once data are screened and errors free then researchers may take preliminary data analysis.

However, there is lack of EDA and relevant published material with the methodological literature (Jebb, 2016) further, Abdulrauf, Abdul Hamid, and Ishak (2016) also pointed out that most researchers did not go through proper data screening and following proper procedure, directly analyze data. Hence, this study is going through detailed examination of data before doing analysis into PLS-SEM.

Therefore, in this study preliminary analysis were conducted; response bias, missing values, calculation of outliers, normality test, and multicollinearity test (Hair, Hult, Ringle & Sarstedt, 2014). Hence, 426 usable questionnaires were entered and coded accordingly into SPSS 23 version.

METHODS

Participants and Procedures

It is noted that, determining an appropriate sample size is very important in a survey research, (Bartlett, Kotrlik, & Higgins, 2001). For reducing the total cost of sampling error a suitable sample size is required. If scholars want to lessen the total cost of sampling error, then the power of a statistical test has to be carried out into consideration. According to experts the power of a statistical test is explained as the possibility that null hypothesis will be rejected when it is in fact false (Cohen, 1992; Faul, Erdfelder, Lang, & Buchner, 2007). Researchers Borenstein, Rothstein, and Cohen, (2001); Kelley and Maxwell, (2003) have largely agreed that the bigger the sample size, the greater the power of a statistical test. Power analysis is a statistical technique for defining an appropriate sample size for a research study (Bruin, 2006).According to one of the guru in Partial Least Square-Structural Equation Modelling,Hair, Hult, Ringle, and Sarstedt (2014) to determine the effective and appropriate sampling size for current study, the recommended approach in PLS-SEM such as G*Power statistical analysis procedure is adopted to get minimum sample size. Hence, to determine the minimum sample for current study, an a priori power analysis is accompanied using G*Power 3.1 software (Faul, Erdfelder, Buchner, & Lang, 2009; Faul et al., 2007). Using the following parameters: Power (1-β err prob; 0.95), an alpha significance level (α err prob; 0.05), medium effect size f^2 (0.15) and seven main predictor variables (i.e. information seeking motivation, entertainment motivation, social utility motivation, personal utility motivation, and escapism motivation, age, and gender), hence, for current study a minimum 153 sample would be

requisite to test a regression based models (Figure 1.1; Cohen, 1992; Faul et al., 2009; Faul et al., 2007).

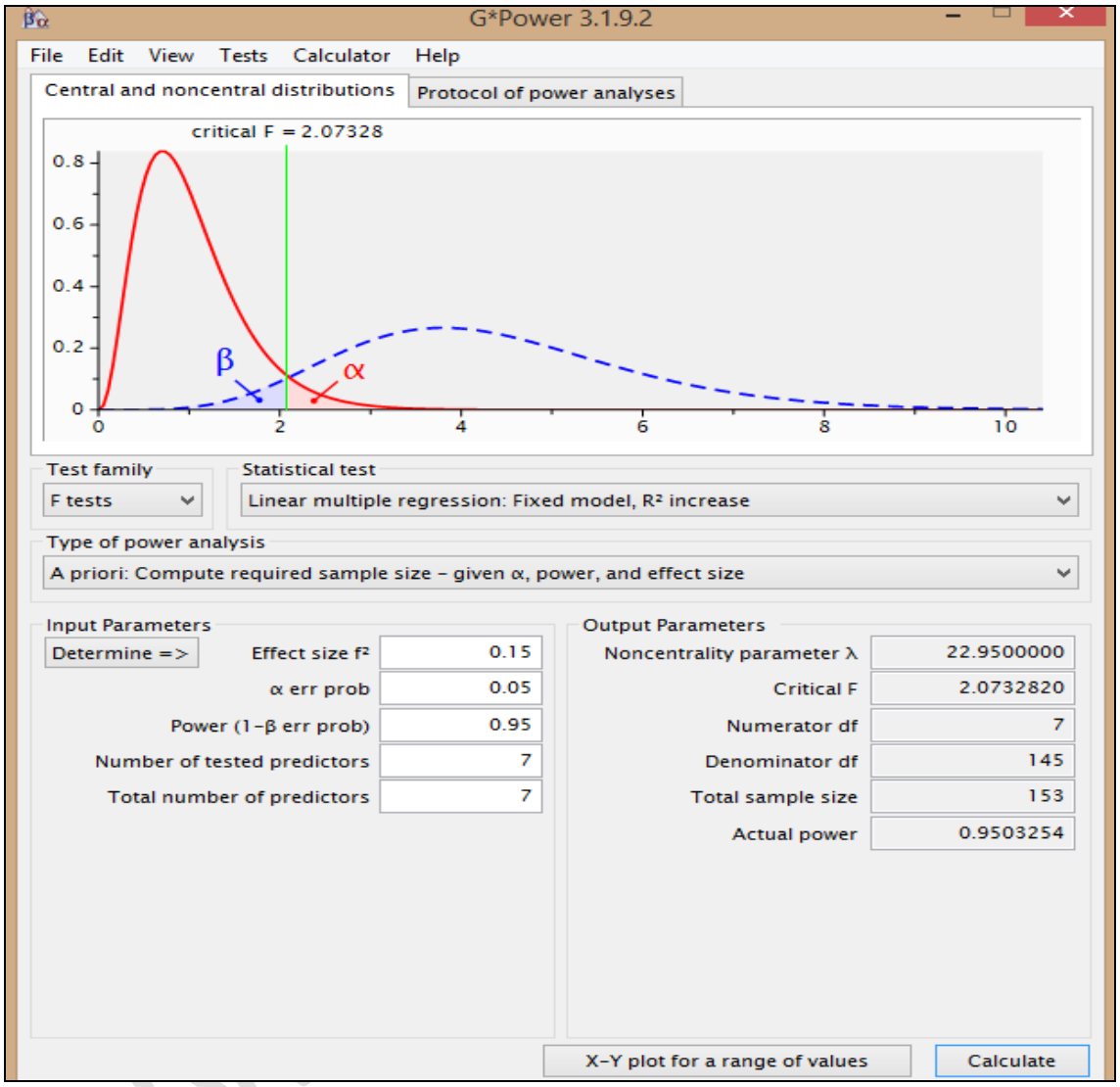

Figure 1.1*The Output of a Priori Power Analysis*

The output of priori power analysis indicates that a minimum of 153 subjects will be required for the present study. It is worth noting that response rate in the Pakistani context is very poor even among universities students and teachers(Memon & Pahore, 2014; Roshan, Pervaiz,& Soomro, 2013). As a result of the poor response rate, the sample size got through prioripower analysis appears to be in adequate in the present study**.** Therefore, for a given population, it is crucial to consider other ways of determining an adequate sample size. On the basis of this argument, Krejcie and Morgan's (1970) sample size determination criteriais preferred over a priori analysis. Above all, sample size determination criteria (Krejcie and Morgan's, 1970) was used to define the representative sample size for current research because it has taken into account the level of confidence and precision, ensuring that sampling error is minimized.

However, on the basis of population for defining adequate sample size, Krejcie and Morgan (1970) approach is followed because, this will give 381 samples for generalizing the results throughout the population. Hence to reduce the small response rate from unhelpful respondents, the sample size of 381 is added by 40% as advised by Salkind (1997). This increase was required to reduce sampling error by the rule of thumb which suggests that researchers should tale as big a sample as possible from the population (Creswell, 2012). Adding this percentage to 381 provide 533 sample size. As a result, this is suitable to be the representative of the population.

Accordingly, the demographic profile of respondents indicated that with respect to gender, 61% of respondents in the study were male while 39% were female, while the age wise distribution of respondents were in, 53.8%respondents were between the age ranges of 18-22 years, 19.4% were between 23-27 years, 11.2% were between 33-37 years, 9.2% were between 28-32 years and while 3.5% were between 38-42 years and then last age group 43 and above years respondents representing 2.7% the least percentages. Furthermore, in terms of highest qualification, out of 403 respondents, 38.2%were undergraduates, 27.8% were Masters, 18.9% were PhD, and 15% were from MPhil/MS. While University wise distribution out of 403, the highest 42.7% respondents were from University of Sindh, 37.7% were from University of Karachi and 19.6% were from Shah Abdul Latif University

RESULTS AND DISCUSSIONS

Response Rate

A total of 533 questionnaires were administered on university students of three public sector general universities in Sindh, Pakistan: University of Sindh (SU), Shah Abdul Latif University (SALU), and Karachi University (KU). There was rigorous administration process used to get a high response rate as possible (Salant & Dillman, 1994). According to Creswell (2012) for the generalization of the findings from the target sample to the population in this study, a high response rate is required from the students. This was moderately accomplished by having students to disseminate the questionnaires, a technique commonly experienced by communication researchers (Keyton, 2015). However, respondents were given instructions on the types of students are required that suit the purpose of current study.

Thus, four hundred and forty six (446) which indicating 83.6% returned rate out of the 533 administered questionnaires were retrieved. This high response rate has been due to several reminders sent to the respondents. This particularly, according Salim Silva, Smith, and Bammer (2002) and Traina, MacLean, Park, and Kahn (2005) to high return rate was achieved due to reminders send through phone calls and text messages. Please refer Table 4.1 for details. Out of the 446 returned questionnaires, 22 were unusable because a most part of the questionnaires were not filled by respondents (Keyton, 2015). Importantly, Hair, Hult, Ringle, and Sarstedt (2013) suggested that an observation should be deleted finally from the analysis when most of the uncompleted questionnaires have 15% or more unfilled items in the overall questionnaire, or from single construct 5% or more unfilled questions from single construct.

Hence, after deletion of unusable questionnaire, only 424 were left. This stood to be 79.5% usable response rate. Hence, this response rate was above the threshold of 30% minimum

recommended by Sekaran (2003). Further, Creswell (2012) recommended 50% or above response rate is adequate for surveys, therefore, the number of valid responses (79.5%) were used for further analysis.

Table **Error! No text of specified style in document.**.1

Response Rate of the Questionnaires

Response	Frequency/Rate
Number of distributed questionnaires	533
Returned questionnaires	446
Returned and usable questionnaires	424
Returned and excluded questionnaires	22
Questionnaires not returned	87
Response rate	83.6%
Valid response rate	79.5%

Response Bias

In this study, the researcher has added 40% to the sample size (Keyton, 2015) construction the final 533 number of sample to avoid the issue of non-response bias. Similarly, Barclay, Todd, Finlay, Grande, & Wyatt, (2002) also affirmed that the researcher can anticipate non response and avoid it by adding to the sample size. Researcher had changed the unit non response bias by adding 152 to its minimum sample 381, making 533 total samples for survey.

Malhotra, Hall, Shaw, and Oppenheim (2006) have explained Non-response bias as the bias that results when respondents differ in significant means from non-respondents which might affect the generalizability of the findings to the population of the research. Further, Creswell (2012) also define response bias as straight away when the answers of questions do not exactlyreplicate the views of the sample and the population. Hence, non-response bias that takes place when respondents answer the questionnaire completely differ in the clear ways from other respondents who did not that might affect the generalizability of the results in this research throughout the population.

Furthermore, Malhortra, Hall, Shaw and Oppenheim (2006) recommended, late respondents were used instead of non-respondents so as to assume the non-response bias rate, because the late respondents to the questionnaires may not return the questionnaire if researchers have not send follow up reminders and requests.

Therefore, questionnaires which were returned within four weeks' time were treated as early responses while those returned after four weeks' time were counted as late responses. Accordingly for data, 263 questionnaires were categorized as early responses and 161 questionnaires were treated as late responses. In this study some 62% respondent had responded to the questions within four weeks' time while the remaining 37.9% responded after four weeks' time.

Specifically, an independent samples T-test was applied to classify any conceivable for non-response bias on this study variables of consumers' motivations for online newspapers consumption. The results present the independent samples T-test taken for the combined respondents in the table 4.2

Table **Error! No text of specified style in document.**.*2*
Results for Independent-Sample T-test for Non-Response Bias

	Response	N	Mean	Std. Deviation	Std. Error Mean	Levene's Test for Equality of Variance F	Sig
IM	Early Response	263	5.4437	1.43892	.08873	.254	.614
	Late Response	161	5.3978	1.36108	.10727		
EM	Early Response	263	3.7004	1.38937	.08567	.847	.358
	Late Response	161	3.7292	1.32125	.10413		
SUM	Early Response	263	4.1819	1.45675	.08983	1.738	.188
	Late Response	161	4.1366	1.35860	.10707		
PUM	Early Response	263	4.3144	1.58543	.09776	.334	.564
	Late Response	161	4.2457	1.49551	.11786		
ECM	Early Response	263	3.6633	1.39860	.08624	.783	.377
	Late Response	161	3.7868	1.33004	.10482		
ONC	Early Response	263	3.5899	1.47179	.09075	1.932	.165
	Late Response	161	3.6364	1.53876	.12127		

Missing Value Analysis

This study contained 17,507 data points, of total points 44 were randomly missed representing 0.25% in SPSS original data set. Especially, information motivation had 9 missing values, personality utility motivation had 3 missing values, escapism motivation had 14 missing values, and online newspapers consumption had 18 missing values. Missing values can be seen in Table 4.3.

However, there is no fix rule of thumb for accepting the number of missing values in the data set for making valid statistical inference, Tabachnick and Fidel (2007) have confirmed that missing values rate of 5% or less is non-significant. Hence, in this study had only 0.25% of missing value that is within acceptable range.
However, before the missing values handling was carried out, the researchers confirmed less than 5% values missing per indicator for all the remaining questionnaires (Hair et al., 2014). Those questionnaires with more than 15% combined missing value for an observation were omitted 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, hence such cases were removed (Hair et al, 2014).Consequently, median of nearby points was used to replace missing data for the study.
Table **Error! No text of specified style in document.**.3

Total and Percentage of Missing Values

Latent Variables	Number of Missing Values
Information Motivation(IM)	9
Personal Utility Motivation (PUM)	3
Escapism Motivation (ECM)	14
Online Newspapers Consumption (ONC)	18
Total	**44 out of 17507 data points**
Percentage	**0.25%**

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

Assessment of Outliers

Outlier assessment was carried out for this study. Barnett and Lewis (1994) have defined outliers as observations or subsets of observations which look to be inconsistent with the rest of the data. In a regression-based analysis, the existence of outliers in the data set can extremely change the estimates of regression coefficients and guide to unreliable results (Verardi & Croux, 2008). To locate observations which seem to behave outside the SPSS value labels may be due to wrong data entry, frequency tables were formulated 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, since this study used multivariate analysis method, Mahalanobis distance (D2) was used to notice multivariate outliers (Osborne & Overbay, 2004; Pallant, 2010). Mahalanobis distance (D2) is defined as "the distance of a case from the centroid of the remaining cases where the centroid is the point created at the intersection of the means of all the variables" (Tabachnick &Fidell, 2007). Hence, to identify outliers, it is important to see the critical chi-square value using the number of separate variables as the degree of freedom (Pallant, 2011). Hence with the omission of demographic and moderating categorical variables the degree of freedom for this study became 40 (41-1=40).

Subsequently, based on the 40 observed items for this research, the suggested threshold of chi-square was 58.12 (p=0.05). Accordingly, after repositioning the Mahanalobis value on the SPSS in descending order, it was revealed that 41 Mahalanobis values exceeded this threshold. Nonetheless these outliers were not removed from the study because scholars (Aguinis,Gottfredson, & Joo, 2013; Burke, 2001; Osborne & Overbay, 2004) recommended keeping outliers.

Moreover, according to Burke (2001) if more than 20% of data are identified as outliers the quality of data collected could be questioned. However since that is not the case in this study the outlier values are only (5%), calculated as number of outliers were divided usable responses and multiplied with hundred (21/424×100), therefore, the data can be used for further analysis. Furthermore, this study is using PLS-SEM non-parametric analysis software so outliers do not affect the normality of data (See Figures 4.1 and 4.2), even though the outliers for this study were deleted and the data set for the study remained 403.

Normality Test

In this research for determine the normality of the data collected, graphical approach was used (Tabachnick & Fidell, 2007). Symmetrical data distribution is called normal data. So,

according to Pallant (2011) the distribution of the scores on the dependent variables produces a bell-shaped curve. Further to explain bell-shaped curve, the highest frequencies of scores appear in the middle with lesser frequencies appear towards the both ends (Gravetter & Wallnau, 2004, p. 48 in Pallant, 2010).

Furthermore, Field (2009) suggested that a study sample larger than 200 should prefer to examine the shape of the distribution graphically rather than look at the value of skewness and kurtosis statistics. . Likewise, Hair et al. (2014) stated the importance of investigating the skewness and kurtosis of a data distribution. According to Field, bigger sample decreases the standard errors thatin resultexpand the value of the kurtosis and skewness statistics. The test of normality for this study was however carried out using histogram and normal probability (Q-Q) plot. The graphical method was suitable for this study because sample size of this study is 403, which is above 200. Hence, graphical method (histogram and normal probability (Q-Q) plot) to test for normality of data for this study is well justified.

Consequently, in this study, histogram and normal probability plot were used to check the assumptions of normality. Figure 4.1 shows that, for this study, data gathered follows a normal shape since all the bars on the histogram were close to a normal curve. Henceforth, the bell-shaped curve shows a normal distribution (Hair et al., 2014). Hence, this study had not violated normality assumptions even though PLS can work with non-normal data.

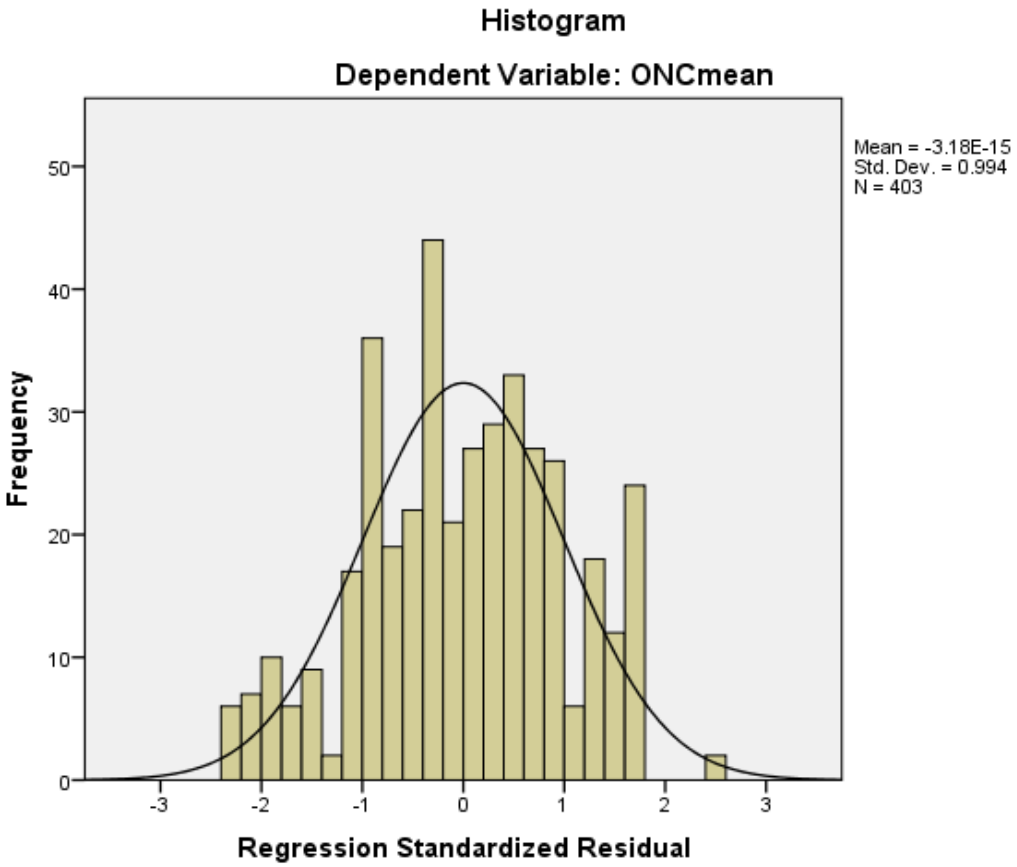

Figure **Error! No text of specified style in document.**.1. Histogram and Normal Probability Plot

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

Scores in figure shows that data is normally distributed. This is also maintained by the normal probability plots where the perceived value for each score is plotted against the expected value from the normal distribution. The straight line in Figure 4.2 shows a normal distribution (Pallant, 2010).

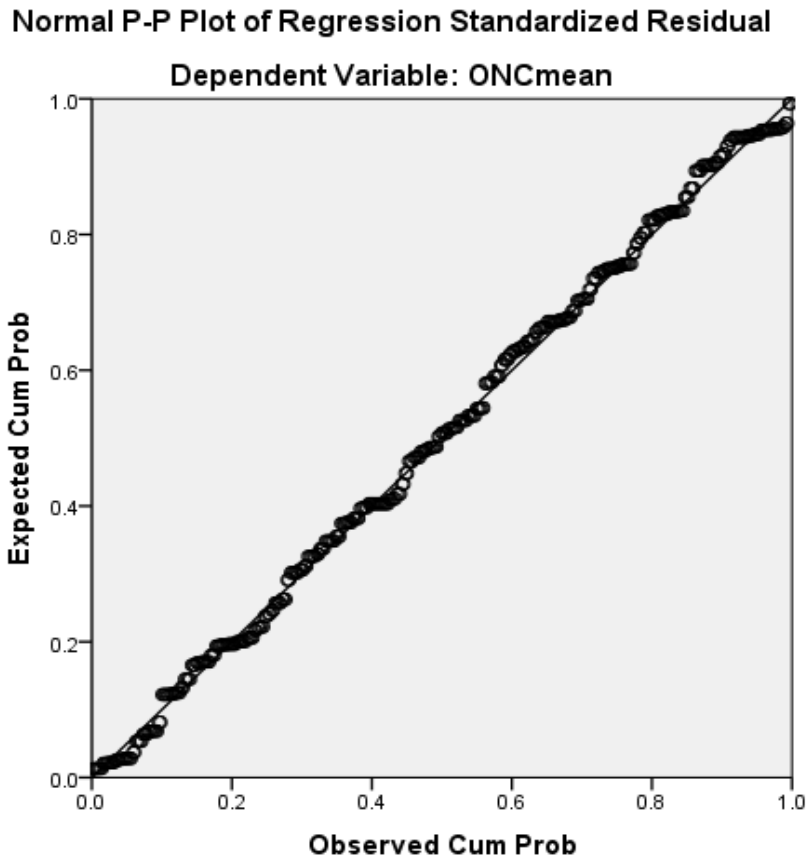

Figure **Error! No text of specified style in document.**.2. Q-Plot

Multi-collinearity Test

To detect multicollinearity, correlation matrix of exogenous latent constructs was used. A correlation coefficient of 0.90 and above shows multicollinearity between exogenous latent constructs. Table 4.12 confirms the correlation matrix of all the exogenous latent constructs.

Table **Error! No text of specified style in document.**.2
Correlation Matrix of the Exogenous Latent Constructs

	IM	EM	SUM	PUM	ECM	ONC
IM	1					
EM	.411**	1				
SUM	.459**	505**	1			
PUM	.437**	.423**	.644**	1		
ECM	.205**	.451**	.339**	.411**	1	
ONC	.338**	.403**	.503**	.444**	.476**	1

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

Results in Table 4.16, the correlations between the exogenous latent constructs were quitelower the suggested threshold values of 0.90 or higher, showing that the exogenous latent constructs were independent and not extremely correlatedin 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, many studies have conducted without considering basics like data screening and preliminary analysis. As a result, this study was conducted to highlight an important part of multivariate analysis which includes calculation of missing values, outliers, normality and multicollinearity. Evidently, these analyses provide better insight into data characteristics of a particular study as well as help in meeting the assumptions of multivariate analysis.

REFERENCES

i. Abdulrauf, A. A., & bin Ishak, S. (2016). An exploratory data analysis on social media and youth online political participation in Nigeria and Malaysia. International Journal of Multidisciplinary Approach & Studies, 3(1).

ii. 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.

iii. Bartlett, J. E. (2001). II, Kotrlik JW, Higgins CC. Organizational research: Determining appropriate sample size in survey research. Inform Tech Learn Perform J, 19(1), 43-50.

iv. Barclay, S., Todd, C., Finlay, I., Grande, G., & Wyatt, P. (2002). Not another questionnaire! Maximizing the response rate, predicting non-response and assessing non-response bias in postal questionnaire studies of GPs. Family practice, 19(1), 105-111.

v. Barnett, V., & Lewis, T. (1994). Outliers in statistical data (Vol. 3, No. 1). New York: Wiley.

vi. Borenstein, M., Rothstein, H., Cohen, J., Schoenfeld, D., & Berlin, J. (2001). SamplePower 2.0. Chicago, IL, SPSS.

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

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

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

x. Cohen, J. (1992). A power primer. Psychological bulletin, 112(1), 155.

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 program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175-191.

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

xv. Gravetter, J. F., & Wallnau, L. B. (2004). Statistics for the Behavior Sciences Thomson: Wadsworth.

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

xvii. 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.

xviii. Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance.

xix. Jebb, A.T., et al., Exploratory data analysis as a foundation of inductive research, Human ResourceManagement

xx. Review (2016), http://dx.doi.org/10.1016/j.hrmr.2016.08.003

xxi. Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: obtaining regression coefficients that are accurate, not simply significant. Psychological methods, 8(3), 305.

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

xxiii. Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. Educational and psychological measurement, 30(3), 607-610.

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

xxv. Memon, B., & Pahore, M. R. (2015). Internet and Online Newspaper Accessing Behaviour of Pakistani Academics: A Survey at Sindh University, Jamshoro. Jurnal Pengajian Media Malaysia, 16(2).

xxvi. Morgenthaler, S. (2009). Exploratory data analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 33-44.

xxvii. 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

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

xxix. Roshan, R., Pervez, M. A., Soomro, B., & Bhutto, R. N. A. (2013). VIOLENCE AGAINST WOMEN: EXPLORING THE EFFECTS OF PAKISTAN TELEVISION URDU DRAMAS VERBAL VIOLENCE ON THE RURAL AND URBAN YOUTH BEHAVIOR. The Women-Annual Research Journal of Gender Studies, 5.

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

xxxi. 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.

xxxii. Silva, M. S., Smith, W. T., & Bammer, G. (2002). Telephone reminders are a cost effective way to improve responses in postal health surveys. Journal of Epidemiology & Community Health, 56(2), 115-118

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

xxxiv. Traina, S. B., MacLean, C. H., Park, G. S., & Kahn, K. L. (2005). Telephone reminder calls increased response rates to mailed study consent forms. Journal of clinical epidemiology, 58(7), 743-746.

xxxv. Uma, S., & Roger, B. (2003). Research methods for business: A skill building approach. book.

xxxvi. Verardi, V., & Croux, C. (2008). Robust regression in Stata.