



A Preliminary Analysis of the Effectiveness of Data in Empirical Research

Bashir Danlami Sarkindaji
(bashirdaji77@gmail.com)

Department of Business Administration, IBB University Lapai, Niger State Nigeria

Abstract

This study X-rays a preliminary investigation of the effectiveness of research data aimed at reducing or eliminating error possibilities bedeviling researchers. The methodology employs survey design through questionnaire administration to elicit responses from a sample of 314 respondents that could inform decisions from the population. Findings revealed that preliminary data screening is purposely to present precise and dependable data. As far as possible, avoid statistical errors, and find a way to deal with everyday research challenges like population, sampling, response rate, data screening and cleansing i.e. missing values, outliers, non-response bias, common method variance, descriptive statistics, data altering, factor analysis, multicollinearity, correlation test, data mining, or developing graphical representation. It is recommended that researchers should as a matter of necessity place pre-test research investigation a priori and center to every research.

Keywords: Preliminary Analysis, Effectiveness, Data, Research, Error Possibilities

1. Introduction

The motive behind preliminary data testing in empirical studies is to present accurate and reliable data. As far as possible, avoid statistical errors, and find a way to deal with everyday research challenges like outliers, missing data, data altering, data mining, or developing graphical representation (Idris & Galoji, 2022). To design, administer and collect dependable data capable of providing a researcher with relevant and reliable findings requires a pre-test investigation of the methodology adopted as well as the research instrument to be used (Samson & Sarkindaji, 2022). Although several studies have argued strongly about the need for preliminary data testing before real data analysis, not many researchers had been observing keenly its due diligence. This manifested in many obstacles in the data analysis processes thereby hindering the effectiveness and reliability of the final output (Hair *et al.* 2019).

Of the varying phases involved in empirical research data collection and analysis though considered one of the most critical components of the research process lack ample attention. Studies have over the years affirmed the relationship between instruments, sources and data analysis techniques employed to have a significant influence on study findings (Solomon & Ringim, 2022). The choice of a statistical technique of data analysis is determined by the research objective, question or hypothesis to be tested (Hair *et al.* 2010).



Analysis of data is as important as any other component of the research process (Pallant, 2011, Sekaran & Bougie, 2010). Hence, to provide accurate, reliable and interpretable results, the method and source of data gathered in the research process must be organized and examined carefully via appropriate planning. Much of the success of the data analysis test lie in the ability of the choice of appropriate method for analysis and observance of a preliminary test (Idrsi & Galoji, 2022; Raykov & Marcoulides, 2006).

Although studies have emphasised extensively the association between data source, reliability and results accuracy, a good number of empirical studies conducted are non-compliant with the norms of data analysis (Hair *et al.* 2019; Sarkindaji, 2011; Newman, 2007; Tabachnick & Fidell, 2007; Yamane, 1967). Further, it is imperative to note that before going into data analysis proper a researcher must place priority on preliminary data investigation to attest to the authenticity and workability of data. Preliminary data investigation, however, emphasised that any data analysis model to be employed must first be justified on pre-test grounds and be consistent with already established theories or results from prior studies in the practical domain under study (Raykov & Marcoulides, 2006). Blind use of data analysis technique could result in a roadmap that leads researchers to deviate from their original substantive goals. However, one would not just expect to enter data into any statistical software and run the analysis to expect magical results, necessary conditions of preliminary investigations must be observed, hence the need for this research. The paper aimed at X-raying the essential preliminary tests expected of every researcher after successful instrument design, data collection, and prior to commencement of data analysis proper. Specifically, the research focused on some grey areas such as population, sampling, response rate, data screening and cleaning such as missing values, detection of outliers, non-response bias, common method variance test, descriptive statistics, factor analysis, multicollinearity, and correlation test. For every empirical study, it is expected that a holistic pre-test investigation must be observed to guarantee accuracy and dependability of output.

2. Literature

Population, Sampling and Response Rate

To ensure generalization and accurate presentation of research findings, it is imperative to determine the population and sample size to be studied in the preliminary investigation stage. The choice of a substantive population is a prerequisite in the determination of an effective sample size. Studies have affirmed the use of empirically established models in determining population and sample and not mere guessing or predictions as used in many studies. For instance, models like Krejcie and Morgan (1970), Yamane (1967) and many others can be deployed. After establishing a determined population, an appropriate sample size, as well as sampling technique, should be used in administering a research instrument e.g. probability sampling that permits all the population elements equal chance of being chosen. A study by Hashim *et al.* (2015) on Institutional Image, utilized a population of 1541 postgraduate students in Malaysia, with an appropriate sample size of 318 students who participated in the survey. 304 questionnaires representing 95.6% response rate were



successfully used. The response rate is appropriate, and perhaps due to the researcher's ability to make follow-up calls, emails, personal visits and SMS texts as a reminder to respondents (Sekaran & Bougie, 2009). An acceptable response rate of 30% is sufficient for survey research (Hair *et al.*, 2010). However, the obtained response rate is consistent with the suggestion of scholars concerning computation of the required sample size for multiple regressions in social sciences (Stevens, 1996; Tabachnick & Fidell, 2007).

Data Screening and Cleaning

Missing Value

Once the administered questionnaires were successfully retrieved by a researcher, the procedure of inserting data into a specific statistical analysis technique commenced e.g. SPSS statistical package. In the preliminary stage, all questions were carefully coded in a fashion that enable the software to identify those that concerns demographic information of respondents, independent, mediating, moderating and dependent variables of the study. However, caution and patience were observed so as to achieve minimal possibility of missing values. Missing values sometimes arise due to respondent's inability to attempt some questions, respond twice to some or attempt some partly. Pallant (2011) stressed that since social science surveys involve mostly dealing with human beings, it is absolutely rare that you obtain complete data from every case, hence it has become necessary for a researcher to check his data file for missing values for each variable. As a precaution to minimize the occurrence of missing data in this study, the researcher and research assistant made sure that at every point of collection careful re-verification is undertaken to ensure questions are adequately answered completely. In cases of serious missing information, we plead with respondents to kindly respond. Even though the task is tough, still it reduces the possibility of missing values or not attempted questions that in the end reduce the quality of study findings. Consequently, as soon as the data had been entered, a preliminary descriptive statistics were observed to find out if there is any missing value. It was observed that missing values exist in the data i.e. about 13.7% but were automatically replaced due to their respective percentage. Studies have suggested when cases of missing values are greater than 15% they should be deleted provided that the returned questionnaires or samples are sufficient for the sample required in the survey (Hair *et al.*, 2010; Tabacnick & Fidell, 2007).

Detecting of Outliers

The next stage of data screening after the detection and replacement of all missing values in a preliminary investigation study is detection of outliers between the responses of respondents. Outliers are cases whose values are well above or below the mainstream of other cases (Pallant, 2011). Detecting outliers assist researchers in identifying cases with extreme scores and whose values may perhaps have considerable influence on the result, when it is well below, well above or present an exceptional combination of values that overlap a number of the study variables. Since the study involves examining the relationships amongst multiple variables through multiple regression analysis, Mahalanobis



Distance (D^2) using the SPSS package was employed to detect the presence of cases with extreme score values. Based on Chi-square table of p-value = 0.001 and a total of 53 questionnaire items any value above 89.27 (Tabachnick & Fidell, 2007) was believed to be an outlier and deleted from subsequent analysis. In all a total of 47 cases representing only 15% of the sampled cases were deleted, with that they cease to constitute a threat to the analysis. Since social science researches involve human behaviours, the possibility of errors of omission or commission attributable to both researcher and respondents becomes inevitable (Pallant, 2011). Therefore, researchers should be careful when treating outliers for you may remove awkward cases to make your results look impressive and also you may lose an opportunity to learn something unexpected (Hair *et al.*, 2019; Pallant, 2011; Dugard *et al.*, 2010). Hence, an acceptable outlier of not more than 15% is considered adequate in empirical research and do not seem to distort the results.

Table 1: Results of Independent Sample T-test for Non-Response Bias Test

Construct	GROUP	N	Mean	Std. Deviation	Levene's Test for Equality of Variances	
					F	Sig.
COR	Early Response	200	2.515	0.767	0.599	0.440
	Late Response	67	2.765	0.694		
CRMO	Early Response	200	1.592	0.564	3.314	0.070
	Late Response	67	1.442	0.509		
KMGT	Early Response	200	1.702	0.588	0.085	0.771
	Late Response	67	1.699	0.638		
TCRM	Early Response	200	2.537	0.679	1.705	0.193
	Late Response	67	2.529	0.730		
CRMM	Early Response	200	1.756	0.474	0.251	0.617
	Late Response	67	1.606	0.460		
CLOY	Early Response	200	4.840	0.823	0.460	0.498
	Late Response	67	4.537	0.745		

Source: Authors Computation

Non-Response Bias

Non-response bias is defined as the changes in the responses among respondents and non-respondents (Lambert & Harrington, 1990). In approximating the likelihood of non-response bias, Armstrong and Overton (1977) recommended a time-trend prediction technique that requires an evaluation of the early and late respondents (i.e., non-respondents). They contended that late respondents exhibit the same features as non-respondents. Therefore, to further reduce non-response bias issues, achieving at least a response rate of 50% was suggested (Lindner & Wingenbach, 2002). In line with Armstrong and Overton's (1977) approach, respondents of this study were categorized into two major groups: group one represented those that responded within 30 days i.e., early respondents, while group two are those that responded after 30 days i.e., late respondents (Vink & Boomsma, 2008). Hence, 200 (74.9%) respondents which constituted a greater part of the sample responded to the questionnaire administered within 30 days, while the



remaining 67, constituting (25.1%) responded after 30 days (see Table 1). Specifically, in order to identify the possibility of any non-response bias in this survey. The table below shows an independent sample t-test observed on each construct comprising COR, CRMO, KMGMT, TCRM, CRMM, and CLOY.

The results of the table above indicated that for all the eight constructs used in this study the equal variance significance values were above 0.05 significance level of Levene's test for equality of variances as recommended by (Field, 2009; Pallant, 2011). Therefore, the assumption of equality of the variances between early and late respondents has not been violated. Thus, it has been established that non-response bias does not pose a threat to this study. In addition, based on the suggestion of Lindner and Wingenbach (2002), since this study had achieved 54.9% response rate, it has further emphasized that the non-response bias issue doesn't pose a threat.

Common Method Variance Test

Common method variance (CMV), also known to as mono method bias, refers to "variance that is attributable to the measurement method rather than to the construct of interest" (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, p. 879). Researchers have generally agreed that common method variance is a major concern for scholars using self-report surveys (Lindell & Whitney, 2001; Podsakoff *et al.*, 2003; Spector, 2006). For example, Conway and Lance (2010) stated that "common method bias inflates relationships between variables measured by self-reports" (p. 325). Similarly, in a meta-analytic review of 55 studies on attitudinal and dispositional predictors of OCB, Organ and Ryan (1995) stated that studies conducted using self-report surveys are associated with spuriously high correlations due to common method variance. This study, however, in an attempt to reduce the effects of CMV adopted numerous procedural solutions (Podsakoff, MacKenzie, & Podsakoff, 2012; MacKenzie & Podsakoff, 2012; Viswanathan & Kayande, 2012; Podsakoff *et al.*, 2003; Podsakoff & Organ, 1986).

In step one, to minimize evaluation anxiety the respondents were made to understand that their responses would be treated with utmost confidentiality and that no wrong or right answer if chosen from the items in the questionnaire. Step two, attempted to reduce common method biases by enhancing the scale items used. This was realised by avoiding the choice of ambiguous terms in the study questionnaire and where such terms become inevitable, simple examples were given. Also to further enhance scale items, questions asked were written in easy, precise and brief language. In addition to the procedural solutions explained above, this study further adopted Harman's single factor test (SFT) suggested by Podsakoff and Organ (1986) to test CMV. Usually, this procedure subjects all the study's variables of interest to exploratory factor analysis and the outcomes of the unrotated factor solution are then assessed to determine the number of factors that are essential to account for the variance in the variables (Podsakoff & Organ, 1986). The major concern of Harman's (1967) SFT assumption is that for a considerable amount of CMV to exist, either a single factor may emerge, or one general factor would account for most of



the covariance in the independent and dependent variables (Podsakoff & Organ, 1986). Following Podsakoff and Organ (1986), all items in this study were subjected to a principal components factor analysis. The outcomes of the analysis revealed eight factors, explaining a cumulative 49.97% of the variance; with the first (largest) factor explaining 18.05% of the total variance, which is less than 50% (Kumar, 2012). Furthermore, the results indicate that no single factor accounted for the majority of covariance in the independent and dependent variables (Podsakoff *et al.*, 2012). Hence, this suggests that common method bias is not a major concern and is unlikely to inflate relationships between variables measured in this study.

Descriptive Statistics

The table below presents the results of the analysis of means and standard deviations of the survey variables. Variables comprising of independent and moderation were measured on a seven-point Likert scale. The mean score criterion used to determine the extent of agreement between the survey variables ranges from 2.99 or less to be “low”, mean score of 3.00 to 4.99 to be “moderate” while mean score 5.00 above to be “high”.

Table 2: Mean Score and Standard Deviation of Study Variables

Variables	Mean (M)	Standard Deviations (SD)
Customer Orientation (COR)	2.515	0.8361
CRM Organization (CRMO)	1.554	0.5535
Knowledge Management (KMGT)	1.721	0.6227
Technology-based CRM (TCRM)	2.500	0.7873
CRM Maintenance (CRMM)	1.693	0.5061

Source: Authors Computation

The above table depicts the means and standard deviations of the five CRM dimensions i.e. COR, CRMO, KMGT, TCRM and CRMM used in this survey. Considering the responses of respondents, COR had a mean score of ($M = 2.515$, $SD = 0.8361$), CRMO ($M = 1.554$, $SD = 0.5535$), KMGT ($M = 1.721$, $SD = 0.6227$), TCRM ($M = 2.500$, $SD = 0.7873$) and CRMM ($M = 1.693$, $SD = 0.5061$). Of all the study variables respondents perceived their implementation by telecommunication operators in Nigeria to be low i.e. none of the mean scores exceeded 2.99 cut-off point. Although CRMO dimension was perceived as having the lowest level of implementation ($M = 1.554$, $SD = 0.5535$) followed by CRMM ($M = 1.693$, $SD = 0.5061$) in that order. Similarly, the moderating effect of switching costs on the relationship between CRM dimensions and customer loyalty to mobile telecommunication services was also perceived to be low by respondents. Overall, results pointed to the fact that CRM application in Nigeria’s mobile telecommunication sector is extremely low, and that is a major factor resulting in the extreme subscribers’ dissatisfaction with mobile services.

Factor Analysis of Measurement Variables

To validate the appropriateness of the study’s sample size and strength of the relationship between the items, exploratory factor analysis was carried out on each independent variable namely COR, CRMO, KMGT, TCRM and CRMM. Overall the variables comprise a total



of 53 items with COR measured by 10 items, CRMO by 10 items, KMGT by 8 items, TCRM by 8 items, and CRMM by 10 items. Principal component analysis based on varimax rotation was carried out to identify those items whose values were suitable for retention for further analysis and who have met the recommended benchmark of above 0.6 and 0.3 for loadings and communalities respectively. After the final principal component test on each variable 14 items were deleted due to the inability to satisfy the recommended threshold values, thus 3 items for COR, 3 items for KMGT, 5 items for TCRM and 3 items for CRMM. All items for CRMO were completely retained. The result of factor analysis of each variable is as follows:

From the table above overall result of factor analysis on CRM dimensions and moderating variable of this study shows that measure of sampling adequacy KMO is 0.837 which has exceeded the value of 0.6, this implies that the data matrix is suitable for factor analysis. Bartlett's test of sphericity of all items revealed a significant Chi-Square value of 6193.27 approx and $p\text{-value} < 0.000$ indicating of howwell the correlated matrix is fit for factor analysis to be carried out. Similarly, communality values for all items after deleting those with loadings below the acceptable minimum threshold was above the recommended 0.3. Consequently, the total variance explained, screen plot and rotated component matrix with varimax rotation analysis revealed the occurrence of six significant factors with their eigen values exceeding 1. These factors were also supported by the direction of the screen plot.

The result of the first factor revealed the factor loading test for COR dimension. The dimension was measured using 10 items and after the initial analysis 3 items (i.e. COR1, COR2, COR9) were deleted due to the inability to load above 0.6 and communalities below 0.3. These items measure respondents' level of agreement to extent of customer orientation application in the mobile telecommunication services sector. The output of 7 items shows factor loadings of above 0.6 with an eigenvalue exceeding 1 (i.e. 9.087), thus indicating that the items are correlated with each other in terms of what they measure. About 21.113% of total variance was explained by items.

The analysis of the second factor shows respondents' opinion on the extent of CRM organization amongst operators in mobile telecommunication services. Based on a 10 items measurement, all indicated values above 0.6 and 0.3 for loadings and communalities respectively, hence the existence of a strong correlation between the items. The percentage of total variation explained by the items is 13.444 with an eigenvalue of 5.781 which is greater than one.

The result of the third factor indicated that KMGT dimension has 8 items. Overall 7 items were retained whose factor loading values range between 0.7 to 0.9 and communalities greater than 0.3. One item (i.e. KMGT6) was deleted due to lower loading of less than 0.6. About 4.916% of total variation explained by the items with an eigenvalue greater than one i.e. 2.114. This factor evaluates respondents' opinion on how the implementation of effective knowledge management in the mobile service industry could help firms retain the loyalty of customers.



The fourth factor shows that TCRM has 8 items out of which only 5 items were successfully retained due to their ability to load above 0.6 with communalities values more than 0.3. All 3 items consisting of TCRM1, TCRM2 and TCRM7 were deleted based on loadings below the minimum threshold of 0.6 and communalities below 0.3. In the same vein, about 4.462% of total variation explained by the items as their eigenvalue exceeded 1 i.e. 1.919. The factor assesses respondents' understanding of the effectiveness of CRM technology application in Nigeria's mobile service industry and how it has achieved in gaining their subscriber's loyalty.

The fifth-factor analysis result was on CRM maintenance. Initially, this dimension has 10 items but was reduced to 7 items after deleting CRMM1, CRMM2, and CRMM6 whose loadings and communalities values are below the recommended 0.6 and 0.3 respectively. About 4.361% of total variation is explained by the items with their eigenvalue greater than 1 i.e. 1.875. The factor measures respondents' perception of the level to which mobile firms are committed to maintaining CRM-related activities in the organizations.

The result of the last factor indicates the analysis for switching cost as a moderating variable amongst CRM dimensions and customer loyalty. Overall a total of 7 items were used and all showed an acceptable value of above 0.6 and 0.3 for factor loadings and communalities respectively. About 3.741% contributed to the total variance explained by the items with an eigenvalue of 1.609. It evaluates respondents' belief as to whether the presence of switching cost could assist in strengthening the relationship between subscribers' satisfaction and their loyalty to mobile telecommunication services.

Multicollinearity Test: Multicollinearity represents the extent to which a variable can be explained by the other variables in the analysis (Hair *et al.*, 2007). As multicollinearity increases, the ability to establish and distinguish a particular variable's effect can be compromised (Sarkindaji, 2011). In detecting the presence of multicollinearity, both tolerance levels and variance inflation factor (VIF) values for each variable were computed. Thus, tolerance level and VIF values of below 0.1 and above 10 respectively indicate the existence of high correlations (Pallant, 2011).

As can be seen above, the result of the analysis shows that both tolerance and VIF values for all predictor variables have met the satisfactory conditions for non-multicollinearity. Tolerance values are all above 0.1 and VIF all below 10, this implies that there is absence of multicollinearity amongst the variables used in this study.

Table 3: Result of CRM Dimensions Variable Test of Multicollinearity

Variables	Tolerance Value	VIF Value
Customer orientation	0.922	1.084
CRM organization	0.358	2.796
Knowledge management	0.376	2.662
Technology-based CRM	0.924	1.083
CRM maintenance	0.488	2.051

Source: Authors Computation



To check for the existence of multicollinearity in this survey, the tolerance value (TV) and variance inflation factor (VIF) is used. Tolerance explains the variability of a specific predictor variable not explained by other predictor variables in the regression model. Tolerance value above 0.1 shows the likelihood of non-multicollinearity and below 0.1 indicates multicollinearity (Pallant, 2011; Sekaran & Bougie, 2010). The VIF values above 10 indicates the possibility of multicollinearity while below 10 shows non-multicollinearity amongst the variables (Pallant, 2011; Sekaran & Bougie, 2010). The residual scatter plot is used to check for linearity amongst the study variables and is presumed that the plots would be rectangularly spread and concentrated or scattered at the centre or nearby 0 point (Pallant, 2011; Flury & Reidwyl, 1998). As deviations from rectangular distribution and centralisation at the centre point occurs, then the assumption of linearity is violated (Pallant, 2011). Outliers can also be discovered from the scatter plot. They are extreme cases of standardized residual above 3.3 or below -3.3 (Tabachnick & Fidell, 2007) which are mostly found with larger samples. The assumption of normality indicates that the disparity between the actual and forecasted independent variable scores achieved normal distribution (Coakes & Ong, 2011). Normality test can be accessed through normal probability plot (NPP) or histogram residual plot (HRP). Based on NPP absence of main deviation from normality occurs where the points lie in a sensible straight diagonal line starting from the bottom left to the top right. HRP explains the nature of distribution of data in situations of continuous variable as well as its correspondence with normal distribution. For the distributions in the histogram to be normal, the bars are to be close to the normal curve, thus the non-violation of its assumption. To check homoscedasticity, as a test of independence of the error term Durbin-Watson was employed. Scholars have suggested a value range of 1.50 to 2.50 as general rule of thumb and when attained, the presumption is not violated. Consequently, all of the assumptions were cautiously observed and followed while using regression analysis in this survey.

Correlation Test

In order to ascertain the relationship between the study variables, a Pearson product-moment correlation test was carried out. Correlation describes the association between two or more continuous variables that are linearly related (Coakes & Ong, 2011). A moderately

Table 4: *Correlation between the Variables*

Variables	COR	CRMO	KMGT	TCRM	CRMM	CSAT	SWC	CLOY
COR	1.000							
CRMO	- 0.175	1.000						
KMGT	- 0.095	0.765**	1.000					
TCRM	0.186	0.106	0.128	1.000				
CRMM	- 0.123	0.672**	0.658**	0.179	1.000			
CLOY	- 0.186	0.338**	0.266**	- 0.011	0.339**	0.242**	0.286**	1.000

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

Source: Authors Computation



weak correlation value of between 0.2 and 0.7 is considered suitable to establish the relationship between variables that are theoretically related (Mayer, 1999). The correlation coefficient is measured using a range of possible values of negative 1 or positive 1 which dictates the direction of the relationship.

From the result of the test in the Table 4.4, 5 predictors i.e. CRMO with $r = 0.338$, $p < 0.01$; KMGT with $r = 0.266$, $p < 0.01$; CRMM, with $r = 0.339$, are positive and significantly related to customer loyalty in Nigeria's mobile telecommunication industry. TCRM variable is affirmed to be insignificantly associated with customer loyalty. Also from the result, CRMM with correlation value of 0.339 contributed most to customer loyalty followed by CRMO, and KMGT. Although COR variable has significant negative influence on customer loyalty with correlation values of $r = - 0.186$, this has been established by numerous studies that Nigerian mobile telecommunication firms are more of product-oriented focused rather than customer-oriented in their operations. In the same vein, some correlations as observed in the table ranges from 0.419 to 0.765 which implies that the association among the variables are not too weak or extremely strong (i.e. > 0.90) to signify a multicollinearity problem. However, a few have indicated extremely weak or zero correlation as their values range between 0.1s06 to $- 0.006$ respectively, thus indicating that the factors are not perfectly correlated.

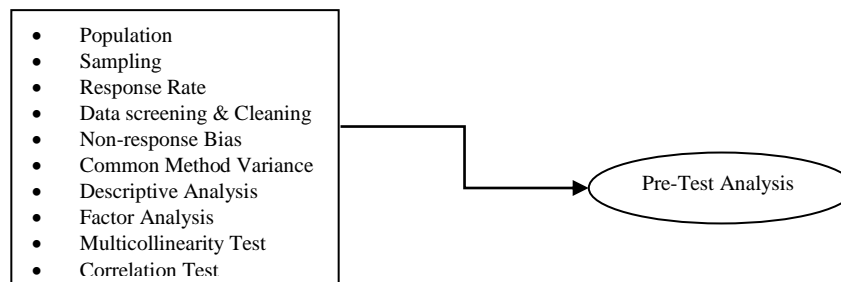


Figure 1: Conceptual Research Model of the Study

3. Methodology

The methodology of this study was a pure review of quantitative empirical studies from varying disciplines to affirm the arguments of scholars that a pre-test analysis is a necessity for effective data analysis. The study is purely survey research based that utilized five independent variables with measurement items selected from previous studies, comprising customer orientation (COR), CRM organization (CRMO), knowledge management (KMGT), technology-based CRM (TCRM), CRM maintenance (CRMM) specifically adapted from Sin & Yim *et al.* (2005). Also, customer loyalty (CLOY) was used as a dependent variable.



The choice of survey study was for three reasons. Survey method utilizes a structured as well as pre-designed questionnaire considered effective in extracting explicit and firsthand information from respondents (Malhotra *et al.*, 2008). The utilization of survey-based research expedites investigation of factors and associations not necessarily measurable directly (Hair *et al.*, 2003) such as customer loyalty as the study focus. Lastly, previous studies have widely employed survey-based research to ascertain consumer brand loyalty (e.g., Brodie *et al.*, 2009; Chaudhuri & Holbrook, 2001; Sirdeshmukh *et al.*, 2002), placing the method suitable for this study.

The quantitative method used to pre-test the analysis of these variables was SPSS software where a survey questionnaire was developed on every construct to measure customers' loyalty to telecommunication services. A population of 1120 telecom managers in Nigeria were chosen given an appropriate sample size of 572 managers who were administered survey questionnaire. In this case, 314 questionnaires representing 54.9% response rate were successfully used for preliminary analysis. After the deletion of outliers, only 267 questionnaires were retained in the final analysis.

4. Results

Testing Hypothesis

To test the proposed hypotheses of this study regression analysis is used. Although two types of regression analysis exist i.e. simple and multiple. Simple regression is used to examine the relationship between one predictor variable and one criterion variable (Dugard, Todman & Staines, 2010). In this study a number of independent variables are used to predict the dependent variable, hence the choice of multiple regression techniques. Multiple regression is a statistical technique employed to examine the association amongst one criterion or dependent variable and more than one independent predictor or variable (Coakes & Ong, 2011; Pallant, 2011). It is aimed to evaluate correlation and provides a more complex examination of the interrelationship between a set of variables. However, one would not just expect to enter data into the SPSS software and run the analysis to expect magical results, necessary conditions must be observed.

Coakes and Ong (2011) identified six underpinning assumptions in the use of regression technique. They involve the test of outliers, multicollinearity/singularity, normality, linearity, homoscedasticity and independence of residuals i.e. evaluated through residual scatter plots and normal probability plot of the regression standardised residual. Similarly, studies have argued for the need for appropriate sample size for multiple regression analysis. For example, approximately 15 respondents per IV (Stevens, 1996), based on formula $N > 50 + 8m$ where $m =$ number of IVs (Tabachnick & Fidell, 2007). With a sample size of 314 and both independent and dependent variables measured on a continuous scale, this study is considered appropriate for regression analysis. Consequently, all of the assumptions were cautiously observed and followed while using regression analysis in this survey.



Customer Relationship Management Dimensions (CRM) and Customer Loyalty

Since this study is aimed at examining the influence of customer relationship management (CRM) dimensions i.e. COR, CRMO, KMGT, TCRM and CRMM on subscribers' loyalty to mobile telecommunication services in Nigeria. Multiple regression analysis was used to evaluate the intensity of each predictor variable's contribution to criterion variable i.e. customer loyalty. Customer loyalty was regressed on respective CRM dimensions to ascertain individual predictor contribution and the output of the final analysis is summarised in the table as follows:

Table 5: Result of Regression Analysis on Customer Relationship Management (CRM) and Customer Loyalty of Unfiltered Data

Predictors (With CRMM)	Model 1 Std. Beta	t-value	Sig.	Decision
CRM Dimensions:				
Customer orientation	-0.032	-0.519	0.302	Not Supported
CRM organization	0.204	2.048	0.021**	Supported
Knowledge management	-0.330	-3.395	0.000***	Not Supported
Technology-based CRM	-0.052	-0.847	0.199	Not Supported
CRM maintenance	0.084	-0.982	0.163	Not Supported
R ²			0.076	
Adjusted R ²			0.058	
R ² Change			0.076	
F-Change			4.270	
Durbin-Watson			1.573	

Note: *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$

Source: Authors Computation

The result above presents the outcome of regression amongst CRM dimensions and customer loyalty before preliminary data filtering. About 7.6% or R^2 of 0.076 of the variation in customer loyalty is predicted by CRM dimensions used in the study (Adj. $R^2 = 0.058$, F-Change= 4.270). Only CRMO was found to positively predicted customer loyalty with beta value of ($\beta = 0.204$, $p < 0.05$), while KMGT established significant negative relationship at beta value of ($\beta = -0.330$, $p < 0.01$). Other dimensions such as COR, TCRM and CRMM were found to be insignificantly related to customer loyalty. Thus, hypothesis H8 is supported while H2, H5, H11, and H14 are not supported.

From the table above, the overall result after the preliminary data filtering revealed that 15.6% i.e. R^2 value of 0.156 of the variation in customer loyalty of telecommunication services was explained by CRM dimensions used in the study model (Adj. $R^2 = 0.140$, F-Change= 0.156, $p < .05$). Of these dimensions only two i.e. CRMO and CRMM have significantly contributed positively to predicting customer loyalty and with beta values of ($\beta = 0.205$, $p < 0.05$) and ($\beta = 0.224$, $p < 0.01$) respectively. The remaining dimensions i.e. COR, KMGT and TCRM were found not to be significant predictors of customer loyalty.



Table 6: Result of Regression Analysis on Customer Relationship Management (CRM) and Customer Loyalty of Filtered Data

Predictors (With CRMM)	Model 1 Std. Beta	t-value	Sig.	Decision
CRM Dimensions:				
Customer orientation	-0.118	-1.997	0.023**	Not Supported
CRM organization	0.205	2.153	0.016**	Supported
Knowledge management	-0.044	-0.470	0.319	Not Supported
Technology-based CRM	-0.045	-0.760	0.225	Not Supported
CRM maintenance	0.224	2.749	0.003***	Supported
R ²			0.156	
Adjusted R ²			0.140	
R ² Change			0.156	
F-Change			9.648	
Durbin-Watson			1.601	

Note: *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$

Source: Authors Computation

Therefore, hypotheses H4 and H13 are supported while hypotheses H1, H7 and H10 are not supported. Although COR was found to establish significant negative relationship at beta value of ($\beta = -0.118$, $p < 0.05$). Hence, the analysis shows strongly the importance of preliminary data filtering and the extent to which a researcher could obtain accurate, reliable and consistent output in the final analysis. About 8% significant influence in customer loyalty was explained by CRM dimensions after preliminary data filtering thereby resulting in additional hypothesis supported.

5. Conclusion and Recommendations

This research specifically centred on the need for pre-test analysis to ascertain the authenticity, reliability as well as accuracy of data in empirical research. The research focused on some grey areas such as population, sampling, response rate, data screening and cleaning such as missing values, detection of outliers, non-response bias, common method variance test, descriptive statistics, factor analysis, multicollinearity, and correlation test. However, the study established that independent variables such as COR, CRMO, KMGT, TCRM, CRMM, and CLOY as dependent variables are all constructs whose items are empirically proven as a measure of their identified constructs. The preliminary investigation established that CRM dimensions as used by prior studies are constructs that significantly influence customers' loyalty to telecommunications services. The findings of the pre-test analysis corroborated with earlier studies by Abdullateef *et al.* (2010) who all found that CRM dimensions as used in this analysis were effective and reliable constructs in the measure of customer relationship management. The regression analysis results for both unfiltered and filtered data show a significant variation of 8% in customer loyalty as explained by CRM dimensions resulting in an additional hypothesis being supported. Further, findings revealed that preliminary investigation prior to real data analysis aid in reducing drastically the possibilities of errors or abnormalities, and showcase all challenges that a researcher may come across in the processes of actual analysis. Hence, studies have



continuously advised on the need for researchers to place a priori pre-test investigation as a condition for building data confidence in research.

Reference

- Abdullateef, A.O., Mokhtar, S., & Yussof, R.Z. (2010). The impact of CRM Dimensions on Call Centre Performance. *International Journal of Computer Science and Network Security*, 10, 184-195.
- Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 396-402.
- Brodie, R., Whittome, J., Brush, G., (2009). Investigating the service brand: a customer value perspective. *Journal of Business Research* 62, 345–355.
- Chaudhuri, A., Holbrook, M.B., (2001). The chain of effects from brand trust and effect to brand performance: the role of brand loyalty. *Journal of Marketing* 65, 81–93.
- Coakes, S. J., & Ong, C. (2011). SPSS version 18.0 for windows: Analysis without anguish.
- Dugard, P., Todman, J., & Staines, H. (2010). *Approaching multivariate analysis: A practical introduction*. London and New York: Routledge Taylor & Francis Group.
- Field, A. (2009). *Discovering statistics using SPSS*: Sage publications.
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (Eds.). (2010). *Multivariate data analysis (7th ed.)*. Uppersaddle River, New Jersey: Prentice Hall.
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (Eds.). (2019). *Multivariate data analysis (7th ed.)*. Uppersaddle River, New Jersey: Prentice Hall.
- Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (Eds.). (2007). *Multivariate data analysis (6th ed.)*. Uppersaddle River, New Jersey: Prentice Hall.
- Hashim, N.A., Abdullateef, A.O., & Sarkindaji, B. D. (2015). The Moderating Influence of Trust on the Relationship between Institutional Image/Reputation, Perceived Value on Student loyalty in Higher Education Institution. *Journal of Review of Management and Marketing*, Vol 5, Issue 3, 5(3), 122-128.
- Idris, A. A. & Galoji, S. I. (2022). Moderating Role of Organizational Justice on the Relationship between Organizational Climate and Employees Turnover Intention: An Empirical Study. *Bayero Business Review*, Vol. 6, No. 1, 1-10.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educ Psychol Meas.*
- Lambert, D. M., & Harrington, T. C. (1990). Measuring nonresponse bias in customer service mail surveys. *Journal of Business Logistics*, 11(2), 5-25.
- Lindner, J. R., & Wingenbach, G. J. (2002). Communicating the handling of nonresponse error in Journal of Extension Research in Brief articles. *Journal of Extension*, 40(6), 1-5.
- Malhotra, N.K., Hall, J., Shaw, M., & Oppenheim, P.P. (2008). *Essentials of Marketing Research: An Applied Orientation*, 2nd ed. Pearson, NSW.
- Neuman, W. L. (2007). *Basics of Social Research: Qualitative and Quantitative Approaches*, Second Edition, Pearson Education, Inc.



- Pallant, J. (2011). SPSS survival manual: A step by step guide to data analysis using the SPSS program (4th Edition). *Australia: Allen & Unwin.*
- Raycov, T. & Marcoulides, G. A. (2006). *A First Course in Structural Equations Modelling*, second Edition.
- Samson, A. T. & Sarkindaji, B. D. (2022). Effects of quality service on customer satisfaction of selected banks in Jalingo metropolis, Taraba State-Nigeria. *International Journal of Intellectual Discourse, Vol. 5 No. 1, pp. 155-168.*
- Sarkindaji, B. D. (2011). An Empirical Investigation of the Relationship between Training and Development and Employee Productivity in Some Selected Local Governments in Niger State. *Lapai Journal of Management Science, Vol. 2 (2), 137-143.*
- Sekaran, U., & Bougie. R. (2010). *Research methods for business; a skill building approach* (5 ed.). *United Kingdom: wiley publishers.*
- Sekaran, U., & Bougie. R. (2009). *Research methods for business; a skill building approach* (5 ed.). *United Kingdom: wiley publishers.*
- Sin, L. Y. M., Tse, A. C. B., & Yim, F. H. K. (2005). CRM: conceptualization and scale development. *European Journal of Marketing, 39(11/12), 1264-1290.*
- Sirdeshmukh, D., Singh, J., & Sabol, B., (2002). Consumer trust, value, and loyalty in relational exchanges. *The Journal of Marketing 66, 15–37.*
- Solomon, A. F. & Ringim, K. J. (2022). An evaluation of organizational effectiveness and efficiency: A case study of Nigerian Television Authority. *International Journal of Intellectual Discourse, Vol. 5 No. 1, pp. 198-225.*
- Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.): Mahwah. NJ: Lawrence Erlbaum.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* Boston: Pearson Education.
- Vink, J. M., & Boomsma, D. I. (2008). A comparison of early and late respondents in a twin-family survey study. *Twin Research and Human Genetics, 11(02), 165-173.*
- Yamane, T. (Ed.). (1967). *Statistics: An introductory analysis* (2nd ed.). *New York: Harper and Row.*
- Yim, F. H., Anderson, R. E., & Swaminathan, S. (2005). Customer relationship management: Its dimensions and effect on customer outcomes. *Journal of Personal Selling and Sales Management, 24(4), 263-278.*