IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN: 2278-487X, p-ISSN: 2319-7668. Volume 23, Issue 4. Ser. II (April 2021), PP 23-34
www.iosrjournals.org

Developing a Research Methodology with the Application of Explorative Factor Analysis and Regression

Dr Priyabrata Panda[1], Sovan Mishra[2], Dr Bhagabat Behera[3]

Abstract
Developing a methodology for an article or thesis has immense importance. Jharotia & Singh (2015) said that methodology eases the work plan of a thesis. Patel & Patel (2019) added that research methodology is the mean to resolve research issues scientifically. The principal purpose of this research work is to develop a methodology when factors are generated after applying explorative factor analysis and regression is to be applied to such factors. This current work also enlightened data screening process and normality assumption. It will help the researchers who are conducting perception studies, behavioural studies etc. with categorical data.
Keywords*: Explorative Factor Analysis, Regression, Methodology, Data Screening, Normality*

Date of Submission: 22-03-2021 Date of Acceptance: 06-04-2021

I. Introduction

The methodology segment of a thesis or article has immense importance as it helps to proceed for further analysis. Jharotia & Singh (2015) emphasised that methodology defines work plan for completion of research. Patel & Patel (2019) added that research methodology is the way to solve research problem scientifically. This part of a research work includes a series of activities in sequence like formulation of statement of the problem, comprehensive design of a research which includes its scope, required data, sampling design etc. Crafting a research methodology chapter largely depends on the type and nature of research. This chapter for time series data is different from panel data. This work is specifically designed for a research work which applied linear regression after generating factors from explorative factor analysis with the help of principal component analysis. The research methodology chapter of an ongoing Ph D thesis of Mishra (2021) has been taken as a background for this research work.

II. Objective of the Study

The main purpose of the study is to develop a methodology of a research work which applies regression on factors extracted from principal component analysis under explorative factor analysis. Data screening and assumption testing are also highlighted.

III. Research Design

Research design answers what, why, where and how (Kothari, 2004). It narrates about the data, type of research, variables, sample etc. which are enumerated below.

Type of Research

Empirical research differs from exploratory research. Researcher has to mention about the type of research here.

Nature of Data

Primary data and or secondary data are applied for hypothesis testing. Moreover, objective of a research specifies the type of data which are to be used. The researcher should theoretically justify the requirement of data in accordance with objectives.

Source of Data

The sources of data must be reliable. Government sources are said to be authentic. CMIE database is widely used by different researchers. Primary data must be collected from targeted population with comprehensive scheduled questionnaire.

---

[1] Assistant Professor of Commerce, Gangadhar Meher University, Amruta Vihar, Sambalpur, Odisha, India, 7978123683, pandapriyabrata@rocketmail.com
[2] Research Scholar, Ravenshaw University, Cuttack, India
[3] Assistant Professor of Commerce, Ravenshaw University, Cuttack, India

Questionnaire and Scale Development

Development of questionnaire depends on objectives. Past literatures must be referred in this regard. The first part must include the demographic information of the respondents. Five-point scale or seven-point scale is desired. Scale development facilitates reliability and validity measures (Tay & Jebb, 2017). Research work of Churchill & Peter (1984), Morgado (2017), Carpenter (2018), Worthington & Whittaker (2006) may be referred for scale development.

Variables

Name of variables and their justification must be mentioned. Dependent variables, independent variables and control variables should be specifically described. Proper citation of earlier literatures must be made.

Sampling Design and Selection of Sample Unit

Probability method of sampling is preferred for collection of primary data. Purpose of the study, population size, sampling error etc. determine the size of sample (Israel, 1992). Selection of sample units from a population has immense importance. The proper determination of sample units reduces standard error and confirms robust inferences. Yamane (1967) formula is quite popular for determining sample size with finite population. Krejcie & Morgan (1970) table has been used by many researchers for sample size determination in "known populations". Other literatures like Jones *et. al.,* (2003), Taherdoost (2017), Blessing & Oribhabor & Anyanwu (2019) have contributed a lot to the sample size determination theory. Sample size also depends on the concerned statistical tool which is being applied in the research work. Like, Kline (1998) infered that sample size may be 10 to 20 times of variables when structural equation modelling is applied. In addition, KMO test confirms sample adequacy when explorative factor analysis is applied.

Scope of the Study

Scope of the study includes geographical scope, parameter scope, time scope etc. The limit and range of the study need to be particularly mentioned.

IV. Data Cleaning

Data cleaning process ensures a comprehensive data set with zero error. It digs out any missing values and outliers. The careless response or unengaged responses are also traced out and ignored before deciding the final data set.

Missing Data

Table 1: Missing Data of Demographic Variables

Valid/Missing	Gender	Age	Income Level	Occupation
Valid	217	217	217	217
Missing	0	0	0	0

Source: Authors own Compilation

It is found in the table that there are no missing values for all the four variables in the above table. Two hundred seventeen responses are observed with no missing values.

Unengaged Responses

There may be possibilities of some careless responses. It may happen that a respondent may select same option for all variables. It will lead to fallacious derivations. Thus, an attempt is made to omit such responses.

Table 2: Unengaged Responses

Responses	Standard Deviation
1.	0.418854
2.	0.455414
3.	0.455562
4.	0.461251
5.	0.466324

Source: Authors own Compilation

Row wise standard deviation is calculated to find such responses. If the value of standard deviation tends to be zero, it means a respondent have selected same option carelessly for all the variables. The above table shows row wise standard deviation of some responses in ascending order which is not zero. Thus, it is inferred that there are no unengaged responses in the data set.

Outliers

Outliers are the extreme values of a data series or data set. If the standardised value exceeds +-3, the data point is said to be an outlier. In the table below, there are no outliers as the standardised value of all the variables are within +-3.

Table 3: Standardised Value and Outliers

Standardised Variables	N	Minimum	Maximum	Desired Value
Z score: X1	22	-1.02203	2.70875	+3 to -3
Z score: X2	22	-.51835	2.22934	
Z score: X3	22	-2.22089	2.12708	
Z score: X4	22	-2.47221	.41740	
Z score: X5	22	-.43054	2.44132	

Source: Authors own Compilation

V. Reliability Analysis

The table below measures reliability statistics of pilot study with the help of Cronbach's alpha (1951). Further, Churchill & Peter (1984) inferred that reliability is necessary for valid research. Such value must be more than .7 (Nunnally, 1978; Nunnally, 1988).

Table 4: Reliability Statistics of Pilot Study

Cronbach's Alpha	No of Items
.775	19

Source: Authors own Compilation

The table above reveals the Cronbach's alpha value .776 for 78 respondents. The value satisfies the recommended criteria.

Table 5: Reliability Statistics of Total Sample

Cronbach's Alpha	No of Items
.786	19

Source: Authors own Compilation

The table shows the reliability statistics through Cronbach's Alpha value for the full sample which consists of 217 responses. Such value is .786 which also satisfies the recommended criterion.

VI. Factor Analysis

Factor analysis may be categorised into following categories.

1. Exploratory factor analysis.
2. Confirmatory factor analysis.
3. Structural equation modelling.

Exploratory factor analysis is analysed in the present study with the help of SPSS software[4].

Kaiser Meyer Olkin (KMO) test of Sampling Adequacy and Bartlett's Test of Sphericity

Table 6: KMO and Bartlett's Test Result

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.749
Bartlett's Test of Sphericity	Approx. Chi-Square	643.795
	df	171
	Sig.	.000

Source: Authors own Compilation

The table above portrays Kaiser-Meyer-Olkin (KMO) and Bartlett statistics. Kaiser-Meyer-Olkin's value measures the adequacy of sampling (Ayuni & Sari, 2018; Hadi, *et. al.,* 2016). KMO statistics is calculated with the following formula (Kaiser, 1970).

$$KMO_j = \frac{\sum_{i\neq j} r_{ij}^2}{\sum_{i\neq j} r_{ij}^2 + \sum_{i\neq j} u_{ij}^2}$$

Where r is simple correlation and u is the partial correlation between i and j.

[4] https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/

Such value ranges from 0 to 1. A value near to one and farer to zero infers that a sample is adequate for factor analysis. To be particular, such value should be more than .5 (Kaiser, 1970; Field, 2000) and .6 (Pallant, 2013, Shree *et al*., (2017). Hutcheson & Sofroniou (1999) opined that "KMO value from .7 to .8 is good, .8 to .9 is great and above .9 is superb". In this research work, KMO value is .749 which is good and within the recommended value according to above all criteria.

On the other hand, Bartlett's Test of Sphericity (Bartlett, 1950,1951) measures the relatedness of variables. The null hypothesis for such test is that variables are uncorrelated. Correlation among some of the variables are required to apply factor analysis. Further, a population matrix is having 1 in diagonal and 0 in non - diagonal, a sample drawn from the population cannot be fit for factor analysis (Tobias & Carlson, 2010). Thus, the null hypothesis must be rejected at least at 5% level of significance (Shree et. al., 2017). Such test is also recommended by Knapp and Swoyer (1967); Gorsuch, (1973). The null hypothesis here is rejected at 1% level of significance as the p value is 0.00 as shown in the above table. It can be inferred that variables are correlated and can be processed for factor analysis.

Factor Extraction

Factor extraction can be carried on by many methods like Scree test (Catell, 1996), Parallel Analysis (Horn, 1965), Principal Component Analysis (PCA) (Pallant, 2013). PCA is widely used. It gives better results (Sehgal *et.al.*, 2014). Mishra *et. al.,* (2017) mentioned that PCA can extract and group inter-correlated variables from a statistical data. Thus, PCA method is applied here.

Communalities

Communalities are the variance or squared factor loadings of variables. It explains the proportion of variability which is explained by the factors and its value is same in spite of using unrotated factor loadings or rotated factor loadings[5]. Its value ranges from 0 to 1. The value closer to one infers that such variable is well explained by the factors. Communalities value decides retaining or removing a variable. But there is a debate in deciding the accepted value of communalities. Osborne (2014) opined that communalities more than 0.4 can be accepted whereas Child (2006) inferred that a variable can be removed if its communality value is less than 0.2. In the present study, variables with communality value nearer 0.5 or more are retained for further analysis.

Table 7: Communalities

SL No	Variables	Initial	Extraction
1.	X1	1.000	.656
2.	X2	1.000	.702
3.	X3	1.000	.643
4.	X4	1.000	.547
5.	X5	1.000	.557
6.	X6	1.000	.286
7.	X7	1.000	.689
8.	X8	1.000	.451
9.	X9	1.000	.585
10.	X10	1.000	.476
11.	X11	1.000	.468
12.	X12	1.000	.484
13.	X13	1.000	.478
14.	X14	1.000	.594
15.	X15	1.000	.493
16.	X16	1.000	.636
17.	X17	1.000	.418
18.	X18	1.000	.635
19.	X19	1.000	.469

Source: Authors own Compilation

[5] https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/factor-analysis/interpret-the-results/all-statistics-and-graphs/.

In the table above, such value for all the variables is more than .4 except X6 which is removed and PCA is again undertaken. The results are shown below.

Table 8: Communalities After Deleting One Variable

SL No	Variables	Initial	Extraction
1.	X1	1.000	.670
2.	X2	1.000	.701
3.	X3	1.000	.640
4.	X4	1.000	.587
5.	X5	1.000	.555
6.	X6	1.000	.657
7.	X7	1.000	.443
8.	X8	1.000	.576
9.	X9	1.000	.487
10.	X10	1.000	.476
11.	X11	1.000	.488
12.	X12	1.000	.489
13.	X13	1.000	.593
14.	X14	1.000	.530
15.	X15	1.000	.700
16.	X16	1.000	.413
17.	X17	1.000	.628
18.	X18	1.000	.470

Source: Authors own Compilation

The variable X16 is having communality .413 which is dropped and EFA is undertaken again. The communality value, Kaiser-Meyer-Olkin Measure of Sampling Adequacy and Bartlett's Test of Sphericity is shown below after dropping such variable.

Table 9: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.720
Bartlett's Test of Sphericity	Approx. Chi-Square	551.584
	df	136
	Sig.	.000

Source: Authors own Compilation

The KMO and Bartlett's Test statistics are matching with the recommended criterion as mentioned above.

Table 10: Communalities of Selected Variables

SL No	Variables	Initial	Extraction
1.	X1	1.000	.676
2.	X2	1.000	.696
3.	X3	1.000	.630
4.	X4	1.000	.580
5.	X5	1.000	.550
6.	X6	1.000	.660
7.	X7	1.000	.486
8.	X8	1.000	.599
9.	X9	1.000	.488
10.	X10	1.000	.478
11.	X11	1.000	.491
12.	X12	1.000	.484
13.	X13	1.000	.616
14.	X14	1.000	.543
15.	X15	1.000	.718
16.	X16	1.000	.626
17.	X17	1.000	.498

Source: Authors own Compilation

The communalities of all the variables in the above table is nearer or more than .5. A total of seventeen variables are processed for further analysis.

In the table below, the total column portrays the eigen value which explains the amount of variance of a factor or component for original variables. Such variance is termed in percentage in the next column followed by the cumulative percentages. A value more than 1 can be accepted for selecting number of components. In this research work, six components are selected for further analysis where eigen value is more than 1. The extracted variance explained for these six factors is 57.64% which more than the recommended value i.e.,50%. Thus, all the seventeen variables explain 57.64% of the total information.

Table 11: Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.37	19.84	19.84	3.37	19.84	19.84	1.87	11.01	11.19
2	1.75	10.33	30.17	1.75	10.33	30.17	1.78	10.50	21.52
3	1.28	7.57	37.75	1.28	7.57	37.75	1.58	9.32	30.84
4	1.20	7.10	44.86	1.20	7.10	44.86	1.57	9.27	40.24
5	1.12	6.60	51.46	1.12	6.60	51.46	1.50	8.83	48.96
6	1.05	6.17	57.64	1.05	6.17	57.64	1.47	8.68	57.64
7	.973	5.72	63.36						
8	.882	5.18	68.55						
9	.748	4.39	72.94						
10	.738	4.34	77.29						
11	.675	3.96	81.25						
12	.644	3.787	85.045						
13	.603	3.544	88.589						
14	.570	3.352	91.942						
15	.512	3.012	94.954						
16	.461	2.713	97.667						
17	.397	2.333	100.000						
Extraction Method: Principal Component Analysis.									

Source: Authors own Compilation

Rotated component matrix, in the table below, explains the factor loadings which is the correlation between the variable and the factor or component. Four factors are grouped under factor 1. Three variables are grouped under factor 2. Third factor includes two variables. Fourth factor includes three variables. Fifth factor includes three variables viz. X13, X14 and X15. X16 and X17 are grouped under the sixth component or factor. Researchers often faced issues while naming such factors or components. With rigorous literature review and theoretical background, factors may be given required names. After generating factors from rotated component matrix, reliability of all variables is again calculated.

Table 12: Rotated Component Matrix

Variables	Component						Factors
	1	2	3	4	5	6	
X1	.745						**Factor 1**
X2	.610						
X3	.587				.471		
X4	.475	.313					
X5		.786					**Factor 2**
X6		.604					
X7		.572			.331		
X8			.798				**Factor 3**
X9			.713				
X10				.827			**Factor 4**

X11				.642			
X12		.346		.423		.367	
X13					.747		Factor 5
X14			.353		.486		
X15	-.327			.318	.377		
X16						.800	Factor 6
X17						.698	

Source: Authors own Compilation

Table 13: Reliability Statistics of Extracted Factors.

Cronbach's Alpha	N of Items
.740	17

Source: Authors own Compilation

The reliability statistics with the help of Cronbach's Alpha of all variables is .740 which is also more than the recommended level.

VII. Test of Normality

Normally distributed data yields robust inferences. Thus, examining such assumption is essential before proceeding to further analysis to avoid inconsistent result (Das & Imon, 2016, Ghasemi & Zahediasl, 2012, Kwak & Park, 2019). Many statistical tests need satisfaction of normality assumption. (Mishra *et. al.,* 2019). Thus, normality of all variables is to be tested.

There are many such tests for measuring normality which are as follows.

Table 14: Different Tests of Normality

Graphical Method	Mathematical Method/ Analytical Method
Histogram	Kolmogorov-Smirnov Test (Kolmogorov, 1933)
Stem-Leaf Plot	Shapiro-Wilk Test (Shapiro and Wilk ,1965)
Box-and-Whisker Plot	Anderson-Darling Test (Anderson and Darling, 1952)
Probability-Probability/Percent-Percent (PP) Plot	D'Agostino-Pearson Omnibus Test (D'Agostino-Pearson, 1973)
Quantile-Quantile (QQ) Plot	Jarqua-Bera Test (Bowman and Shenton, 1975))
Detrended Probability Plot	

Source: Authors own Compilation

For selection of test and method of normality, Das & Imon, 2016 deduced that analytical test of normality is more preferable than graphical test and recommended Shapiro-Wilk Test. Razali & Wah (2011) compared four normality tests with ten thousand samples and also deduced that Shapiro-Will test of normality is most powerful followed by Anderson-Darling test, Lilliefors test & Kolmogorov-Smirnov test. In the present study, Shapiro-Wilk test, Kolmogorov-Smirnov test along with histogram, PP plot, QQ plot and descriptive statistics are applied for assessing normality in different situations. The null hypothesis for normality is that data tend to be normal. Thus, sig./p value more than .05 is desired to accept such hypothesis. Normality of continuous variables are desired.

Table 15: Normality of Variables

Variables	Kolmogorov-Smirnov				Shapiro-Wilk			
	Statistic	df	Sig.	Remark	Statistic	df	Sig.	Remark
V1	.270	7	.133	Normal	.822	7	.066	Normal
V2	.428	7	.000	Non-Normal	.564	7	.000	Non-Normal
V3	.180	7	.200	Normal	.956	7	.780	Normal
V4	.240	7	.200	Normal	.889	7	.272	Normal
V5	.252	7	.199	Normal	.765	7	.018	Non-Normal

Source: Authors own Compilation

Normality with Kolmogorov-Smirnov and Shapiro-Wilk test of different financial parameters is shown in the above table. All the parameters are normal except the second variable under Kolmogorov-Smirnov test. Similar results are also drawn from Shapiro-Wilk test statistics. Further, Shapiro-Wilk test reveals that variable 5 is not normally distributed.

VIII. Regression Assumptions

Normality of Residuals

In linear regression analysis, normality of residuals is an essential assumption. Chan (2004) narrated such assumptions quite clearly. The normality of regression equations is shown below.

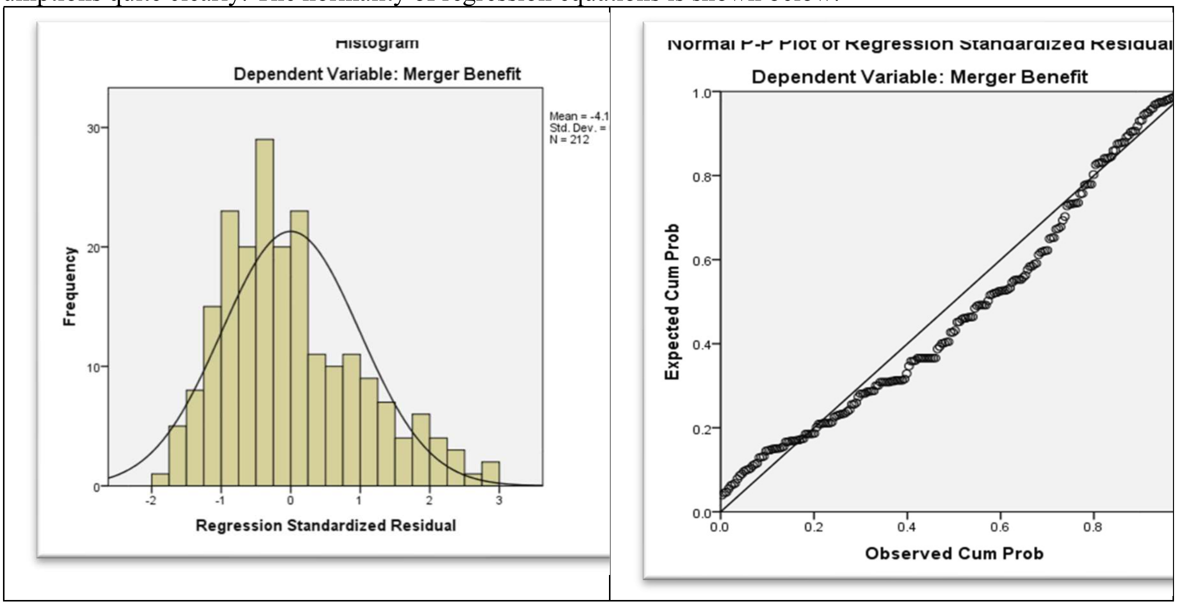

Figure 1: Normality of Residuals with Histogram and PP Plot of Regression Model 1
Source: Authors own Compilation

The distribution of residuals shows that all the data points are within the histogram. The data points are also nearer to the PP Plot. Both the pictures above infer that residuals are normally distributed. The residuals of other models are tested in the similar way and found a normal distribution.

Table 16: Normality of Std. Residuals with Descriptive Statistics

Models	Mean	SD
1	.000	.993
2	.000	.988
3	.000	.991
4	.000	.986

Source: Authors own Compilation

The table above shows normality of residuals of regression models with mean and standard deviation. Mean is zero and standard deviation is very nearer to one. It ensures that residuals of all regression models are normally distributed.

R squared value explains the degree of variability in the dependent variable by the independent variable. It infers about the goodness of fit of the model to the observed data. "An R-squared of 60% reveals that 60% of the data fit the regression model"[6]. Further, 60% of the dependent variable variance has been affecting independent

[6] https://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/#:~:text=The%20most%20common%20interpretation%20of,better%20fit%20for%20the%20model.

variable. But difference of opinion has been observed for the reliability of this test result. It is opined that small R square value does not always mislead and high R-square value may not be always essentially good[7].

Table 17: R- squared and Adjusted R-Squared

Models	R Square	Adjusted R Square
1	.138	.126
2	.143	.122
3	.206	.190
4	.213	.191
5	.159	.143
6	.166	.142

Source: Authors own Compilation

All the models have low R-squared value. Researchers opined that research analysis can be proceeded even with low R-squared value. Akossou & Palm, (2013) opined that R squared is a biased estimate. Filho *et. al*. (2011) substantiated that coefficient of determination fails to draw a meaningful conclusion. Dodge (1999) added that such value can be increased by subsequent addition of variables. Thus, other parameters like F statistics, D-W statistics can be examined in regression analysis along with R-squared value.

Durbin-Watson (D-W) Statistics

Durbin-Watson (1950) Statistics measures auto-correlation. It is said that "the Durbin Watson (DW) statistic is a test for autocorrelation in the residuals from a statistical regression analysis[8]". Maxwell & David, (1995) and White (1992) opined that such statistics should be within 1.5 to 2.5 so that there will be no autocorrelation at lag 1.

Table 18: Durbin-Watson (D-W) Statistics

Models	D-W Statistics
1	1.725
2	1.728
3	2.129
4	2.122
5	1.749
6	1.766

Source: Authors own Compilation

The Durbin Watson statistics of all the models are within the recommended level. Thus, it can be inferred that the model is not affected by auto-correlation.

F-statistics

Table 19: F-statistics and Model Fit

Models	F Value	Sig./P value
1	11.138	.000
2	6.882	.000
3	13.458	.000
4	9.318	.000
5	10.010	.000
6	6.966	.000

Source: Authors own Compilation

7 https://statisticsbyjim.com/regression/interpret-r-squared-regression/

8 https://www.investopedia.com/terms/d/durbin-watson-statistic.asp

F statistics assumes that the regression model may not have predictive efficiency. The null hypothesis is that there are zero regression coefficients. There are no variables affecting the target variable except the intercept or constant. The null hypothesis must be rejected. In the above table, the null hypothesis for all the models is rejected 1% level of significance. The model is said to be fit and having some value of regression coefficients.

Test of Multicollinearity

There should not be multicollinearity issue in a regression model. Multicollinearity is a state where two or more independent variables are highly inter-correlated with each other. The independency of the predictors is violated. It occurs when multiple factors are correlated.[9] It leads to overestimation of the results. Daoud (2017) inferred that the correlation among dependent variables is undesired. Such problem increases the value of standard error of the suffered coefficient. Type II error is invaded to the model. Lindner (2020) found that multicollinearity does not lead to bias however it violates regression assumption. Detection of such issue can be made by made by assessing tolerance level and Variance Inflation Factor (VIF). No multicollinearity is detected when tolerance level is more than .7 and VIF is within 3.

Table 20: Collinearity Statistics

Factors	Tolerance	VIF
Factor 1	.942	1.061
Factor 2	.927	1.079
Factor 3	.968	1.033
Factor 4	.882	1.134
Factor 5	.910	1.099

Source: Authors own Compilation

In the above table, the variables which are derived from explorative factor analysis for regression purpose have no multicollinearity problem as both tolerance and VIF are within the recommended value.

Homoscedasticity or Constant Variance of Residuals

Panda *et al.,* (2020) reported that "there must be constant variance among residuals" when regression analysis is applied. Heteroscedasticity is a problem because ordinary least squares (OLS) regression assumes that all residuals are drawn from a population should have a constant variance (homoscedasticity)[10]. The plots must be scattered across the area, then it is said that there is constant variance of residuals.

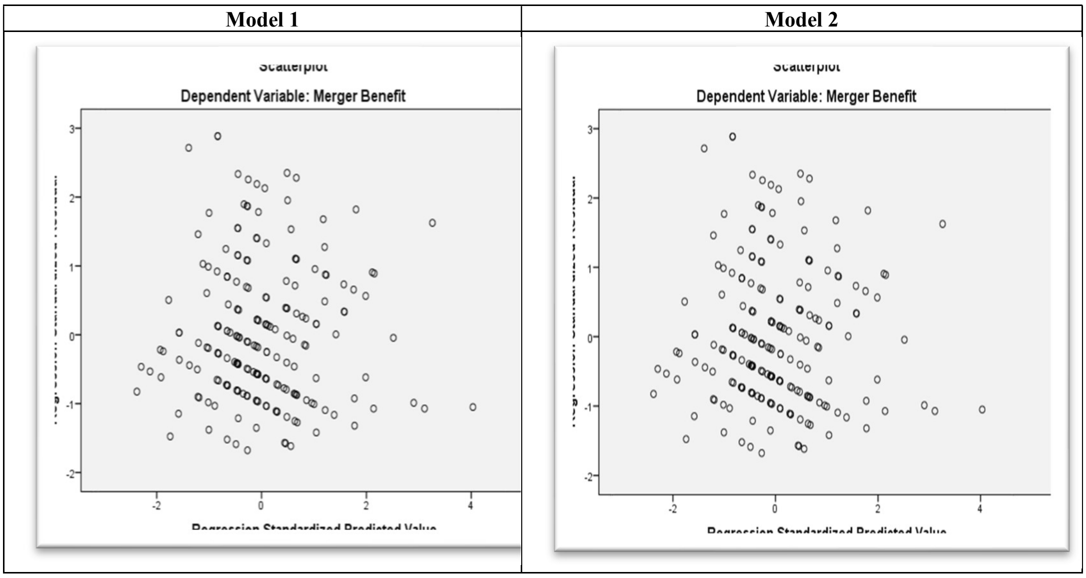

Fig 2: Constant Variance of Residuals

Source: Authors own Compilation

[10]https://statisticsbyjim.com/regression/heteroscedasticityregression/#:~:text=Heteroscedasticity%20is%20a%20problem%20because,should%20have%20a%20constant%20variance.

The table above shows the scatter plot of residuals of model 1 and model 2. The plots are spread across the scatter plot without forming a clear pattern. Thus, it can be inferred that there is a constant variance of residuals. Similar inferences are to be drawn for other models also.

IX. Conclusion

The way or mean of achieving research objectives is carefully decided in this chapter. Research design is framed. A rigorous screening is undertaken for the data set so that unengaged responses are to be ignored. Moreover, outliers are checked with missing values. Statistical tools are selected on the basis of objective and nature of data. Due care is taken to test the assumptions of a tool before applying it.

References

[1]. Akossou, A., Y., J., & Palm, R. (2013). Impact Data Structure on the Estimators R-Square and Adjusted R-square in Linear Regression, *International Journal of Mathematics and Computation*, *20*(3), 84-93.

[2]. Anderson, T. W., and Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *The Annals of Mathematical Statistics, 23*(2), 193-212.

[3]. Ayuni, N.W.D., & Sari, I. G. A. M. K. K. (2018). Analysis of factors that influencing the interest of Bali State Polytechnic's students in entrepreneurship. *Journal of Physics: Conference Series*, 1–10.

[4]. Bartlett, M. S. (1950). Tests of significance in factor analysis, *The British Journal of Psychology*, *3* (Part II), 77-85.

[5]. Bartlett, M. S. (1951). The effect of standardization on a Chi-square approximation in factor analysis. *Biometrika*, *38*(3/4), 337-344.

[6]. Bowman, K. O., and Shenton, B. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika 64*: 243-50.

[7]. Carpenter, S. (2018). Ten Steps in Scale Development and Reporting: A Guide for Researchers. *Communication Methods and Measures*, *12*(1), 25–44.

[8]. Catell, R. R. (1966). The screen test for number of factors, *Multivariate Behavioral Research*, 1, 245-276.

[9]. Chan, Y. H. (2004). Biostatistics 201: Linear Regression Analysis. *Singapore Med J*, *45*(2), 55–61.

[10]. Child, D. (2006) The essentials of factor analysis. Continuum, London, 1-106.

[11]. Churchill, G. A., & Peter, J. P. (1984). Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research*, *XXI*, 360–375.

[12]. Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests, *Psychometrika*, *16* (3) 297–334, doi:10.1007/bf02310555.

[13]. D'Agostino R, and Pearson E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika*. *60*(3), 613-622.

[14]. Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 1–6. doi :10.1088/1742-6596/949/1/012009

[15]. Das, K. R., & Imon, A. H. M. R. (2016). A Brief Review of Tests for Normality. *American Journal of Theoretical and Applied Statistics*, *5*(1), 5–12.

[16]. Dodge, Y. (1999). Analyse De Regression Appliquee, Dunod, Paris.

[17]. Durbin, J. & Watson, G. S. (1950). "Testing for Serial Correlation in Least Squares Regression, I". *Biometrika*. *37* (3–4): 409–428.

[18]. Field, A. (2000). Discovering Statistics using SPSS for Windows. London – Thousand Oaks – New Delhi: Sage publications.

[19]. Filho, D., B., F., Silva, J., A., & Rocha, E. (2011). What is R2 all About? *Leviathan – Cadernos de Pesquisa Política*, (3), 60-68.

[20]. Ghasemi, A., & Zahediasl, S. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology & Metabolism*, *10*(2), 486–489.

[21]. Gorsuch, R. L. (1973). Using Bartlett's Significance Test to Determine the Number of Factors to Extract. *Educational and Psychological Measurement*, (33), 361–364.

[22]. Hadi, N. U., Abdullah, N., & Ilham, S. (2016). An Easy Approach to Exploratory Factor Analysis: Marketing Perspective. *Journal of Educational and Social Research*, *6*(1), 215–223.

[23]. Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, (30), 179-185.

[24]. Hutcheson, G. D., and Sofroniou, N. (1999). The Multivariate Social Scientist: an introduction to generalized linear models. Sage Publications.

[25]. Israel, G., D. (1992). Sampling the Evidence of Extension *Program Impact. Program Evaluation and Organizational Development*, IFAS, University of Florida. PEOD-5. October.

[26]. Jharotia, A. K., & Singh, S. (2016). Use of Research Methodology in Research: An Overview. *International Journal of Social Science, Journalism & Mass Communication*, *2*(2), 44–51.

[27]. Kaiser, H. (1970). A Second-Generation Little Jiffy, *Psychometrika*, *35*, 401-415.

[28]. Kaiser, H. (1974). An index of factorial simplicity. Psychometrika, *39*, 31-6.

[29]. Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*. New York.

[30]. Knapp, T. R. and Swoyer, V. H. (1967). Some empirical results concerning the power of Bartlett's test of the significance of a correlation matrix. *American Educational Research Journal*, *4*, 13-17.

[31]. Kolmogorov, A. (1933) Sulla determinazione empirica di una legge di distribuzione.'' *G. Ist. Ital. Attuari*, *4*, 83–91.

[32]. Kothari C. R. (2004). *Research Methodology: Methods and Techniques*. New Age International Publisher, Second Edition, 1-401.
[33]. Krejcie, R. V., & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*, *30*(3), 607–610.
[34]. Kwak, S. G., & Park, S. (2019). Normality Test in Clinical Research. *Journal of Rheumatic Diseases*, *26*(1), 5–11.
[35]. Lindner, T., Puck, J., & Verbeke, A. (2020). Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies. *Journal of International Business Studies*, *51*, 283–298. https://doi.org/10.1057/s41267-019-00257-1.
[36]. Maxwell, L., K and & David C., H. (1995): The Application of the Durbin-Watson Test to the Dynamic Regression Model Under Normal and Non-Normal Errors, *Econometric Reviews, 14*(4), 487-510.
[37]. Mishra (2021). *An analysis of Merger and Acquisition in Indian Banking Sector*. (Unpublished doctoral dissertation). Ravenshaw University, Cuttack, India.
[38]. Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, *22*(1), 67–72.
[39]. Mishra, S., Sarkar, U., Taraphder, S., & Datta, S. (2017). Principal Component Analysis. *International Journal of Livestock Research*, *7*(5), 60–78.
[40]. Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexao e Critica*, *30*(1), 1–20.
[41]. Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill Publishing.
[42]. Nunnally, J. C. (1988). *Psychometric Theory*. New Jersey: McGraw-Hill, Englewood Cliffs.
[43]. Oribhabor, C. B., & Anyanwu, C. A. (2019). Research Sampling and Sample Size Determination: A practical Application. *Journal of Educational Research (Fudjer)*, *2*(1), 47–57.
[44]. Osborne, J. W. (2016 or 2014). *Best Practices in Exploratory Factor Analysis*. Scotts Valley, CA: CreateSpace Independent Publishing. ISBN-13: 978-1500594343,
[45]. Pallant, J. (2013). *SPSS Survival Manual. A step-by-step guide to data analysis using SPSS,* 4th edition. Allen & Unwin, www.allenandunwin.com/spss.
[46]. Panda, P., Das, K., K. & Mohanty, M., K. (2020). Direct Tax Reform in India: An Impact Analysis with Special Reference to Government Revenue. *The Orissa Journal of Commerce*, *41*(1), 66-86.
[47]. Patel, M., & Patel, N. (2019). Exploring Research Methodology: Review Article. *International Journal of Research and Review*, *6*(3), 48–55.
[48]. Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.
[49]. S. S. Shapiro, & M. B. Wilk. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika.52*, (3/4), 591-611.
[50]. Sehgal, S., Singh, H., Agarwal, M., & Shantanu, V. B. (2014). Data Analysis Using Principal Component Analysis. *International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, 45–48.
[51]. Shree, S. V., Pugazhenthi, R., & Chandrasekaran, M. (2017). Statistical Investigation of the Performance Evaluation in Manufacturing Environment. *International Journal of Pure and Applied Mathematics*, *114*(12), 225–235.
[52]. Taherdoost, H. (2017). Determining sample size; How to calculate survey sample size. *International Journal of Economics and Management Systems*, *2*(2), 237–239.
[53]. Tay, L., & Jebb, A. (1996). Scale development. In *Journal of Health and Social Policy* (Vol. 8, Issue 1). https://doi.org/10.1300/J045v08n01_02
[54]. Tay, L., & Jebb, A. (2017). Scale Development. In S. Rogelberg (Ed), *The SAGE Encyclopedia of Industrial and Organizational Psychology*, 2nd edition. Thousand Oaks, CA: Sage.
[55]. Tobias, S., & Carlson, J. E. (2010). Brief Report: Bartlett' S Test of Sphericity and Chance Findings in Factor Analysis. *Multivariate Behavioral Research*, 375–377.
[56]. White, K. (1992). The Durbin-Watson Test for Autocorrelation in Nonlinear Models. The Review of Economics and Statistics, *74*(2), 370-373.
[57]. Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, *34*(6), 806–838.
[58]. Yamane, T. (1967). Statistics, An Introductory Analysis, 2nd Ed., New York: Harper and Row.

Panda, P., Mishra, S. & Behera, B. (2021). Developing a Research Methodology with the Application of Explorative Factor Analysis and Regression. *IOSR Journal of Business and Management (IOSR-JBM)*, *23*(04), 23-34.