Original Research Article

Application of Sensitivity & Principal Component Analyses for Modelling of Safety Parameters for Oil & Gas Companies in Niger Delta

Abstract

This study assesses the sensitivity analysis in classical multiple regression models as well as those of the principal component regression, PCR models. The study is a cross-sectional research design and data collection was through questionnaire distribution to employees of Oil & Gas indigenous (OGI) and multinational (OGM) companies in Niger Delta. The questionnaire was designed to obtain information from respondents on 7 safety constructs as independent variables against employee productivity as dependent variable. The questionnaire was designed using 5-point Likert scale, strongly agree, agree, disagree, strongly disagree and undecided with corresponding weighting of 5 – 1. The descriptive statistics featured the calculation of weighted arithmetic mean which served as input data for the calibration of both simple linear regression models, multiple principal component regression models and classical multiple regression models, respectively. SPSS was employed to calculate Cronbach's alpha coefficients, assessing the internal consistency of the research constructs; while XLSTAT, was used for principal component analysis, PCA & principal component regression, PCR. The relationship between the slope of simple linear regression equation for 7 independent variables/constructs were examined with respect to the corresponding goodness of fit. The result showed that the simple linear regression models of the Employee Productivity as dependent variable against each of the 7 independent variables (Employee Involvement, Management Commitment, Safety Compliance, Safety Knowledge, Safety Participation, Safety Promotional Policies & Safety Training) yielded positive slope values for both OGI & OGM. In contrast to the classical multiple regression model, for which the total goodness of fit value for the 7 constructs for OGI = 44.9 & 48.4% for OGM; the simple linear models for the same 7 construct yielded total goodness of fit of 81.2 & 171.1% for OGI & OGM, respectively. The total goodness of fit values for the simple linear regression were not only greater than those of classical multiple regression models, but exceeded 100% limit in classical multiple regression models. The principal component regression models established in this study is a variable reduction technique. Given that all the R^2 values of PCR models (i.e. 3 & 4 constructs) are superior to the equivalent R^2 values extracted from sensitivity values of classical multiple regression models; the reason is on the manner of solving the resulting simultaneous equations arising from those of PCR and classical multiple regression. The PCR modelling adopts the optimization technique, using XLSTAT version 16 in this study; while the classical multiple regression modelling utilizes the Gaussian method of solving the simultaneous linear equations. Therefore, the optimization method of solution of simultaneous equations using XLSTAT software is recommended.

Key words: **Sensitivity Analysis, Principal Component Regression Models, Goodness of fit Employee Productivity, Safety Variables, Oil & Gas Companies, Niger Delta**

1. INTRODUCTION

The Niger Delta, south of Nigeria is home to oil & gas operations which dominate the economy and it is pertinent to access the safety of operations of the industries. These industries are often associated with high-risk environments, making safety an integral factor not only for employee well-being but also for enhancing productivity (Sala et al., 2024). Safety programmes are designed to mitigate risks, reduce accidents, and create a safe working environment, which, in turn, contributes to employee productivity (Burke et al., 2006).

This study focuses on seven critical safety parameters: Employee Involvement (EI), Management Commitment (MC), Safety Knowledge (SK), Safety Promotional Policies (SPP), Safety Training (ST), Safety Compliance (SC) and Safety Participation (SP). These constructs or parameters serve as the foundation for understanding the relationship between safety programmes and employee productivity (Sala et al. 2024):

1. Employee Involvement (EI)

Employee involvement in safety-related decisions fosters a sense of ownership and responsibility for maintaining a safe work environment. Hale et al. (2010) emphasized that involving employees in safety planning and decision-making processes results in higher levels of compliance and productivity. In the context of the Niger Delta, where industrial activities pose significant safety risks, encouraging employee involvement can enhance both safety outcomes and operational efficiency.

2. Management Commitment (MC)

Management commitment is crucial to fostering a safety culture within an organization. Studies have shown that when management actively demonstrates a commitment to safety, employees are more likely to adhere to safety policies and practices, leading to fewer accidents and higher productivity (Hofmann & Morgeson, 1999). In the high-risk industries of the Niger Delta, such commitment is essential for establishing trust and ensuring that safety standards are upheld consistently.

3. Safety Knowledge (SK)

Safety knowledge refers to the understanding that employees have regards to safety protocols and risks in their work environment. Griffin and Neal (2006) found that employees with higher safety knowledge are more efficient in their work, as they can identify and avoid potential hazards. In high-risk sectors such as oil and gas, well-informed employees contribute not only to a safer workplace but also to improved productivity, as they can perform their tasks with greater confidence and efficiency.

4. Safety Promotional Policies (SPP)

Safety promotional policies are strategies implemented by organizations to encourage safe behavior among employees. These policies often include rewards for adhering to safety standards or public recognition of safe practices. Vredenburgh (2002) highlighted that safety promotional policies create a positive safety climate, which leads to fewer workplace injuries and improved productivity. In regions like the Niger Delta, where hazardous work environments are common, effective promotion of safety awareness is vital for reducing accident rates and ensuring continuous operations.

5. Safety Training (ST)

Safety training equips employees with the knowledge and skills necessary to perform their tasks safely. While training is essential for minimizing workplace accidents, it can also impact productivity if not implemented effectively. Burke et al. (2006) noted that well-structured safety training programmes lead to better safety outcomes and improved employee performance. However, excessive or poorly timed training sessions can detract from productive time. In the Niger Delta, where industries operate under tight deadlines, striking a balance between training and productivity is key.

6. Safety Compliance (SC)

Safety compliance refers to the extent to which employees follow established safety regulations and procedures. Compliance is a direct indicator of the effectiveness of safety programmes. Clarke (2006) found that strict adherence to safety rules reduces workplace incidents, which translates to fewer work interruptions and enhanced productivity. In industries where safety risks are prevalent, such as oil and gas, compliance is non-negotiable for maintaining operational efficiency.

7. Safety Participation (SP)

Employee participation in safety-related activities is another critical factor that influences productivity. Griffin and Neal (2000) demonstrated that when employees actively engage in safety programmes, such as through reporting hazards or suggesting improvements, there is a direct impact on reducing safety risks and improving overall organizational performance. In the Niger Delta, where industries rely heavily on manual labor, fostering employee participation in safety decisions can significantly enhance productivity.

2. MATERIALS AND METHODS

2.1 Study Area

The study area is the delta of the Niger River at the Gulf of Guinea on the Atlantic Ocean in Nigeria; Latitude: 5° 19' 20.40" N, Longitude: 6° 28' 8.99" E, (Figure 1). It has a population of about 30 million people. It houses oil wells and operational locations of most oil and gas companies in Nigeria. The core states are Akwa Ibom, Bayelsa, Delta, Imo & Rivers where the copies of our questionnaires were distributed to Oil & Gas workers (Indigenous & Multinational Industries).

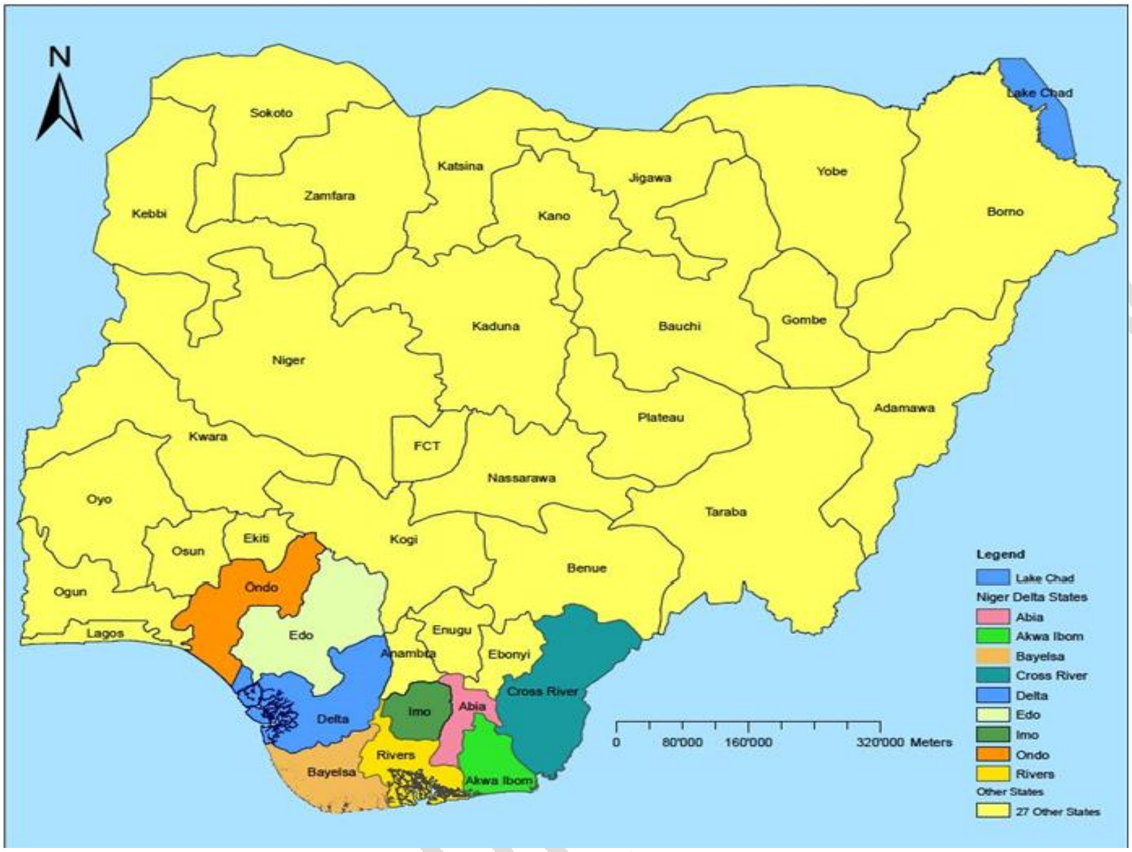

Figure 1: Map of Nigeria showing the Niger Delta Study Area

Structured questionnaire served as the major instrument to facilitate data collection through responses on assessment of safety programmes on employee's productivity in selected Oil & Gas industries in Niger Delta Area. Sample size of 390 was determined using **(Cochran, 1963)**, Equation (1):

$$N = \frac{Z^2 p(1-p)}{T^2} \quad (1)$$

Where N = Sample size; Z = the abscissa of the normal curve that cuts off an area α at the tails; (1 – p) = the desired confidence level (i.e. 95%); T= Tolerance error (or the desired level of precision); and p = prevalence or estimated proportion of an attribute that is present in the population without considering the finite population correction factor (fpc).

Three hundred & ninety (390) copies of the structured questionnaire were administered by random sampling to the respondents for data collection in the selected Oil & Gas companies in Niger Delta region using Google Forms, an online survey tool. The purposive method was adopted for the selection of Oil & Gas companies in Niger Delta region.

Most of the questions were structured using Likert scale (5-Strongly agree, 4- agree, 3-Disagree, 2-Strongly disagree, 1-Undecided). A retrieval of 350 out of the 390 questionnaires were made given rise to 89.74% retrieval rate.

The study is focused on the comparative analyses of safety programmes on employee's productivity in selected Oil & Gas industries in Niger Delta Area. Two categories of companies were assessed, namely: Multinational and Indigenous Oil & Gas companies.

2.2 QUESTIONNAIRE INSTRUMENTS

Information captured in the questionnaire included Gender, Age, Level of education, Marital status and Years of experience. The administered questionnaire also includes information on seven constructs considered herein as independent variables: Management Commitment, Safety Participation, Safety Compliance, Safety Promotional Policies, Safety Training, Safety Knowledge, Employee Involvement and Employee Productivity considered as dependent variable.

There are 7 broad questions on Management Commitment, 6 questions on Safety Participation, 5 questions on Safety Compliance, 6 questions on Safety Promotional Policies, 5 questions on Safety

Training, 6 questions on Safety Knowledge, 4 questions on Employee Involvement and 5 questions on Employee Productivity using the 5-point Likert Scale (Cochran, 1963).

2.3 DATA ANALYSIS AND PROCEDURES

The statistical analyses in this research were conducted using two powerful statistical software packages: SPSS version 26 and XLSTAT version 16. The Scale Reliability Analysis tool in SPSS was employed to calculate Cronbach's alpha coefficients, assessing the internal consistency of the research constructs. XLSTAT, as an Excel add-in, complemented SPSS by offering advanced statistical capabilities. The Principal Component Analysis (PCA) module in XLSTAT was utilized, incorporating its built-in Bartlett's sphericity test and Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The software's rotation options, particularly the Varimax rotation, were employed to enhance the interpretability of the factor structure. Also, the multiple regression were carried out using the questionnaire weighted responses as input data for the 7 constructs as independent variables against the Employee Productivity as dependent variable; and the resulting goodness of fit for each developed model was used as a basis for model performance evaluation.

2.4 The Concept of Goodness of fit in Regression Modelling

To carry out the test of goodness of fit for multi–variable regressions, the starting point is the sum of the square of the residuals, which is independent of the value of the unknown common variance (Herren, 2023). In evaluating a fitted regression equation our interest lies in three areas of variations namely, **total variation, explained and unexplained variations (Davis, 1973; Nwaogazie, 2021)**. The three types of variations are related as given in Equation (2):

$$\sum(y-\bar{y})^2 = \sum(y-y_{est})^2 + \sum(y_{est}-\bar{y})^2 \quad (2a)$$

Or $$SS_T = SS_D + SS_R \quad (2b)$$

in which:

y= a variable; $\bar{y}$ = mean value; y_{est} =y-estimate.

(SS_T)= Total sum of squares =$\sum(y-\bar{y})^2$;

(SS_D)= Sum of squares due to deviation or residual & outliers = $\sum(y-y_{est})^2$;

(SS_R)= Sum of squares due to regression= $\sum(y_{est}-\bar{y})^2$;

Dividing Equation (2b) by SS_T yields a quotient equation as

$$R^2 \text{ or } \frac{SS_R}{SS_T} = 1 - \frac{SS_D}{SS_T} \quad (3)$$

According to MUC (2023) no one has come up with a perfect measure of goodness of fit for statistical models, although there has been and continues to be much research in the area. However, a given linear regression may be accorded with the following assumptions: i) exact linearity of all relationships, ii) normality of residuals or errors from the model, iii) constant residual variance throughout the range, etc. For Equation (3), if SS_D = SS_T, then R^2 = 0, and model is not useful. If SS_D = 0, then R^2 = 1, and model fits all points perfectly. In other words, the range of goodness of fit varies from 0 – 1 (i.e. 100%).

3. RESULTS AND DISCUSSION

3.1 Respondents Distribution in Oil & Gas Industries

The distribution of respondents across different company types is presented in Table 1. A total of 255 out of 350 retrieved questionnaires are for Oil & Gas industries used specifically for all the model development exemplified in this study i.e. Multinational = 158 and Indigenous = 97, respectively. The remaining 95 retrieved questionnaires are for the Construction industries which are not utilized for the purpose of this manuscript.

The extracted responses from the 255 returned questionnaires (correctly filled), served as input data for computation of weighted averages for each question. These averages were employed in multiple regression model development for both indigenous and Multinational Oil & Gas companies.

Table 1: Respondents in the various companies

Company Type	Number	Percentage (%)	Cumulative Percentage (%)
Oil and Gas (Indigenous)	97	38.0	38.0
Oil and Gas (Multinational)	158	62.0	100.0

Total	255	100

3.2 Reliability of Constructs

The reliability of the safety programme constructs, and productivity measure was assessed using Cronbach's alpha, with the results presented next. The analysis revealed good internal consistency across all measures, with Cronbach's alpha values ranging from 0.630 to 0.916. Management Commitment demonstrated the highest reliability with a Cronbach's alpha of 0.916 (standardized 0.920), followed closely by Safety Training at 0.895 (standardized 0.898). Safety Knowledge and Safety Compliance both showed strong reliability with Cronbach's alpha values of 0.867 (standardized 0.877) and 0.862 (standardized 0.861), respectively. Employee Involvement and Safety Promotional Policies exhibited good reliability, both with Cronbach's alpha values of 0.849 (standardized 0.850). Safety Participation, while having the lowest Cronbach's alpha of 0.630, still indicated good reliability when considering its standardized value of 0.741. The Employee Productivity measure also demonstrated good reliability with a Cronbach's alpha of 0.785 (standardized 0.787).

These results are particularly noteworthy as they all meet or exceed the generally accepted threshold of 0.70 for Cronbach's alpha, with most constructs showing excellent internal consistency ($\alpha > 0.80$). Even Safety Participation, which had the lowest unstandardized alpha, still demonstrates good reliability when its standardized value is considered. This consistency across all constructs suggests that the measurement scales used for assessing safety programme constructs and employee productivity in this study are robust and reliable.

The high reliability scores across these constructs indicate that the questionnaire items within each construct are closely related and effectively measure the same underlying concept. This strong internal consistency enhances the validity of the study's findings, as it suggests that the data collected through these measures are likely to provide a consistent and accurate representation of the safety programmes and productivity in the selected industries in Niger Delta Area.

3.3 Principal Component Analysis

The principal component analysis (PCA) was applied to assess the relationship between safety programme construct. The analyses are presented in Tables 2 & 3. The KMO test produced a value of 0.859, signifying a strong pattern in the data, thus confirming that PCA is suitable for this analysis.

Next, the eigenvalues and proportion of variance were calculated to identify the number of components to retain. The results showed that the first principal component (PC1) had an eigenvalue of 4.277 and accounted for 53.46% of the variability in the data. The second component (PC2) had an eigenvalue of 1.230 and accounted for 15.37% of the variability, resulting in a cumulative variability of 68.83% before rotation. These two components were retained based on the eigenvalue-one criterion, as they both had eigenvalues greater than 1, and the cumulative proportion of variance was above the common threshold of 70%.

The retained components were subjected to varimax rotation to enhance interpretability. After the rotation, the first two components explained 61.49% of the variability in the data. Component 1 had a variability of 36.03%, while Component 2 accounted for 25.46%. Factor loadings after the varimax rotation indicated that four organizational factors loaded significantly on the first component, while three loaded on the second component. Factors with loadings greater than 0.50

were considered significant, suggesting that these variables were highly correlated with the respective components.

Table 2: Eigenvalue and Proportion of Variance before and After Varimax rotation

		Before Varimax Rotation		After Varimax Rotation	
Components	Eigenvalue	Variability (%)	Cumulative Variability (%)	Variability (%)	Cumulative Variability (%)
PC1	4.277	53.456	53.456	36.03	36.03
PC2	1.230	15.374	68.830	25.461	61.49
PC3	0.623	8.158	76.988	15.498	76.98

Table 3: Factor Loading after varimax rotation

Safety Constructs	D1	D2	D3
MC	0.863	0.165	0.078
SP	0.189	0.791	0.199
SC	0.309	0.809	0.137
SPP	0.823	0.186	0.182
ST	0.759	0.385	0.160
SK	0.226	0.683	0.441
EI	0.827	0.203	0.112
EP	0.132	0.198	0.954
Industry you work in? -Construction (Indigenous)	-0.410	-0.036	0.003
Industry you work in? -Oil and Gas (Multinational)	0.415	-0.097	0.016
Industry you work in? -Construction (Multinational)	-0.234	0.101	0.108
Industry you work in? -Oil and Gas (Indigenous)	0.001	0.044	-0.113

As presented in Table 3 column 2 for D1, column 3 for D2 & column 4 for D3, two separate relationships are possible, viz:

i. Employee Productivity (EP) = function (MC, SPP, ST & EI) (4a)

ii. EP = function (SP, SC & SK) (4b)

The multiple regression models of Equations 4a & b are calibrated using response values as input data for the various constructs and for two different companies, viz: Oil & Gas Indigenous (OGI) and Oil & Gas Multinational (OGM).

Table 4 exemplifies response distribution of questions taken from a selected construct (Employee Productivity, EP) for computation of weighted mean with respect to a given question and this procedure is repeated for all questions in each of the remaining 6 constructs. The participants responses with respect to the seven organizational variables, given that the responses are of five options (Likert scale) with individual weighting, e.g. Strongly Agree for W1= 5, Agree, W2=4; Disagree, W3= 3; Strongly Disagree, W4= 2 and Undecided, W5= 1, respectively. In effect, the weighted average value is computed as:

Weighted average $=\sum_{i=1}^{5}(Ri\ X\ Wi)/\sum_{i=1}^{5} Ri$ (5)

Where R_i= No. of respondents for Strongly Agree, Agree… or Undecided.

W_i= weighting value assigned to Strongly Agree, … or Undecided.

i = counter 1 – 5.

The overall average of the weighted averages in column 8 equals 4.03, which represents the dependent variable EP.

Table 4: Computation of weighted average of Employee Productivity, EP with respect to 5-Point Likert scale

Responses against 5-point Likert scale							**Weighted Ave. =**
Questions	**SA, W_1=5**	**A, W_2=4**	**D, W_3=3**	**SD, W_4=2**	**UD, W_5=1**	**Total Scores = $\sum_{i=1}^{5}(Ri\ X\ Wi)$**	**(Col. 7/ $\sum_{i=1}^{5} Ri$)**
Q.1	29	53	12	0	3	396	4.08
Q.2	24	62	9	0	2	397	4.09
Q.3	33	56	0	0	8	397	4.09

Q.4	32	57	3	0	5	402	4.14
Q.5	16	59	12	0	10	362	3.73
$\sum_{i=1}^{5} \boldsymbol{Col.8}$ /5							**=20.13/5 = 4.03**

For each of the two industries namely: Oil & Gas Indigenous, OGI and Oil & Gas Multinational, OGM, three multiple regression equations were calibrated giving rise to a total of 6 multiple regression equations for option 1: linear models. Similarly, two multiple regression models (option 2: quadratic models) were calibrated, giving rise to a total of 4 multiple regression non-linear models (See Table 5). The essence of the non-linear quadratic option is to assess any improvement in the value of R^2 with respect to the linear option. The Goodness of fit for the 6 classical multiple regression equations in Table 5, range from 25.9 – 48.4%. According to Nwaogazie (2021), PCA is a variable reduction technique that groups responses with similar measures into principal factors/components. For instance, in Table 5 the model containing 4 constructs classified as principal component D1 has a goodness of fit value of 32% as compared to R^2 of 25.9 % for 3 constructs, both are of OGI. However, for OGM the R^2 are 36.6 & 39.5% representing classical multiple regression for 4 constructs and 3 constructs, respectively. Apparently, the higher R^2 value between 4 and 3 constructs for OGI & OGM represent the preferred multiple regression models for Principal component regression, PCR model.

On the other hand, when the 7 constructs are employed in the classical multiple regression model for OGI & OGM, the resulting goodness of fit are 44.9 & 48.4%, respectively. It is important to note that the 7-construct classical multiple regression model does not belong to PCR model. As one would expect, the resulting goodness of fit for the 7 constructs will usually be greater than those resulting from the principal component regression models. Similar research work in literature to support the factor reduction nature of principal component regression modelling are those of Mbaluka et al. (2022) and Susilawati & Didiharyono (2023). In order to evaluate the sensitivity values of each construct or independent variable, the sensitivity analysis procedure was adopted. Given the R^2 values in Table 5 for OGI & OGM with respect to linear and non-linear PCR models, the observable difference ranges from 0.1 – 1.1 % indicating no improvement on the R^2 values for the non-linear option.

The contribution of each construct to the resulting goodness of fit was ranked with first and last position assigned to the largest & smallest. The 3 or 4 most contributing values to the overall goodness of fit are comparable to the principal components of Table 5. Illustrative examples of sensitivity analysis result were compared with those of the principal components (i.e. 3 & 4 constructs) in Table 5 in section 3.4.

Table 5: Comparative analyses of Goodness of fit for Principal component regression models

s/n	Type of company	No. of independent variables	Multiple Regression Equation	Goodness of fit, R^2
Option 1: Linear Models				
1.	**OGI**	**4**	$EP = 2.156 + 0.623ST + 0.182SPP – 0.120EI – 0.259MC$	0.320
2.	**OGI**	**3**	$EP = 0.353 – 0.161SP + 0.446SC + 0.525SK$	0.259
3.	**OGI**	**7**	$EP = −0.045−0.299·MC−0.303·SP+0.557·SC+0.078·SPP+0.527·ST+0.325·SK+0.011·EI$	0.449
4.	**OGM**	**4**	$EP = 1.275 – 0.008ST + 0.316SPP + 0.242EI + 0.217MC$	0.366
5.	**OGM**	**3**	$EP = -0.366 + 0.541SP – 0.041SC + 0.505SK$	0.395
6.	**OGM**	**7**	$EP = −0.043+0.116·MC+0.401·SP−0.159·SC+0.238·SPP−0.336·ST+0.504·SK+0.214·EI$	0.484
Option 2: Quadratic Models				
1.	**OGI**	**4**	$EP = 3.27174 - 0.27603MC + 0.12982SPP - 0.00493EI^2 + 0.07646ST^2$	0.309
2.	**OGI**	**3**	$EP = 1.60590 + 0.40442SC - 0.16501SP + 0.06451SK^2$	0.270
3.	**OGM**	**4**	$EP = 2.33251 + 0.13381MC - 0.01624ST + 0.04172SPP^2 + 0.03006EI^2$	0.357
4.	**OGM**	**3**	$EP = 0.78268 - 0.03168SC + 0.50241SK + 0.06145SP^2$	0.394

3.4 Sensitivity Analysis by simple & multiple-linear regression modelling

3.4.1 Simple Regression Modelling for OGI & OGM

Simple linear regression modelling involves a direct relationship between a dependent variable and one independent variable, which will yield a goodness of fit that showcases the degree of relationship (Nwaogazie, 2021). Adopting the weighted mean values of the applicable questions for the 7 constructs as input data, the simple linear regression models were developed for the OGI & OGM, respectively (see Figures 2a – g & Figures 3a – g). A close examination of Figures 2a – g and 3a – g, indicates positive relationship between Employee Productivity, EP (dependent variable) and each of the 7 constructs as independent variables (i.e., EI, MC, SC, SK, SP, SPP & ST) in the range of 0.11 – 0.67 for OGI and 0.40 – 0.76 for OGM, respectively. The value of the slope seems to indicate the rate of increase of dependent variable per unit increase of the applicable independent variable. There seems to be an observable trend with respect to incremental slope and goodness of fit (Figure 4). Figure 4 clearly indicates linear positive relationship and very noticeable are the higher values of R^2 for OGM with respect to those of OGI. Of interest is the fact that the goodness of fit, R^2 value in Figure 4 indicate the degree of sensitivity of independent variable with each other.

In Table 6 the rank order of R^2 values places the largest & least values in positions 1 and 7, respectively. The total value of R^2 for OGI and OGM are 81.2 & 171.1%, respectively. However, the maximum value of goodness of fit for any multiple regression model is equal or less than 100%. This is in accordance with conceptual definition of goodness of fit, R^2 (Equation (3)). The R^2 value of 171.1% violates the maximum limit of goodness of fit in a simple linear regression model. Very noticeable are the number of respondents for OGI and OGM (97 and 158, respectively). Given the larger number for OGM, the advantage is seen in the improved R^2 value for OGM as against OGI for 6 out of the 7 constructs (Table 6). This unequal number of respondents between OGI & OGM may be seen as a study limitation. For sake of comparison, it is necessary to have equal number of input data or responses as per questionnaire study for two industries to be compared.

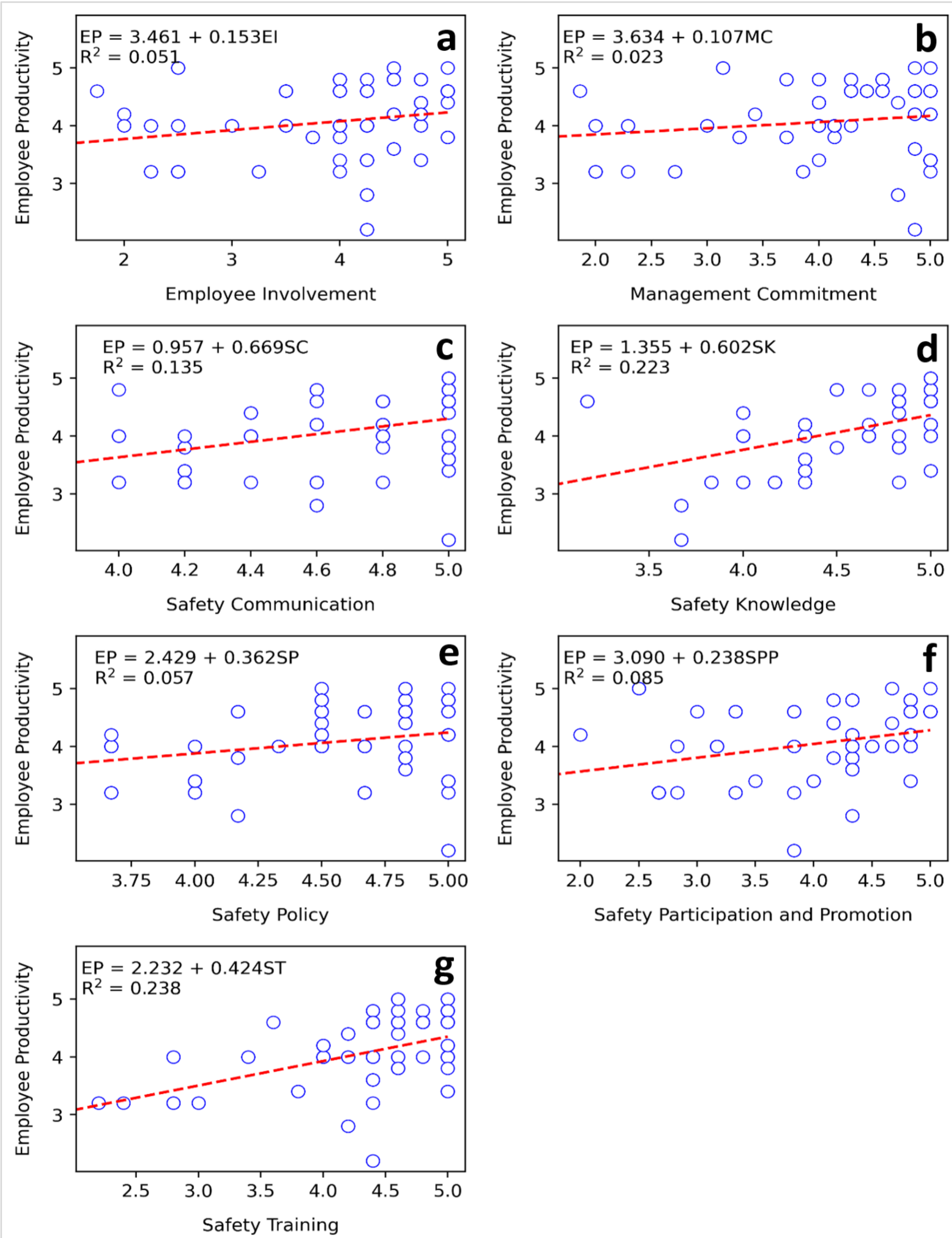

Figure 2: Simple linear regression for 7 constructs for Oil & Gas indigenous

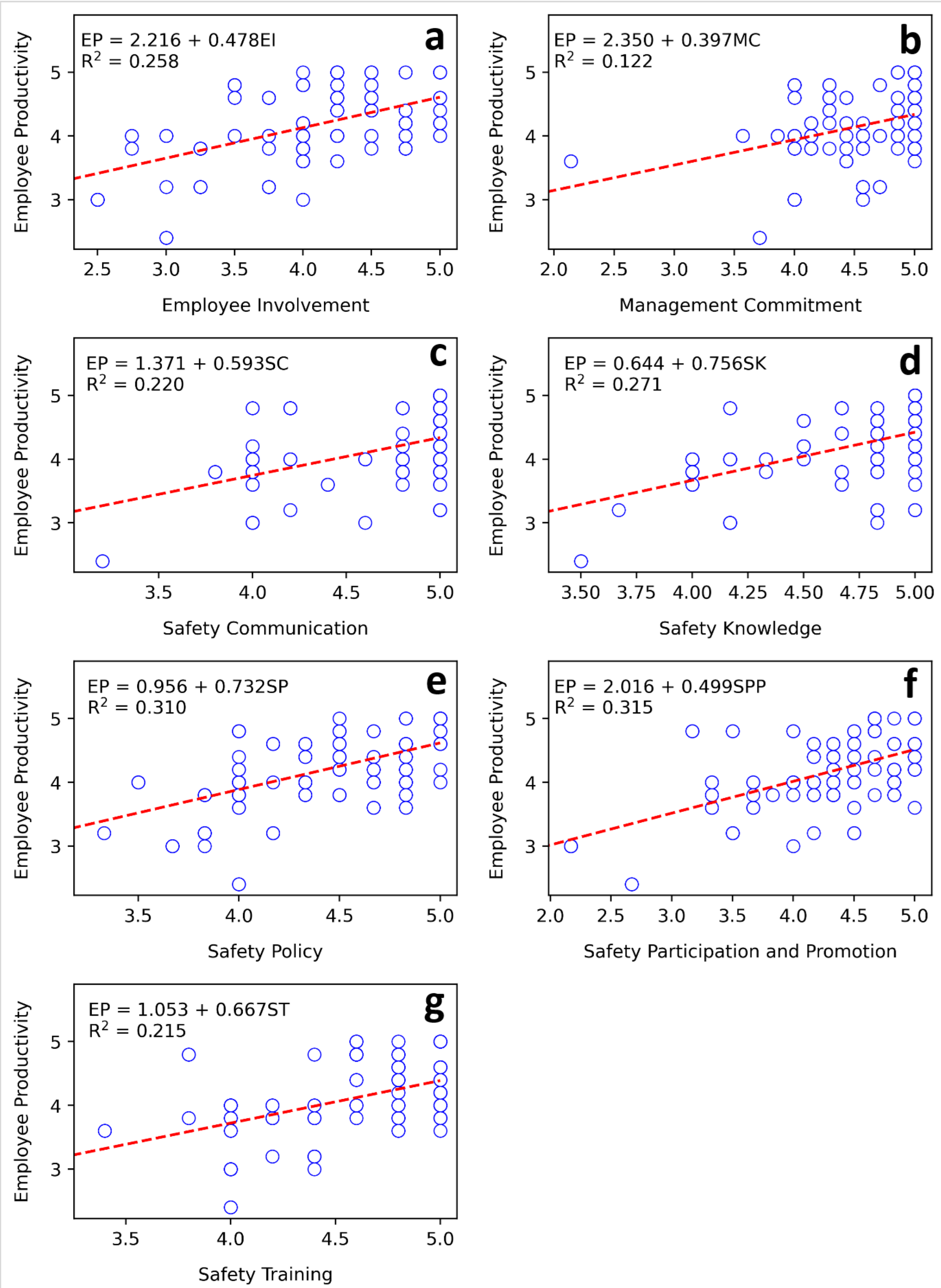

Figure 3: Simple linear regression for 7 constructs for Oil & Gas Multinational

Table 6: Sensitivity analysis part 1 by simple linear regression modelling

s/n	**Oil & Gas Indigenous Company**			**Oil & Gas Multinational Company**		
	Simple linear regression	**Goodness of fit, R^2 (%)**	**Rank Order**	**Simple linear regression**	**Goodness of fit, R^2 (%)**	**Rank Order**
1.	EP=3.461 + 0.153EI	5.1	6th	EP=2.216 + 0.478EI	25.8	4th
2.	EP=3.634 + 0.107MC	2.3	7th	EP=2.350 + 0.397MC	12.2	7th
3.	EP=0.957 + 0.669SC	13.5	3rd	EP=1.371 + 0.593SC	22.0	5th
4.	EP=1.355 + 0.602SK	22.3	2nd	EP=0.644 + 0.756SK	27.1	3rd
5.	EP=2.429 + 0.362SP	5.7	5th	EP=0.956 + 0.732SP	31.0	2nd
6.	EP=3.090 + 0.238SPP	8.5	4th	EP=2.016 + 0.499SPP	31.5	1st
7.	EP=2.232 + 0.424ST	23.8	1st	EP=1.053 + 0.667ST	21.5	6th
8.		$\sum 81.2$			$\sum 171.1$	

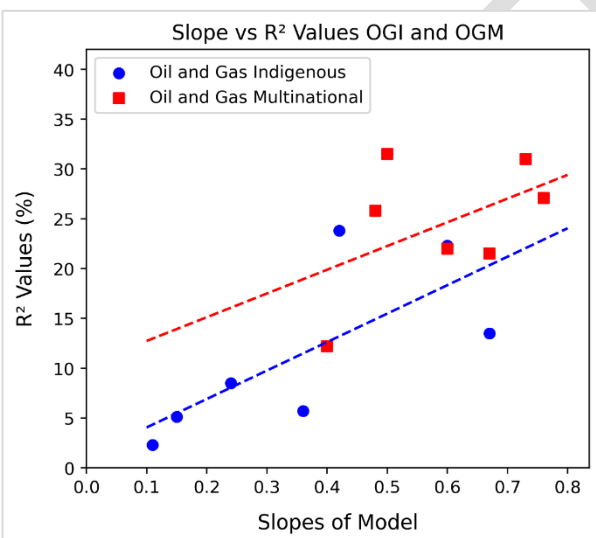

Figure 4: Graph of slope against Goodness of fit of simple linear regression models of OGI & OGM

3.4.2 Classical Multiple Regression Modelling for OGI & OGM

The conventional method of factor (variable) reduction in multiple regression modelling is by sensitivity analysis. The method offers advantage of dropping the less sensitive independent variables and thus, the resulting model becomes very handy for predictive purposes. A situation where few independent variables are involved in a calibrated model gives a sense of time and cost saving, as compared to larger number of independent variables in the model. Table 7 shows a case of sensitivity analysis for both OGI & OGM for each of the independent variables and their order of ranking (i.e. highest to the least value). The total R^2 value with respect to sensitivity analysis for the multiple regression models for OGI & OGM are 44.9 and 48.4%, respectively. These values are in agreement with the multiple regression models containing 7 constructs in Table 5. The difference in R^2 values between 100 & 44.9 or 100 & 48.4 explain the impact of the other organizational safety variables not involved in this study.

A comparative analyses of R^2 values as presented in Tables 5, 6 & 7 are given as follows:

i) Of the 7 simple linear regression models for OGI in Table 6 is 81.2% as against 44.9% in Table 7 credited to multiple regression model;

ii) Similarly, the total R^2 value for OGM in Table 6 is 171.1% as compared to 48.4% of Table 7. The lesson from this result as mentioned earlier in subsection 3.4.1 is that the total R^2 values of simple linear regression will not only be greater than those of multiple linear models, but are most likely to exceed the maximum limit of 100% in multiple regression models;

iii) A second lesson from this study is from the multiple regression resulting from the principal component regression of Table 5 in which the total R^2 value for 3 and 4 constructs are 25.9 & 32.0%; while those of the OGM are 39.5 & 36.6 %, respectively. Similar studies on PCR modelling of water quality index against physicochemical parameters & heavy metals by Odia et al. (2016) and Mbachu & Nwaogazie (2023) yielded R^2 values of 95.5 & 92.9% for 5 constructs; while, 40.0 & 99.0% for 6 constructs, respectively. Therefore, the advantage of the PCR modelling for principal factors is very handy and economical. In effect, PCA as factor reduction in this study indicate that for OGI, the principal components of MC, SPP, ST & EI is superior to the equivalent model of the 3 constructs (i.e. SP, SC & SK). However, for the OGM the 3 constructs multiple regression model yielded R^2 of 39.5% as against 36.6% for the 4 constructs which makes it a superior PCR model;

iv) On the other hand, if the R^2 values of the 3 and 4 constructs of PCR are extracted from Table 7, in preference, we obtain the following values: 16.2 & 28.7% for OGI; while those of the OGM are 15.0 & 33.4%. In other words, the 4 constructs of PCR models evaluated based on sensitivity values of Table 7 and are superior to the equivalent 3 constructs PCR model for OGI & OGM, respectively. This finding is in agreement with that of Ming-ming & Jing-lian (2015) that opined that principal component regression analysis is more accurate than the classical regression analysis;

v) Given that all the R^2 values of PCR models (i.e. 3 & 4 constructs) of Table 5 are superior to the equivalent R^2 values extracted from sensitivity values of Table 7; the reason is on the manner of solving the resulting simultaneous equations arising from those of PCR and multiple linear regression. The PCR modelling adopts the optimization technique, using XLSTAT version 16 in this study; while the classical multiple regression modelling of Table 7 utilizes the Gaussian method of solving the simultaneous linear equations. The optimization method of solving simultaneous equations (linear or non-linear) involves iterative method in which the guess values are assigned to the unknowns and then the process of simultaneous equations begin where the difference between the previous assumed value and the current value is equal or less than a specified error value, and thus solution is achieved. In contrast, the Gaussian solution approach is limited to linear set of simultaneous equations which is usually solved by matrix method and a common option is Gauss elimination technique.

In essence, the classical multiple regression model development should adopt the method of optimization exemplified in XLSTAT Version 16 for improved model coefficient values as well as goodness of fit.

Table 7: Sensitivity analysis part 2 by classical multiple regression modelling

s/n	Oil & Gas Indigenous Company				Oil & Gas Multinational Company			
	multiple regression function	**Goodness of fit, R_i^2 (%)**	**Sensitivity value, x^+_i (%)**	**Rank Order**	**Simple regression function**	**Goodness of fit, R_i^2 (%)**	**Sensitivity value, x_i (%)**	**Rank Order**
1.	$Y = f(x_1)$	$R_1^2=5.1$	$x_1 = R_1^2=5.1$	3rd	$Y = f(x_1)$	$R_1^2=25.8$	$x_1 = 25.8$	1st
2.	$Y=f(x_1,x_2)$	$R_2^2=6.8$	$x_2 = R_2^2-R_1^2=1.7$	6th	$Y=f(x_1,x_2)$	$R_2^2=26.9$	$x_2=1.1$	5th
3.	$Y=f(x_1,x_2,x_3)$	$R_3^2=9.5$	$x_3 = R_3^2-R_2^2=2.7$	4th	$Y=f(x_1,x_2,x_3)$	$R_3^2=27.4$	$x_3=0.5$	7th
4.	$Y=f(x_1,x_2,x_3,x_4)$	$R_4^2=21.2$	$x_4 = R_4^2-R_3^2=11.7$	2nd	$Y=f(x_1,x_2,x_3,x_4)$	$R_4^2=31.1$	$x_4=3.7$	4th
5.	$Y=f(x_1,x_2,x_3,x_4,x_5)$	$R_5^2=23$	$x_5 = R_5^2-R_4^2=1.8$	5th	$Y=f(x_1,x_2,x_3,x_4,x_5)$	$R_5^2=41.9$	$x_5=10.8$	2nd
6.	$Y=f(x_1,x_2,x_3,x_4,x_5,x_6)$	$R_6^2=23.3$	$x_6 = R_6^2-R_5^2=0.3$	7th	$Y=f(x_1,x_2,x_3,x_4,x_5,x_6)$	$R_6^2=47.8$	$x_6=5.9$	3rd

7.	$Y=f(x_1,x_2,x_3,x_4,x_5,x_6,x_7)$	$R_7^2=44.9$	$x_7 = R_7^2-R_6^2=21.6$	1st	$Y=f(x_1,x_2,x_3,x_4,x_5,x_6,x_7)$	$R_7^2=48.4$	$x_7=0.6$	6th
8.			$\sum 44.9$				$\sum 48.4$	

x^+_i : x_1=EI, x_2=MC, x_3= SC, x_4=SK, x_5=SP, x_6=SPP & x_7=ST

4. CONCLUSION

Based on the results of this study carried out, the following conclusions are drawn:

i) The simple linear regression models of the Employee Productivity as dependent variable against each of the 7 independent variables (Employee Involvement, Management Commitment, Safety Compliance, Safety Knowledge, Safety Participation, Safety Promotional Policies & Safety Training) yielded positive slope values for both OGI & OGM. The range of slope values are 0.11 – 0.67 for OGI and 0.40 – 0.76 for OGM which are positively related with their corresponding goodness of fit.

ii) An observable trend with respect to incremental slope and goodness of fit was established in the plot of slope as independent variable against goodness of fit as dependent variable for both OGI & OGM.

iii) In contrast to the multiple regression model, for which the total goodness of fit value for the 7 constructs for OGI = 44.9 & 48.4% for OGM; the simple linear models for the same 7 construct yielded total goodness of fit of 81.2 & 171.1% for OGI & OGM, respectively. The result showed that the predicted R^2 value for each of the 7 constructs with respect to simple linear regression are compliant to Equation (3). However, the total value for OGM of 171.1% is greater than 100% which is not compliant to Equation (3). The total goodness of fit values for the simple linear regression are not only greater than those of multiple linear models, but are most likely to exceed the maximum limit of 100% in multiple regression models.

iv) The procedure for sensitivity analysis of the 7 independent variables were demonstrated using classical multiple regression model and the resulting sensitivity values were ranked in the order of importance. The 1st – 4th position (EI, SP, SPP, SK) for which the total goodness of fit, R^2 = 41.1%. The remaining three less-sensitive variables i.e. MC, ST, SC yielded R^2 value of 3.8% and these three can be dropped. Similarly, for OGM, the four most ranked independent variables are ST, SK, EI & SC with total R^2 value of 46.2% with three less-sensitive variables (SP, MC, SPP) yielded R^2 value of 2.2% can equally be dropped.

v) The principal component regression models established in this study is a variable reduction technique. Given that all the R^2 values of PCR models (i.e. 3 & 4 constructs) are superior to the equivalent R^2 values extracted from sensitivity values of classical multiple regression models; the reason is on the manner of solving the resulting simultaneous equations arising from those of PCR and multiple linear regression. The PCR modelling adopts the optimization technique, using XLSTAT version 16 in this study; while the classical multiple regression modelling utilizes the Gaussian method of solving the simultaneous linear equations.

5. RECOMMENDATION

Based on the comparative analysis of sensitivity analysis and principal component regression modelling, the goodness of fit resulting from PCR models has an edge over those of multiple regression models. It is therefore recommended that any simultaneous equation resulting from multiple regression formulation should be solved using optimization software in XLSTAT Version 16 (preferably current version).

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

References

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health, 96*(2), 315-324.

Cochran, W. G. (1963). Sampling Techniques, 2nd Ed., New York: John Wiley and Sons, Inc.

Clarke, S. (2006). The relationship between safety climate and safety performance: A meta-analytic review. *Journal of Occupational Health Psychology, 11*(4), 315-327.

Davis, J. C. (1973): Statistics and Data Analysis in Geology, 1st edition, John Wiley and Sons, New York.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology, 5*(3), 347-358.

Griffin, M. A., & Neal, A. (2006). A Study of the Lagged Relationships Among Safety Climate, Safety Motivation, Safety Behavior, and Accidents at the Individual and Group

Levels. *Journal of Applied Psychology*, *91*(4), 946-953. https://doi.org/10.1037/0021-9010.91.4.946

Hale, A. R., Guldenmund, F. W., van Loenhout, P. L., & Oh, J. (2010). Evaluating safety management and culture interventions to improve safety: Effective intervention strategies. *Safety Science*, *48*(8), 1026-1035.

Herren, Abby Breann (2023). Goodness of Fit for Multivariable Linear Regression. Honors Capstone Projects and Theses. 805. https://louis.uah.edu/honors-capstones/805

Hofmann, D. A., & Morgeson, F. P. (1999). Safety-related behavior as a social exchange: The role of perceived organizational support and leader-member exchange. *Journal of Applied Psychology*, *84*(2), 286-296.

Mbachu, Frank C. and Nwaogazie, Ify. L. (2023). Modelling and Prediction of Water Quality Index of Selected Pond Water in Aboh Mbaise Local Government Area, Imo State, Nigeria. J. Eng. Res. Rep., vol. 25, no. 11, pp. 13-21; Article no. JERR.108942. DOI: 10.9734/JERR/2023/v25i111016.

Mbaluka, Morris Kateeti, Muriithi, Dennis K. & Njoroge, Gladys G. (2022). Application of Principal Component Analysis and Hierarchical Regression Model on Kenya Macroeconomic Indicators. EJ-MATH, 3(1), 26 – 38, DOI: http://dx.doi.org/10.24018/ejmath.2022.3.1.74.

McGill University Canada (2023). Goodness of Fit in Linear Regression, available at:

http://www.medicine.mcgill.ca › joseph › EPIB-621

Ming-ming, Chen & Jing-lian, Ma (2015). Application of Principal Component Regression Analysis in Economic Analysis. International Conference on MSETASSE, published by Atlantis Press, pp. 1205 – 1208, DOI: 10.2991/msetasse-15.2015.255

Nwaogazie, I.L. (2021): Probability and Statistics for Science and Engineering Practice, 4th edition, De-Adroit Innovation, Enugu, Nigeria pp 363. ISBN: 978-8137-47-4, available at: http://Selar.co/9o5y.

Odia Mamuyovwi, Nwaogazie Ify L., Nwachukwu Eunice O. and Avwiri Gregory O. (2016). ”Modeling Leachate Pollution Index and Potential for Selected Municipal Solid Waste Dump Sites: A Case Study. British Journal of Applied Science and Technology. Vol. 18, Issue 6, pp. 1-16. Article no. BJAST.31198. ISSN: 2231-0843, NLM ID: 101664541. Available at: www.sciencedomain.org.

Sala, Lazarus Tsado, Nwaogazie, Ify L. and Ugbebor, John N. (2024). A Comparative Analysis of Safety Programs and Employee Productivity in Oil and Gas Versus Construction Industries in the Niger Delta Region. Asian Journal of Advanced Research and Reports Volume 18, Issue 12, Page 10-24. DOI: https://doi.org/10.9734/ajarr/2024/v18i12802.

Susilawati, Sumarni & Didiharyono (2023). Application of Principal Component Regression in Analyzing Factors Affecting Human Development Index, Vol. 6, No. 2, April 2023, Hal. 199 208 e-ISSN: 2581-2017, pp. 199 208, DOI: https://doi.org/10.30812/varian.v6i2.2366.

Vredenburgh, A. G. (2002). Organizational safety: Which management practices are most effective in reducing employee injury rates? *Journal of Safety Research*, *33*(2), 259-276.