Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.25
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Full Length Research Article

ELIMINATING THE EFFECT OF CONCOMITANT VARIABLES ON SALES DATA USING THE ANALYSIS OF COVARIANCE TECHNIQUE

*[1]Yisa Yakubu, [2]Daniel Bitrus Dajel, [1]Usman Abubakar , [3]Usman Yahayah Baba

[1]Department of Statistics, School of Physical Sciences, Federal University of Technology, Minna, Nigeria
[2]Drugs and Medical Commodities Management Agency, Jos, Nigeria
[3]Department of Mathematics and Statistics, Federal Polytechnic, Idah, Nigeria

*Corresponding Author Email Address: yisa.yakubu@futminna.edu.ng

ABSTRACT
In marketing research, the understanding of relationships between variables by statistically controlling for extraneous sources of variation is refined by eliminating covariates. This work examines and eliminates the covariate effect on the sales of grand cereals oil. The data set for the work was on yearly sales (in millions of Naira) of grand cereals oil, Managers' Age, Gender, and Operating Cost (the covariate) for a period of thirty (30) years (1991 - 2020). A general linear model (GLM) was first fitted to the sales data without the covariate. Then, another GLM with the covariate included (analysis of covariance (ANCOVA)) was fitted. Then, the proportion of total variability in the data accounted for by the covariate alone and by both the dummy variable and the covariate was investigated. The GLM result shows a highly significant effect of both the gender and age, with $P< 0.0001$, R^2 of 49.8%, and mean squared error (MSE) of 70.782. The ANCOVA model showed a highly significant age effect with $P< 0.0001$, R^2 of 80.5%, and a MSE of 27.028, indicating a sharp increase in the power of the test. The study has established the potential of the ANCOVA technique in controlling and eliminating the effect of the concomitant variable.

Keywords: Sales Data, General Linear Model, Analysis of Covariance, Regression Mean Squared Error.

INTRODUCTION

For any business to be successful, there is a need for strong strategic marketing research for sales improvement by the firm. In marketing research, the understanding of relationships between variables by statistically controlling for extraneous factors is refined by analyzing covariates. The three important principles often used by experimenters in statistics to increase precision are randomization, replication, and blocking. However, Fisher (1935) observed a second means by which precision may be much increased in designed experiments, which involves the elimination of causes of variation (concomitant variables or covariates) that the experimenter cannot control.

With these measurements on covariates, the treatment effects can be corrected or adjusted. The unexplained variability (that is, experimental error) in most experiments with a response y is due to the effect of the covariate, which is linearly related to, and can only be observed along with, y. Analysis of covariance (ANCOVA) will eliminate the effect of the covariate (X) on the response variable (y), reduce the experimental error variance, thereby revealing a precise independent variable effect (Montgomery, 2012).

To effectively eliminate covariates from sales data and isolate the impact of specific variables, ANCOVA or randomization and blocking techniques can be employed during the study design or analysis phase. ANCOVA allows adjusting for the influence of measured covariates, while randomization and blocking aim to distribute the effects of both known and unknown covariates evenly across different groups. ANCOVA combines features of analysis of variance (ANOVA) and regression.

A first restriction of covariates in experimental designs listed by Field (2011) is an independent relationship between the covariates and conditions. If this assumption is violated, there is no reliable way to distinguish the factor effect from the covariate effect on the response (Kutner *et al*., 2005; Miller & Chapman, 2001). Regression slopes homogeneity is another crucial assumption, where the covariate effect on the dependent variable across all treatment conditions is assumed to be constant (Field, 2011; Howell, 2010; Kutner *et al*., 2005). Violation of this assumption leads to inaccurate treatment effect size estimates, since the between-groups differences will depend on the covariate value (Engqvist, 2005).

This study aims to examine and eliminate the covariate effect on the sales of grand cereals oil**.** The work makes use of the data on yearly sales (millions of Naira) of the oil from 1991 to 2020 (30 years). The factors examined were the sales managers' Age and Gender, as the independent variables, while the covariate is the operating cost (millions of naira), and is continuous.

Each of the two factors was first tested alone for effect significance on oil sales without the covariate by fitting a general linear model (GLM). Then, another GLM was fitted to the sales data with the covariate included so as to test for its effect on sales. Then, the proportion of total variability in the data accounted for by the covariate alone and by both the dummy variable and the covariate is investigated.

MATERIALS AND METHODS

Data Collection

The factors examined in this work were the sales managers' Age and Gender, as the independent variables, while the covariate is the operating cost (in millions of naira), and is continuous. The Age variable consists of three categories, which include 20-28 years (young), 29-37 years (middle-aged), and 38-46 years (elderly). The Gender variable consists of two categories, which include 1= male and 2 = female.

Model Fitting

For a one-factor and one covariate designed experiment, the ANCOVA model is

$$y_{ij} = \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij} \quad (1)$$

Where y_{ij} is the j^{th} observation for the i^{th} treatment, μ is the overall mean, τ_i is the i^{th} treatment effect, ε_{ij} is the random error, β is the covariate effect on y_{ij}, and x_{ij} is the covariate score corresponding to the dependent variable y_{ij}. Thus, the ANCOVA general linear model (1) combines the ANOVA and regression general linear model features.

Now, if we replace βx_{ij} in equation (1), with $\beta(x_{ij} - \bar{x}_{..})$, we have the one-factor experimental design model for the general ANCOVA with one covariate as

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_{..}) + \varepsilon_{ij} \quad (2)$$

where $\bar{x}_{..}$ is the grand covariate mean.

Now to estimate the parameters in (2), we have:

$$L = \sum_{i=1}^{t}\sum_{j=1}^{n} e_{ij}^2 = \sum_{i=1}^{t}\sum_{j=1}^{n}[y_{ij} - \mu - \tau_i - \beta(x_{ij} - \bar{x}_{..})]^2 \quad (3)$$

Taking partial derivative of (3) with respect to the parameters μ, τ_i, and β, and equating to zero, we have

$$\boldsymbol{\mu}: y_{..} = tn\hat{\mu} + n\sum_{i=1}^{t}\hat{\tau}_i + \beta\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = tn\hat{\mu}$$

Thus $\hat{\mu} = \bar{y}_{..}$ (4)

$$\boldsymbol{\tau_i}: \sum_{j=1}^{n} y_{ij} = \sum_{j=1}^{n}\hat{\mu} + \sum_{j=1}^{n}\hat{\tau}_i + \beta\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) \quad (5)$$

That is,

$$y_{i.} = n(\hat{\mu} + \hat{\tau}_i) + \beta\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}), i = 1,2,\dots,t \quad (6)$$

Thus

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \frac{\beta\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})}{n} \quad (7)$$

That is,

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} - \beta(\bar{x}_{i.} - \bar{x}_{..}) \quad (8)$$

$$\boldsymbol{\beta}: \sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}(x_{ij} - \bar{x}_{..}) = \mu\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) + \sum_{i=1}^{t}\sum_{j=1}^{n}\tau_i(x_{ij} - \bar{x}_{..}) + \beta\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})^2 \quad (9)$$

That is,

$$\sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}(x_{ij} - \bar{x}_{..}) = \sum_{i=1}^{t}\tau_i\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) + \beta\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})^2 \quad (10)$$

Now substituting for $\hat{\tau}_i$ from equation (8) in equation (10), we have

$$\sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}(x_{ij} - \bar{x}_{..}) = \sum_{i=1}^{t}[\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})]\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) + \hat{\beta}\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})^2 \quad (11)$$

That is,

$$\sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}(x_{ij} - \bar{x}_{..}) = \sum_{i=1}^{t}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) - \hat{\beta}\sum_{i=1}^{t}(\bar{x}_{i.} - \bar{x}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) + \hat{\beta}\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})^2 \quad (12)$$

Thus,

$$\hat{\beta} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}(x_{ij} - \bar{x}_{..}) - \sum_{i=1}^{t}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})}{\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})^2 - \sum_{i=1}^{t}(\bar{x}_{i.} - \bar{x}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})} \quad (13)$$

From the numerator of equation (13), we have

$$\sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij}(x_{ij} - \bar{x}_{..}) = \sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}y_{ij} - \frac{(\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij})(\sum_{i=1}^{t}\sum_{j=1}^{n} y_{ij})}{tn} = \sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{tn}$$

And

$$\sum_{i=1}^{t}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = n\sum_{i=1}^{t}(\bar{y}_{i.} - \bar{y}_{..})(\bar{x}_{i.} - \bar{x}_{..}) = \frac{1}{n}\sum_{i=1}^{t}(x_{i.})(y_{i.}) - \frac{(x_{..})(y_{..})}{tn}$$

Thus, this numerator equals

$$\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}y_{ij} - \frac{(x_{..})(y_{..})}{tn} - \left[\frac{1}{n}\sum_{i=1}^{t}(x_{i.})(y_{i.}) - \frac{(x_{..})(y_{..})}{tn}\right] = \sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}y_{ij} - \frac{1}{n}\sum_{i=1}^{t}(x_{i.})(y_{i.}) = n\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}y_{ij} - \sum_{i=1}^{t}(x_{i.})(y_{i.})$$

Similarly, the denominator of (13) can be expressed as

$$\sum_{i=1}^{t}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}^2 - \frac{x_{..}^2}{tn}$$

and

$$\sum_{i=1}^{t}(\bar{x}_{i.} - \bar{x}_{..})\sum_{j=1}^{n}(x_{ij} - \bar{x}_{..}) = \frac{1}{n}\sum_{i=1}^{t} x_{i.}^2 - \frac{x_{..}^2}{nt}$$

Thus, the denominator equals,

$$\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}^2 - \frac{x_{..}^2}{tn} - \left(\frac{1}{n}\sum_{i=1}^{t} x_{i.}^2 - \frac{x_{..}^2}{nt}\right) = \sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{n}\sum_{i=1}^{t} x_{i.}^2$$

$$= n\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{t} x_{i.}^2 \quad (15)$$

Now putting equations (14) and (15) back into equation (13), we have

$$\hat{\beta} = \frac{n\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}y_{ij} - \sum_{i=1}^{t}(x_{i.})(y_{i.})}{n\sum_{i=1}^{t}\sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{t} x_{i.}^2} \quad (16)$$

Equation (16) gives the estimate of the effect of the covariate on the response variable in a one-way designed experiment containing one factor and one covariate.

Now, for an experimental design with *m* experimental factors $x_1, x_2, \dots, x_m$ and c covariates, we have a main-effects model:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \gamma_1 z_1 + \cdots + \gamma_c z_c + \varepsilon = \beta_0 + \sum_{i=1}^{m}\beta_i x_i + \sum_{i=1}^{c}\gamma_i z_i + \varepsilon \quad (17)$$

In situations where some of the experimental factors interact, model (17) becomes inadequate, and a more appropriate model is the expanded one that captures the effect of such interaction. This may be

$$Y = \beta_0 + \sum_{i=1}^{m}\beta_i x_i + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\beta_{ij}x_i x_j + \sum_{i=1}^{c}\gamma_i z_i + \varepsilon \quad (18)$$

where all the terms are as defined above, and β_{ij} is the effect of the interaction between the i^{th} and the j^{th} levels of x.

Situations of interaction between experimental factors and covariates may arise. In such situations, model (18) has to be expanded to give:

$$Y = \beta_0 + \sum_{i=1}^{m}\beta_i x_i + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\beta_{ij}x_i x_j + \sum_{i=1}^{c}\gamma_i z_i + \sum_{i=1}^{m}\sum_{j=1}^{c}\beta_{ij^c}x_i z_{j^c} + \varepsilon \quad (19)$$

where β_{ij^c} is the effect of the interaction between the ith level of the experimental factor and the j^c level of the covariate.

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.25
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Situations also may arise where interaction between the covariates exists. Then we have the model for the designs in such situations as

$$Y = \beta_0 + \sum_{i=1}^{m}\beta_i x_i + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\beta_{ij}x_i x_j + \sum_{i=1}^{c}\gamma_i z_i + \sum_{i=1}^{c-1}\sum_{j=i+1}^{c}\gamma_{ij}z_i z_j + \sum_{i=1}^{m}\sum_{j=1}^{c}\beta_{ij}x_i z_j + \varepsilon \quad (20)$$

where γ_{ij} is the interaction effect between the i^{th} and the j^{th} levels of the covariates.

Ignoring covariates out of the model for experimental designs that it is supposed to be present may result in increased error variance, reduced statistical power, biased estimates, and invalid comparisons.

In matrix notation, these models (2, 17, 18, 19, and 20) can be expressed as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon} \quad (21)$$

where, for the main-effects model (17),

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{m1} \\ 1 & x_{12} & \dots & x_{m2} \\ . & . & & . \\ . & . & \dots & . \\ . & . & & . \\ 1 & x_{1n} & \dots & x_{mn} \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & z_{11} & \dots & z_{c1} \\ 1 & z_{12} & \dots & z_{c2} \\ . & . & & . \\ . & . & \dots & . \\ . & . & & . \\ 1 & z_{1n} & \dots & z_{cn} \end{bmatrix}$$

$$\boldsymbol{\beta} = [\beta_0\ \beta_1 \dots \beta_m]'$$

and

$\boldsymbol{\gamma} = [\gamma_0\ \gamma_1 \dots \gamma_c]'$.

For the model (18) where the experimental factors interact, we have

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{m1} & \dots & x_{11}x_{21} & \dots & x_{m-1,1}x_{m1} \\ 1 & x_{12} & x_{22} & \dots & x_{m2} & \dots & x_{12}x_{22} & \dots & x_{m-1,2}x_{m2} \\ . & . & . & \dots & . & \dots & . & \dots & . \\ . & . & . & \dots & . & \dots & . & \dots & . \\ . & . & . & \dots & . & \dots & . & .. & . \\ 1 & x_{1n} & x_{2n} & \dots & x_{mn} & \dots & x_{1n}x_{2n} & \dots & x_{m-1,n}x_{mn} \end{bmatrix}$$

And when the covariates also interact, we have

$$\boldsymbol{Z} = \begin{bmatrix} 1 & z_{11} & z_{21} & \dots & z_{c1} & \dots & z_{11}z_{21} & \dots & z_{c-1,1}z_{c1} \\ 1 & z_{12} & z_{22} & \dots & z_{c2} & \dots & z_{12}z_{22} & \dots & z_{c-1,2}z_{c2} \\ . & . & . & \dots & . & \dots & . & \dots & . \\ . & . & . & \dots & . & \dots & . & \dots & . \\ . & . & . & \dots & . & \dots & . & .. & . \\ 1 & z_{1n} & z_{2n} & \dots & z_{cn} & \dots & z_{1n}z_{2n} & \dots & z_{c-1,n}z_{cn} \end{bmatrix}$$

while

$$\boldsymbol{\beta} = [\beta_0\ \ \beta_1 \dots \beta_m\ \ \beta_{12} \dots \beta_{m-1,m}]'$$

and

$$\boldsymbol{\gamma} = [\gamma_0\ \ \gamma_1 \dots \gamma_c\ \ \gamma_{12} \dots \gamma_{c-1,c}]$$

ANOVA and Regression approaches are used in this work to analyze the collected data. With the ANOVA approach, a GLM is first fitted to the data using only the independent variables (sales managers' age and gender) without the covariate (operating cost), so as to assess the impact of each of the factors on the oil sales. Next, we perform the GLM univariate analysis on the two factors (Age and Gender) and the covariate (operating cost) combined. The GLM results for the two analyses are then compared in terms of power.

With the regression approach, first, each value of the variables (Age and Gender) is represented in the model with an indicator (dummy) variable consisting of only codes, leaving one of the levels of every variable out of the regression model (as a reference category) to avoid perfect multicollinearity, which will prevent a solution. For the Age variable, the elderly-age category is the reference category, while for the Gender variable, the female category is the reference category. For each of these variables, the differential effect estimates for other levels (or categories) are then compared with the reference category.

Now, we first regress the response (sales) variable on the dummy variables alone (that designated the treatments) and obtain $R^2_{y\,D_{Age1},D_{Age2},D_{Gender1}}$, which is the proportion of the total variability in the response (sales) accounted for by these dummies.

Next, we regress the response (sales) variable on the covariate (operating cost) alone and obtain $R^2_{y,X}$, which is the proportion of the total variability in the response accounted for by the fitted model. Thirdly, we regress the response (sales) variable on the dummy variables scores (that designate the treatments) and the covariate combined, and also obtain $R^2_{y\,D_{Age1},D_{Age2},D_{Gender1},X}$, which is the proportion of the total variability in the response accounted for by both these variables.

In each case, the coefficient of multiple determination (R^2) is given as

$$R^2 = \frac{Model\ SS}{Total\ SS} = \frac{Total\ SS - Residual\ SS}{Total\ SS} \quad (22)$$

where,

$Total\ SS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$, $Residual\ SS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, and $Model\ SS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$

Then, we compute the difference $R^2_{y\,D_{Age1},D_{Age2},D_{Gender1},X} - R^2_{y,X}$, which gives the unique contribution of the dummy variables to the first computed coefficient and also reflect the proportion of the total variation that is uniquely accounted for by the independent variables.

RESULTS

The collected data were analyzed, and the ANOVA and ANCOVA results are presented in Tables 1 and 2, respectively.

(1) The ANOVA and ANCOVA results from the GLM technique.

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.25
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Table 1: Analysis of Variance (ANOVA) Model

Dependent Variable: Sales

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12238.294[a]	5	2447.659	34.58	0.000
Intercept	158360.672	1	158360.672	2237.308	0.000
Age	6433.878	2	3216.939	45.449	0.000
Gender	3440.939	1	3440.939	48.613	0.000
age * gender	2363.478	2	1181.739	16.696	0.000
Error	12316.033	174	70.782		
Total	182915.000	180			
Corrected Total	24554.328	179			

a. R Squared = .498 (Adjusted R Squared = .484)

Table 2: Analysis of Covariance (ANCOVA) Model

Dependent Variable: sales

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	19774.762[a]	6	3295.794	119.294	0.000
Intercept	537.307	1	537.307	19.448	0.000
op_cost	7536.468	1	7536.468	272.788	0.000
Age	2010.672	2	1005.336	36.389	0.000
Gender	2375.67	1	2375.67	85.989	0.000
age * gender	1125.013	2	562.506	20.36	0.000
Error	4779.566	173	27.628		
Total	182915	180			
Corrected Total	24554.328	179			

a. R Squared = .805 (Adjusted R Squared = .799)

(2) The ANCOVA results through the linear regression procedure

Table 3: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.706[a]	.498	.484	8.41319

a. Predictors: (Constant), gender dummy code by age grp2 dummy code interaction, gender dummy code by age grp1 dummy code interaction, age1, age2, gender1.

Table 4: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12238.294	5	2447.659	34.580	.000[b]
	Residual	12316.033	174	70.782		
	Total	24554.328	179			

a. Dependent Variable: sales

b. Predictors: (Constant), gender dummy code by age grp2 dummy code interaction, gender dummy code by age grp1 dummy code interaction, age1, age2, gender1.

Table 5: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	35.7	1.536		23.242	0.000
	age1	-13.433	2.172	-0.542	-6.184	0.000
	age2	8.433	2.172	0.34	3.882	0.000
	gender1	-5.433	2.172	-0.233	-2.501	0.013
	Gender1 x Age grp1	3.433	3.072	0.11	1.118	0.265
	Gender1 x Age grp2	-13.367	3.072	-0.427	-4.351	0.000

a. Dependent Variable: Sales

Table 6: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.759[a]	0.576	0.574	7.64549

a. Predictors: (Constant), operating cost

Table 7: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14149.605	1	14149.605	242.066	.000[b]
	Residual	10404.723	178	58.454		
	Total	24554.328	179			

a. Dependent Variable: sales

b. Predictors: (Constant), operating cost

Table 8: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.159	1.554		4.606	0.000
	operating cost	0.189	0.012	0.759	15.558	0.000

a. Dependent Variable: Sales

Table 9: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.897[a]	0.805	0.799	5.25619

a. Predictors: (Constant), gender dummy code by age grp2 dummy code interaction, gender dummy code by age grp1 dummy code interaction, operating cost, gender1, age1, age2

Table 10: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19774.76	6	3295.794	119.294	.000[b]
	Residual	4779.566	173	27.628		
	Total	24554.33	179			

a. Dependent Variable: Sales

b. Predictors: (Constant), gender dummy code by age grp2 dummy code interaction, gender dummy code by age grp1 dummy code interaction, operating cost, gender1, age1, age2.

Table 11: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	15.851	1.538		10.306	0.000
	operating cost	0.196	0.012	0.783	16.516	0.000
	age1	-10.995	1.365	-0.444	-8.054	0.000
	age2	-6.279	1.623	-0.253	-3.868	0.000
	gender1	-9.253	1.377	-0.396	-6.721	0.000

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.25
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

gender1 age grp1	X	8.818	1.947	0.281	4.529	0.000
gender1 age grp2 d	X	-2.989	2.02	-0.095	-1.48	0.141

a. Dependent Variable: Sales

DISCUSSION

Table1 is a reduced model where only the variations due to the two factors (Age and Gender) are shown. The Table shows that each of the two factors (Age and Gender) and their interaction has a significant effect on the sales of the grand cereals oil. The ANOVA model accounted for about 49.8% of the total variability in the sales data, as indicated by $\boldsymbol{R^2 = 0.498}$.The ANCOVA model in Table 2 shows that each of the two factors and their interaction, as well as the covariate, has a significant effect on the sales of the oil. This fitted ANCOVA model outperforms the ANOVA model as it caused a sharp increase in the sum of squares due to regression, a drastic reduction in the error sum of squares and thereby in the error variance (MSE) and it accounted for about 80.5% of the total variability in the sales data with $\boldsymbol{R^2 = 0.805}$. This result also indicates a high and positive association between the covariate and the response (sales) variable.

(i) Tables 3 – 5 give the regression summary results of the sales scores on the dummy variables scores alone (that designate the treatments). Table 3 gives the summary of the fitted model in Table 5, showing the presence of a significant association (r = 0.706) between the oil sales scores and the dummy variable scores, with the fitted model accounting for about 49.8% of the total variability in the sales data. This fairly low proportion indicates the presence of a covariate that is also highly correlated with the sales variable.

Table 5 is the coefficient table. From this table, the fitted linear regression model is

$$\hat{y} = 35.700 - 13.433U_1 + 8.433U_2 - 5.433V_1 + 3.433U_1V_1 - 13.367U_2V_1 \quad (23)$$

Where: U_1= young age category, U_2= middle age category, and V_1= male category, $\hat{y}$ is the estimated sales of the Grand Cereals Oil, 35.700 is the expected sales (in millions of naira) of the oil by the sales agents in the elderly-age category (that is, 38-46 years) and those who are females by gender. This model shows that the sales agents within 20 to 28 years age group made sales of about 13½ million naira less than that of the sales agents who are within 38 to 46 years age group, while those within the middle-age category (that is, 29-37 years) made sales of about 8½ million naira more than that of their counterparts that are within 38 to 46 years of age. From Table 5, we observed that each of these two coefficients is statistically significant (*P-value=0.000)*, indicating a significant difference in the sales of the grand cereals oil between the agents in each of these two age groups and those in the 38-46 years age category.

The model further shows that the male sales agents made about 5½ million naira less than their female counterparts. This coefficient is also statistically significant at 5% level with a *p-value* of 0.013, which indicates that there is a significant difference in sales of the oil between the male and the female sales agents. The interaction effect (3.433) between the young sales agents (that is, sales agents between 20-28 years of age) and the male sales agents(U_1V_1) indicates that the effectiveness of AGE variable on sales of grand cereals oil increased by about 3½ million naira for males sales agents compared to the female sales agents. From the coefficient table, we see that this coefficient is not statistically significant as its p-value is 0.265.

The interaction effect (-13.367) between the middle-aged sales agents and the male sales agents(U_2V_1) indicates that the effectiveness of the GENDER variable on sales of grand cereals oil decreased by about 13.4 million naira for sales agents within the 29-37 years age group compared to those within the 38-46 years age group. From the coefficient table, we see that this coefficient is statistically significant as its p-value is 0.000.

(ii) Tables 6 – 8 give the ANCOVA results of the regression of the sales scores on the covariate scores alone.

Table 6 gives the summary of the fitted model in Table 8 and shows the presence of a significant association between the Grand Cereals Oil sales scores and the operating cost scores, as given by the high correlation coefficient of about 0.759. The Table shows that the fitted model in equation (23) accounted for about 57.6% of the total variability in the sales data. This is further confirmed by the ANOVA in Table 7, where the sum of squares of the fitted regression model is shown to be significantly higher than the error sum of squares, and with a p-value of 0.000.

Table 8 is the coefficient table. From this table, the fitted linear regression model is

$$\hat{y} = 7.159 + 0.189X \quad (24)$$

where $\hat{y}$ is the estimated sales of the Grand Cereal Oil and X is the corresponding operating cost (the covariate). The model shows that the company makes sales of 7.159 million naira per year in the absence of operating cost, and that for every unit change in the operating cost incurred by the company, there is a change (an increase) of 0.189 million naira in the sales of the oil. The table shows that the operating cost has a significant effect on the sales as indicated by the significant probability value (p-value) of 0.000.

(iii) Tables 9 – 11 give the ANCOVA results of the regression of the sales scores on the dummy variable scores (that designate the treatments) and the covariate scores.

Table 9 gives the summary of the fitted model in Table 11 and shows a significant association between the Grand Cereals Oil sales scores and the dummy variables and the covariate scores, as given by the high correlation coefficient of 0.897. The Table shows that the fitted model in Table 11 accounted for about 80.5% of the total variability in the sales data. This result is due to the removal of the effect of the covariate (operating cost) in this analysis. This is further confirmed by the ANOVA in Table 10, where the sum of squares of the fitted regression model is shown to be higher than that of the error. The table further confirms that the model is highly significant at both 1% and 5% levels of significance, with a probability value (*p-value*) of 0.000.

Table 11 is the coefficient table. From this table, the fitted ANCOVA (linear regression) model is

$$\hat{y} = 15.851 + 0.196X - 10.995U_1 - 6.279U_2 - 9.253V_1 + 8.818U_1V_1 - 2.989U_2V_1 \quad (25)$$

where $\hat{y}$ is the estimated sales of the Grand Cereals Oil, 15.851 is the expected sales (in millions of naira) of the oil by the sales agents in the elderly-age category (that is, 38-46 years old) and those who are females by gender when there is no effect of the covariate.

This model shows that a unit change in the covariate (operating cost) scores will result in an increase of 0.196 million naira in the sales of the oil. The model shows that the sales agents within 20 to

28 years age group made sales of about 11 million naira less than that of the sales agents who are within 38 to 46 years age group, while those within the middle-age category (that is, 29-37 years old) made sales of about 6.3 million naira less than that of their counterparts that are within 38 to 46 years of age. From Table 11, we observed that each of these two coefficients is statistically significant (*P-value=0.000)*, indicating a significant difference in the sales of the grand cereals oil between the agents in each of these two age groups and those in the 38-46 years age category.

The model further shows that the male sales agents made about 9.3 million naira less than their female counterparts. The table shows that this coefficient is also statistically significant with a *p-value* of 0.000, which indicates that there is a highly significant difference in sales of the oil between the male and the female sales agents. The interaction effect (8.818) between the young sales agents (i.e., sales agents between 20 and 28 years of age) and the male sales agents (U_1V_1) indicates that the effectiveness of AGE variable on sales of grand cereals oil increased by about 8.8 million naira for males' sales agents compared to the female sales agents. From the coefficient table, we see that this coefficient is also significant with a p-value of 0.000.
The interaction effect (-2.989) between the middle-aged sales agents and the male sales agents (U_2V_1) indicates that the effectiveness of the GENDER variable on sales of grand cereals oil decreased by about 3 million naira for sales agents within the 29-37 years age group compared to those within the 38-46 years age group. From the coefficient table, we see that this coefficient is significant with a p-value of 0.000.

Now from Table 6,

$$R^2_{y,X} = 0.576$$

While from Table 9,

$$R^2_{y\,D_{Age1},D_{Age2},D_{Gender1},X} = 0.805$$

Thus, computing the difference, we have

$$R^2_{y\,D_{Age1},D_{Age2},D_{Gender1},X} - R^2_{y,X} = 0.805 - 0.576 = 0.229,$$

This difference indicates the unique contribution of the dummy variables to the first computed coefficient and also reflects the proportion of the total variation that is uniquely accounted for by the independent variables. The research findings indicate an equal mean and standard deviation of each group's sales points, with a sample size of 30.

Conclusions

The study revealed that the male sales agents achieved sales approximately 5½ million naira lower than their female counterparts. The interaction effect between young sales agents (20-28 years) and male sales agents (U_1 V_1) suggests that the effectiveness of the AGE variable on oil sales increased by approximately 3½ million naira for male sales agents compared to female sales agents. However, this coefficient is not statistically significant (*p =0.265*). The interaction effect between middle-aged sales agents and male sales agents (U_1 V_1) indicates that the effectiveness of the GENDER variable on oil sales decreased by approximately 13.4 million naira for sales agents in the 29-37year age group compared to those in the 38-46year age group. This work has established the potential of the ANCOVA technique in controlling and eliminating the concomitant variable effect.

These results imply that young male agents are more skilful (and thus high performers) in boosting oil sales than female agents of the same age range, and that the gender effect on sales performance is not the same for all age groups.

REFERENCES

Cox, D.R. & McCullagh, P. (1982). Some aspects of analysis of covariance. Biometrics 38, 541-561.

Engqvist, L. (2005). 'The Mistreatment of Covariate Interaction Terms in Linear Model Analyses of Behavioural and Evolutionary Ecology Studies', Animal Behavior 70, 967-971.

Field, A. (2011). Discovering Statistics Using SPSS. London, Sage Publications Ltd.

Fisher, R.A. (1935). The Design of Experiments. Edinburgh: Oliver & Boyd.

Howell, D.C. (2010). Statistical Methods for Psychology. Belmont, Wadsworth.

Kutner, M.H., Nachtsheim C.J., Neter, J. & Li, W. (2005). Applied Linear Statistical Models. New York, McGraw-Hill.

Miller, G.A. & Chapman, J.P. (2001). 'Misunderstanding Analysis of Covariance', Journal of Abnormal Psychology 110(1), 40-48.

Montgomery, S. (2012). Object-oriented information engineering: analysis, design, and implementation. Academic Press.