International Journal of Humanities and Social Science Research
ISSN: 2455-2070; Impact Factor: RJIF 5.22
Received: 23-01-2019; Accepted: 25-02-2019
www.socialsciencejournal.in
Volume 5; Issue 2; March 2019; Page No. 82-90

Applications of multifactorial experiments in petroleum distribution: NNPC as case study

Akintunde OA[1], Kazeem OO[2]
[1, 2] Department of Mathematics, Faculty of Science, Federal University, Oye-Ekiti, Ekiti State, Nigeria

Abstract
The petroleum industry, also known as the oil industry includes the global processes of exploration, extraction, refining, transporting, and marketing of petroleum products of the industry. At this time, crude oil is one of the most present and essential resources in everyday life. The oil industry is one of the most powerful branches in the world economy.
In the research study, two-level factorial design was performed to evaluate the effects of three variables and their interaction on the oil extraction. The variables included the petroleum products, years of distribution, and time (quarters of month) of distribution. The statistical analysis of the experimental data shows that the extracted products (PMS, HHK, ATK, and AGO) depends on all the examined variables. It also depends on the interactions between the petroleum products, quarters, and years. The experimental design method enables to determine the most important variables, their interaction and the most significant variables. The experimental data were in good agreement with those predicted by the model. The use and application of multiple regression model was also examined.

Keywords: petroleum, factorial design, interaction, regression model

1. Introduction

Mathematical Statistics is concerned not just with the analysis of data, but also with how data are collected. Such concern is quite natural, for analysis cannot be truly successful unless the data are informative, and the best way to assure informative data is to apply statistical principles to the way in which they are collected. One area in which statistics has made notable contributions to collecting informative data is Experimental Design [1]. However, with modern technological advances, products and processes are becoming exceedingly complicated, as the cost of experimentation rises rapidly. Hence, it is becoming increasingly difficult to investigate the numerous factors that affect these complex processes. Instead, a technique is needed that identifies the "vital few" factors in the most efficient manner [2].

Design of Experiment provides powerful and efficient methods to achieve these objectives, because it is the tool to develop an experimentation strategy that maximizes learning using minimum resources. Design of experiment is much more efficient than one-factor-at-a-time experiments, which involves changing a single factor at a time to study the effect of the factor on the process. Experimental design is the process of planning a study to meet specified objectives and can be used at the point of greatest leverage to reduce design costs by speeding up the design process, reducing late engineering design changes, and reducing product material and labor complexity. The design and analysis of experiments revolves around the understanding of the effects of different variables on another variable. In technical terms, the objective is to establish a cause-and-effect relationship between a number of *independent variables* and a *dependent variable* of interest. The dependent variable, in Design of Experiments, is called the *response*, and the independent variables are called *factors*. Experiments are run at different factor values, called *levels*. Each run of an experiment involves a combination of the levels of the investigated factors, and each of the combinations is referred to as *treatment.* When the same numbers of response observations are taken for each of the treatments of an experiment, the design of the experiment is said to be balanced. Repeated observations at a given treatment are called *replicates* [3].

This research work focuses its interest on factorial experiments, because it emphasizes the usefulness of varying several factors simultaneously in an experiment rather than just one factor at a time during the test. Factorial experiments are designed to evaluate multiple factors set at multiple levels and can be used when there are more than two levels of each factor. The factorial experiments, where all combinations of the levels of the factors are run, are usually referred to as full factorial experiments. Full factorial experiments with two level experiments are also referred to as 2^k designs where k denotes the number of factors being investigated in the experiment. In design of experiments, these designs are referred to as 2- Level Factorial Experiments.

Moreover, factorial experiments are designed to evaluate multiple factors set at multiple levels and can be used when there are more than two levels of each factor. When a design is denoted by a 2^3 factorial, this identifies the number of factors (3); and each factor having (2) levels with 8 experimental conditions.

Two levels factorial experiments are factorial experiments in which each factor is investigated at only two levels. Two level factorial experiments are used during experimentation to quickly filter out unwanted effects so that attention can then be focused on the important ones. The choice of two levels of factor depends on the factor; some factors naturally have two levels. For example, if gender is a factor, then male and female are the two levels. The two levels of the factor in the design are usually represented by either 0 (for the first level) and 1 (for the second level) or -1(for the first level) and 1 (for the second level).

2. Analysis of petroleum distribution in global world

The distribution of oil and gas reserves is majorly among the world's 50 largest oil companies and the reserves of the privately

owned companies are grouped together. The oil produced by the "supermajor" companies accounts for less than 15 percent of the total world supply. Over 80 percent of the world's reserve of oil and natural gas are controlled by national oil companies. The petroleum industry, also known as the oil industry includes the global processes of exploration, extraction, refining, transporting, and marketing of petroleum products of the industry are fuel oil and gasoline (petrol).
At this time, crude oil is one of the most present and essential resources in everyday life. The oil industry is one of the most powerful branches in the world economy. More than four (4) billion metric tons of oil is produced worldwide annually. Nearly one third of these amount is generated in the Middle East region, Saudi Arabia and the United State are the world's leading oil producers, each responsible for around 13 percent of the total global production. Russia is the third-largest producer, generating over 12 percent of the world's total oil production [4]. Oil account for a large percentage of the world's energy consumptions, ranging from a low of 32 percent for Europe and Asia, to a high of 53 percent for the Middle East other geographic consumption patterns are as follows; South and Central America (44 percent),Africa(41 percent), and North America (40 percent). The world consume about 30 billion barrels (4:8km3)of oil per year, with developed nations being the largest consumer. The United State consumed 25 percent of the oil produced in 2007.

3. Design and analysis of two level 2^k factorial experiments

3.1 2^2: Two –level two-factor factorial design

The complete interaction model for a two-factor design is:

$$y_{ijk} = \hat{\mu}_{...} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} + \varepsilon_{ijk} \; i = 1,2 \; j = 1,2,3 \; k = 1,2,\dots,n \quad (1)$$

Two factors $A \; and \; B$ are investigated at (a, b) levels. The mean response when A at the i^{th} level is denoted by $\mu_{i.}$ and B at the j^{th} level is denoted by $\mu_{.j}$

Thus $\mu_{i.} = \frac{\sum_j \mu_{ij}}{b}, \mu_{.j} = \frac{\sum_i \mu_{ij}}{a}, \mu_{..} = \frac{\sum_i \sum_j \mu_{ij}}{ab}$

The main effects in equation (1) are defined as: $\alpha_i = \mu_{i.} - \mu_{..} \; and \; \beta_j = \mu_{.j} - \mu_{..}$

with the two factors interaction in equation (1) is defined as:

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}$$

It should be noted that:

$$Y_{ij.} = \sum_k Y_{ijk} \; and \; \bar{Y}_{ij.} = \frac{Y_{ij.}}{n}, Y_{i..} = \sum_j \sum_k Y_{ijk} \; and \; \bar{Y}_{i..} = \frac{Y_{i..}}{bn},$$

$$Y_{.j.} = \sum_i \sum_k Y_{ijk} \; and \; \bar{Y}_{.j.} = \frac{Y_{.j.}}{an},$$

$$Y_{...} = \sum_i \sum_j \sum_k Y_{ijk} \; and \; \bar{Y}_{...} = \frac{Y_{...}}{abn}$$

$$\hat{Y}_{ijk} = \bar{Y}_{ij} = \frac{Y_{ij.}}{n}, \hat{\mu}_{ij} = \bar{Y}_{ij}, e_{ijk} = Y_{ijk} - \bar{Y}_{ijk} = Y_{ijk} - \bar{Y}_{ij}$$

Note: $\hat{\mu}_{...} = \bar{Y}_{...}, \hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}, \hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...} \; and \; \widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$

Thus it can be derived from the model in the equation (1) above that the Sum of Squares for Total (SST) will be:

$$SST = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - y_{...})^2$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (\hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} + \varepsilon_{ijk})^2$$

$$= \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \hat{\alpha}_i^{\,2} + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \hat{\beta}_j^{\,2} + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \widehat{(\alpha\beta)}_{ij}^{\,2} + \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \varepsilon_{ijk}^{\,2}$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{Y}_{i..} - \bar{Y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{Y}_{.j.} - \bar{Y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{ij})^2$$

$$= \left(\frac{Y_{i..}^{\,2}}{bn} - \frac{Y_{...}^{\,2}}{abn}\right) + \left(\frac{Y_{.j.}^{\,2}}{an} - \frac{Y_{...}^{\,2}}{abn}\right) + \left(\frac{Y_{ij.}^{\,2}}{n} - \frac{Y_{i..}^{\,2}}{bn} - \frac{Y_{.j.}^{\,2}}{an} + \frac{Y_{...}^{\,2}}{abn}\right) + \left(Y_{ijk}^{\,2} - \bar{Y}_{ij}^{\,2}\right)$$

$$\therefore SST = SSA + SSB + SSAB + SSE \quad (2)$$

Where $SSA = \frac{Y_{i..}^{\,2}}{bn} - \frac{Y_{...}^{\,2}}{abn}, SSB = \frac{Y_{.j.}^{\,2}}{an} - \frac{Y_{...}^{\,2}}{abn},$

$$SSAB = \frac{Y_{ij.}^{\,2}}{n} - \frac{Y_{i..}^{\,2}}{bn} - \frac{Y_{.j.}^{\,2}}{an} + \frac{Y_{...}^{\,2}}{abn} \ and$$

$$SSE = Y_{ijk}^{\,2} - \bar{Y}_{ij}^{\,2}$$

and $MSA = {SSA}/{(a-1)}, MSB = {SSB}/{(b-1)}, MSAB = {SSAB}/{(a-1)(b-1)}\ and\ MSE = {SSE}/{ab(n-1)}$

Using F-test to test equality of A, B and AB effect:

$F_{oa} = {MSA}/{MSE}, F_{ob} = {MSB}/{MSE}\ and\ F_{oab} = {MSAB}/{MSE}$

This is explained by the Analysis of Variance (ANOVA) shown in Table 1:

Table 1: ANOVA Table for Two –Level Two-Factor Factorial Design

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F-ratio
Factor A	SSA	a-1	MSA	$F_{oa\sim F_{(a-1),ab(n-1),\alpha}}$
Factor B	SSB	b-1	MSB	$F_{ob}\sim F_{(b-1),ab(n-1),\alpha}$
Factor (AB)	SSAB	(a-1)(b-1)	MSAB	$F_{oab\sim F_{(a-1)(b-1),ab(n-1),\alpha}}$
Error	SSE	ab(n-1)	MSE	
Total	SST	abn-1		

3.2 2^3: Two –level three-factor factorial design

The completely interaction model for a three-factor design is:

$$y_{ijkm} = \hat{\mu}_{....} + \hat{\alpha}_i + \hat{\beta}_j + \gamma_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk} + \widehat{(\alpha\beta\gamma)}_{ijk} + \varepsilon_{ijkm}\ \ i = 1,2,\dots,a \quad (3)$$

$j = 1,2,\dots,b$ $k = 1,2,\dots,c$ $m = 1,2,\dots n$

Three factors A, B and C are investigated at (a, b and c) levels. The main response when A is at the i^th^ level is denoted by $\mu_{i..}$, B at the j^th^ level is denoted by $\mu_{.j.}$ and C at the k^th^ level is denoted by $\mu_{..k}$

Thus $\mu_{ij.} = \frac{\sum_k \mu_{ijk}}{c}, \mu_{i.k} = \frac{\sum_j \mu_{ijk}}{b}, \mu_{.jk} = \frac{\sum_i \mu_{ijk}}{a}, \mu_{i..} = \frac{\sum_j\sum_k \mu_{ijk}}{bc}, \mu_{.j.} = \frac{\sum_i\sum_k \mu_{ijk}}{ac}, \mu_{..k} = \frac{\sum_i\sum_j \mu_{ijk}}{ab}$, and $\mu_{...} = \frac{\sum_i\sum_j\sum_k \mu_{ijk}}{abc}$

Main Effects: Main effects in a three-factor study are defined analogously to those for a two-factor study. Thus the main effects are defined as:

$$\alpha_i = \mu_{i..} - \mu_{...}\ \beta_j = \mu_{.j.} - \mu_{...}\ and\ \gamma_j = \mu_{..k} - \mu_{...}$$

Two-Factor Interaction Effects: The two-factor interaction effects in a three-factor are defined in the same manner as for a two-factor study:

$$(\alpha\beta)_{ij} = \mu_{ij.} - \mu_{i..} - \mu_{.j.} + \mu_{...}, (\alpha\gamma)_{ik} = \mu_{i.k} - \mu_{i..} - \mu_{..k} + \mu_{...}\ and\ (\beta\gamma)_{jk} = \mu_{.jk} - \mu_{.j.} - \mu_{..k} + \mu_{...}$$

Three-Factor Interaction Effect: $(\alpha\beta\gamma)_{ijk}$ is defined as the difference between the treatment mean μ_{ijk} and the value that would be expected if the main effect and first-order are sufficient to account for every factor effects.

It should be noted that:

$$Y_{ijk.} = \sum_{m} Y_{ijkm}\,, \bar{Y}_{ijk.} = \frac{Y_{ijk.}}{n}, Y_{i.k.} = \sum_{j}\sum_{m} Y_{ijkm}\,, \bar{Y}_{i.k.} = \frac{Y_{i.k.}}{bn}$$

$$Y_{ij..} = \sum_{k}\sum_{m} Y_{ijkm}\,, \bar{Y}_{ij..} = \frac{Y_{ij..}}{cn}, Y_{.jk.} = \sum_{i}\sum_{m} Y_{ijkm}\,, \bar{Y}_{.jk.} = \frac{Y_{.jk.}}{an}$$

$$Y_{.j..} = \sum_{i}\sum_{k}\sum_{m} Y_{ijkm}\,, \bar{Y}_{.j..} = \frac{Y_{.j..}}{acn}, Y_{i...} = \sum_{j}\sum_{k}\sum_{m} Y_{ijkm}\ \bar{Y}_{i...} = \frac{Y_{i...}}{bcn}$$

$$Y_{..k.} = \sum_{i}\sum_{j}\sum_{k} Y_{ijkm}\,, \bar{Y}_{..k.} = \frac{Y_{..k.}}{abn}, Y_{....} = \sum_{i}\sum_{j}\sum_{k}\sum_{m} Y_{ijkm}\ \bar{Y}_{....} = \frac{Y_{....}}{abcn}$$

$$\hat{\mu}_{ijk} = \bar{Y}_{ijk}, \bar{Y}_{ijkm} = \bar{Y}_{ijk}, e_{ijkm} = Y_{ijkm} - \bar{Y}_{ijk} = Y_{ijkm} - \bar{Y}_{ijkm}$$

$$\overline{\hat{\mu}_{....} = Y_{....}}, \hat{Y}_k = \bar{Y}_{..k.} - \bar{Y}_{....}, \hat{\alpha}_i = \bar{Y}_{i...} - \bar{Y}_{....}, \hat{\beta}_j = \bar{Y}_{.j..} - \bar{Y}_{....}, \hat{\gamma}_k = \bar{Y}_{..k.} - \bar{Y}_{....}$$

$$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{.j..} + \bar{Y}_{....}, \widehat{(\alpha\gamma)}_{ik} = \bar{Y}_{i.k.} - \bar{Y}_{i...} - \bar{Y}_{..k.} + \bar{Y}_{....}, \widehat{(\beta\gamma)}_{jk} = \bar{Y}_{.jk.} - \bar{Y}_{.j..} - \bar{Y}_{..k.} + \bar{Y}_{....}$$

$$\widehat{(\alpha\beta\gamma)}_{ijk} = \bar{Y}_{ijk..} - \bar{Y}_{ij..} - \bar{Y}_{i.k.} - \bar{Y}_{.jk.} + \bar{Y}_{i...} + \bar{Y}_{.j..} + \bar{Y}_{..k.} - \bar{Y}_{....}$$

Thus it can be derived from the model in the equation (3) above that the Sum of Squares for Total (SST) will be:

$$SST = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\mathrm{y}_{\mathrm{ijkm}} - \bar{\mathrm{y}}_{....})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\hat{\alpha}_{\mathrm{i}} + \hat{\beta}_{\mathrm{j}} + \gamma_{\mathrm{k}} + \widehat{(\alpha\beta)}_{\mathrm{ij}} + \widehat{(\alpha\gamma)}_{\mathrm{ik}} + \widehat{(\beta\gamma)}_{\mathrm{jk}} + \widehat{(\alpha\beta\gamma)}_{\mathrm{ijk}} + \varepsilon_{\mathrm{ijkm}})^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\hat{\alpha}_{\mathrm{i}})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\hat{\beta}_{\mathrm{j}})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\gamma_{\mathrm{k}})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} \widehat{(\alpha\beta)}_{\mathrm{ij}}^{\ 2} + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} \widehat{(\alpha\gamma)}_{\mathrm{ik}}^{\ 2}$$
$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} \widehat{(\beta\gamma)}_{\mathrm{jk}}^{\ 2} + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} \widehat{(\alpha\beta\gamma)}_{\mathrm{ijk}}^{\ 2} + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\varepsilon_{\mathrm{ijkm}})^2$$

$$= bcn\sum_{i=1}^{a} (\hat{\alpha}_{\mathrm{i}})^2 + acn\sum_{j=1}^{b} (\hat{\beta}_{\mathrm{j}})^2 + abn\sum_{k=1}^{c} (\gamma_{\mathrm{k}})^2 + cn\sum_{i=1}^{a}\sum_{j=1}^{b} \widehat{(\alpha\beta)}_{\mathrm{ij}}^{\ 2} + bn\sum_{i=1}^{a}\sum_{k=1}^{c} \widehat{(\alpha\gamma)}_{\mathrm{ik}}^{\ 2} + an\sum_{j=1}^{b}\sum_{k=1}^{c} \widehat{(\beta\gamma)}_{\mathrm{jk}}^{\ 2}$$
$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \widehat{(\alpha\beta\gamma)}_{\mathrm{ijk}}^{\ 2} + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (\varepsilon_{\mathrm{ijkm}})^2$$

$$= bcn\sum_{i=1}^{a} (\bar{Y}_{i...} - \bar{Y}_{....})^2 + acn\sum_{j=1}^{b} (\bar{Y}_{.j..} - \bar{Y}_{....})^2 + abn\sum_{k=1}^{c} (\bar{Y}_{..k.} - \bar{Y}_{....})^2 + cn\sum_{i=1}^{a}\sum_{j=1}^{b} (\bar{Y}_{ij..} - \bar{Y}_{i...} - \bar{Y}_{.j..} + \bar{Y}_{....})^2$$
$$+ bn\sum_{i=1}^{a}\sum_{k=1}^{c} (\bar{Y}_{i.k.} - \bar{Y}_{i...} - \bar{Y}_{..k.} + \bar{Y}_{....})^2 + an\sum_{j=1}^{b}\sum_{k=1}^{c} (\bar{Y}_{.jk.} - \bar{Y}_{.j..} - \bar{Y}_{..k.} + \bar{Y})^2$$
$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} (\bar{Y}_{ijk..} - \bar{Y}_{ij..} - \bar{Y}_{i.k.} - \bar{Y}_{.jk.} + \bar{Y}_{i...} + \bar{Y}_{.j..} + \bar{Y}_{..k.} - \bar{Y}_{....})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} (Y_{ijkm} - \bar{Y}_{ijkm})^2$$

$$= \left(\frac{Y_{i\ldots}^2}{bcn} - \frac{Y_{\ldots\ldots}^2}{abcn}\right) + \left(\frac{Y_{.j..}^2}{acn} - \frac{Y_{\ldots\ldots}^2}{abcn}\right) + \left(\frac{Y_{..k.}^2}{abn} - \frac{Y_{\ldots\ldots}^2}{abcn}\right) + \left(\frac{Y_{ij..}^2}{cn} - \frac{Y_{i\ldots}^2}{bcn} - \frac{Y_{.j..}^2}{acn} + \frac{Y_{\ldots\ldots}^2}{abcn}\right) + \left(\frac{Y_{i.k.}^2}{bn} - \frac{Y_{i\ldots}^2}{bcn} - \frac{Y_{..k.}^2}{abn} + \frac{Y_{\ldots\ldots}^2}{abcn}\right)$$
$$+ \left(\frac{Y_{.jk.}^2}{an} - \frac{Y_{.j..}^2}{acn} - \frac{Y_{..k.}^2}{abn} + \frac{Y_{\ldots\ldots}^2}{abcn}\right) + \left(\frac{Y_{ijk.}^2}{n} - \frac{Y_{ij..}^2}{cn} - \frac{Y_{i.k.}^2}{bn} - \frac{Y_{.jk.}^2}{an} + \frac{Y_{i\ldots}^2}{bcn} + \frac{Y_{.j..}^2}{acn} + \frac{Y_{..k.}^2}{abn} - \frac{Y_{\ldots\ldots}^2}{abcn}\right)$$
$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n}(Y_{ijkm} - \bar{Y}_{ijkm})^2$$

$\therefore\ SST = SSA + SSB + SSC + SSAB + SSAC + SSBC + SSABC + SSE$ (4)

Where $SSA = \frac{Y_{i\ldots}^2}{bcn} - \frac{Y_{\ldots\ldots}^2}{abcn}$, $SSB = \frac{Y_{.j..}^2}{acn} - \frac{Y_{\ldots\ldots}^2}{abcn}$, $SSC = \frac{Y_{..k.}^2}{abn} - \frac{Y_{\ldots\ldots}^2}{abcn}$, $SSAB = \frac{Y_{ij..}^2}{cn} - \frac{Y_{i\ldots}^2}{bcn} - \frac{Y_{.j..}^2}{acn} + \frac{Y_{\ldots\ldots}^2}{abcn}$, $SSAC = \frac{Y_{i.k.}^2}{bn} - \frac{Y_{i\ldots}^2}{bcn} - \frac{Y_{..k.}^2}{abn} + \frac{Y_{\ldots\ldots}^2}{abcn}$, $SSBC = \frac{Y_{.jk.}^2}{an} - \frac{Y_{.j..}^2}{acn} - \frac{Y_{..k.}^2}{abn} + \frac{Y_{\ldots\ldots}^2}{abcn}$, $SSABC = \frac{Y_{ijk.}^2}{n} - \frac{Y_{ij..}^2}{cn} - \frac{Y_{i.k.}^2}{bn} - \frac{Y_{.jk.}^2}{an} + \frac{Y_{i\ldots}^2}{bcn} + \frac{Y_{.j..}^2}{acn} + \frac{Y_{..k.}^2}{abn} - \frac{Y_{\ldots\ldots}^2}{abcn}$ and $SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n}(Y_{ijkm} - \bar{Y}_{ijkm})^2$

and $MSA = {SSA}/{(a-1)}$, $MSB = {SSB}/{(b-1)}$, $MSC = {SSC}/{(c-1)}$, $MSAB = {SSAB}/{(a-1)(b-1)}$, $MSAC = {SSAC}/{(a-1)(c-1)}$, $MSBC = {SSBC}/{(b-1)(c-1)}$, $MSABC = {SSABC}/{(a-1)(b-1)(c-1)}$, $MSE = {SSE}/{abc(n-1)}$

Using F-test to test equality of A, B and AB effect:

$F_{oa} = {MSA}/{MSE}$, $F_{ob} = {MSB}/{MSE}$, $F_{oc} = {MSC}/{MSE}$,

$F_{oab} = {MSAB}/{MSE}$, $F_{oac} = {MSAC}/{MSE}$, $F_{obc} = {MSBC}/{MSE}$ *and*

$F_{oabc} = {MSABC}/{MSE}$

This is explained by the Analysis of Variance (ANOVA) shown in Table 2:

Table 2: ANOVA table for two –level three-factor factorial design

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F-ratio
Factor A	SSA	a-1	MSA	$F_{oa \sim F_{(a-1),abC(n-1),\alpha}}$
Factor B	SSB	b-1	MSB	$F_{ob \sim F_{(b-1),abc(n-1),\alpha}}$
Factor C	SSC	c-1	MSC	$F_{oc \sim F_{(c-1),abc(n-1),\alpha}}$
Interaction AB	SSAB	(a-1)(b-1)	MSAB	$F_{oab \sim F_{(a-1)(b-1),abc(n-1),\alpha}}$
Interaction AC	SSAC	(a-1)(c-1)	MSAC	$F_{oac \sim F_{(a-1)(c-1),abc(n-1),\alpha}}$
Interaction BC	SSBC	(b-1)(c-1)	MSBC	$F_{obc \sim F_{(b-1)(c-1),abc(n-1),\alpha}}$
Interaction ABC	SSABC	(a-1)(b-1)(c-1)	MSABC	$F_{oabc \sim F_{(a-1)(b-1)(c-1),abc(n-1),\alpha}}$
Error	SSE	abc(n-1)	MSE	
Total	SST	abcn-1		

This can be extended to 2^k factorial designs and principles of Mathematical induction can as well be applied, as previously discussed in [5].

3.3 Regression approach to two –level three-factor factorial design

It is noted that the cell means model in equation (3) is a linear model. To develop a regression model, it is noted that each of the three factors is at two levels [6]. Hence, one indicator variable is required for each factor. The full regression model is:

$$Y_{ijkm} = \mu_{\ldots} + \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3} + (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3} \quad (5)$$

where X_{ijkm1} denotes the value of indicator variable X_1 for the jth case from ith factor levelX_{ijkm2} denotes the value of indicator variable X_2 for the jth case from ith factor level and so on Thus as shown/used in Tables 1 and 2,

$$X_1 = \begin{cases} 1\ if\ case\ from\ level\ 1\ for\ factor\ A \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ A \end{cases}$$

$$X_2 = \begin{cases} 1\ if\ case\ from\ level\ 1\ for\ factor\ B \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ B \end{cases}$$

$$X_3 = \begin{cases} 1 \ if\ case\ from\ level\ 1\ for\ factor\ C \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ C \end{cases}$$

The regression parameters in equation (7) are ANOVA/factorial design model parameters in equation (3). It is also noted that the sum of squares total and error (i.e SST and SSE) obtained from both the regression model and design model are the same, while the sum of squares regression equals the addition of SSA, SSB, SSC, SSAB, SSAC, SSBC and SSABC.
As regression approach can be used for all two-level multi-factor factorial design.

4. Analysis and discussion

This research work was carried our mainly on the annual statistical bulletin of Nigerian National Petroleum Corporation (NNPC). The data used for this study were drawn from the 2013 to 2016 NNPC monthly Petroleum Products Distribution. Four (4) products out of many products that are in annual statistical bulletin were used as resource data. It is assumed that the all the NNPC annual statistical bulletin of any year had an equal chance of being selected. The data were chosen by using the stratified sampling method. The four products used are:

1. PMS (Premium Motor Spirit) which represents the first product,
2. HHK(Household Kerosene) which represents the second product,
3. ATK (Aviation Turbine Kerosene) which represents the third product, and
4. AGO (Automobile Gasoline) which represents the fourth product

Also 2^3 factorial experiments (both balanced and coded) and regression analysis were done on the data.

4.1.1 Analysis from regression model

The regression model is given as:

$$Y_{ijkm} = \mu_{...} + \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3} + (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3}$$

with i = 1; 2; 3; 4 j = 1; 2; 3; 4 k = 1; 2; 3; 4 l = 1; 2; 3

Where α stands for the products distribution, β stands for the quarter of the products distribution and γ stands for the year of the products distribution, and αβ, αγ, βγ and αβγ stands for the interactions.

The parameters were thus estimated, resulting into

Y_{ijkl} = 1127603:560 - 2073426:442X_{1i} - 182646:800X_{2i} - 157400:435X_{3i} + 18753:588$X_{1i}X_{2i}X_{3i}$

It was tested whether there is relationship between the variables (products, quarters and years). This resulted in the use of the Table 3:

Table 3: Anova table for 2^3 regression analysis

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-ratio
Model	4	1.408×10^{13}	3.523×10^{12}	30.507
Error	187	2.159×10^{13}	1.155×10^{11}	
Total	191	3.568×10^{13}		

Since $F_{calculated}$ (= 30.507) > $F_{tabulated}$(= 2.37), it is thereby concluded that there is relationship among the products, quarters and years.

4.1.2 Experimental analysis for 2^3 factorial design

The complete interaction model for a three factor experimental design is:

$$y_{ijkm} = \hat{\mu}_{....} + \hat{\alpha}_i + \hat{\beta}_j + \gamma_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk} + \widehat{(\alpha\beta\gamma)}_{ijk} + \varepsilon_{ijkm} \quad i = 1,2,\dots,a \qquad (3)$$

$j = 1,2,\dots,b \ k = 1,2,\dots,c \ m = 1,2,\dots n$

with Table 4 below.

Table 4: Anova table for experimental analysis for 2^3 factorial design

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-ratio
A	3	1.877×10^{13}	6.2569×10^{13}	2.911×10^{3}
B	3	8.082×10^{9}	2.6943×10^{9}	1.254
C	3	6.289×10^{12}	2.0933×10^{12}	9.7539×10^{2}
AB	9	3.1917×10^{10}	3.5463×10^{9}	1.6501
AC	9	1.0251×10^{13}	1.139×10^{12}	5.299×10^{2}

BC	9	1.6917 x 10^{10}	1.8797 x 10^{9}	0.8746
ABC	27	4.3083 x 10^{10}	1.5957 x 10^{9}	0.7424
Error	128	2,751 x 10^{11}	2.1452 x 10^{9}	
Total	191	3.563 x 10^{13}		

It **was then concluded that**

1. There is relationship among the four products distributed (A).
2. There is relationship among the quarters of the distribution (B).
3. There is relationship among the years of the distribution (C).
4. There is relationship between the products and the quarters of the distribution (AB).
5. There is relationship between the products and the year of distribution (AC).
6. There is relationship between the quarters and years of distribution (BC).
7. There is relationship between the products, quarters, and years of distribution (ABC).

4.2.1. Analysis for 2^3 coded regression model

Like it was done in section 4.1.1, the resulting analysis yields the table blow using the principle of section 3.3:

Model:

$$Y_{ijkm} = \mu_{...} + \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3} + (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3}$$

Where

$$X_1 = \begin{cases} 1\ if\ case\ from\ level\ 1\ for\ factor\ A \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ A \end{cases}$$

$$X_2 = \begin{cases} 1\ if\ case\ from\ level\ 1\ for\ factor\ B \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ B \end{cases}$$

$$X_3 = \begin{cases} 1\ if\ case\ from\ level\ 1\ for\ factor\ C \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ C \end{cases}$$

and

$$X_4 = \begin{cases} 1\ if\ case\ from\ level\ 1\ for\ factor\ D \\ -1\ if\ case\ from\ level\ 2\ for\ factor\ D \end{cases}$$

which now gives

Y_{ijkl} = 296086:533 + 183093:546X_{1i} - 2974:078X_{2i} + 179956:466X_{3i} - 763:834$X_{1i}X_{2i}$ - 1982:783$X_{1i}X_{3i}$ + 512:347$X_{1i}X_{2i}X_{3i}$

With Table 5 below.

Table 5: Anova table for 2^3 coded regression model analysis

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-ratio
Model	6	1.266 X 10^{13}	2.109 X 10^{12}	16.947
Error	185	2.303 X 10^{13}	1.245 X 10^{11}	
Total	191	3.568 X 10^{13}		

Since $F_{calculated}$ (= 16.947) > $F_{tabulated}$ (- 2.10), it is thereby concluded that there is relationship among the products, quarters, and years.

4.2.2. Experimental design analysis for 2^3 coded factorial design

The complete interaction model for a three factor experimental design is:

$$y_{ijkm} = \hat{\mu}_{....} + \hat{\alpha}_i + \hat{\beta}_j + \gamma_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk} + \widehat{(\alpha\beta\gamma)}_{ijk} + \varepsilon_{ijkm}\ i = 1,2,\dots,a \quad (3)$$
$$j = 1,2,\dots,b\ k = 1,2,\dots,c\ m = 1,2,\dots n$$

with Table 6 below.

Table 6: Anova table for experimental analysis for 2^3 coded factorial design

Source of Variation	Degree of Freedom	Sum of Squares	Mean Squares	F-ratio
A	1	6.436×10^{12}	6.436×10^{12}	51.43583
B	1	1.704×10^{9}	1.704×10^{9}	0.013618
C	1	6.218×10^{12}	6.218×10^{12}	49.6936
AB	1	1.120×10^{8}	1.120×10^{8}	0.000895
AC	1	7.548×10^{8}	7.548×10^{8}	0.006032
BC	1	5.040×10^{7}	5.040×10^{7}	0.000403
ABC	1	5.040×10^{7}	5.040×10^{7}	0.000403
Error	184	2.302×10^{13}	1.251×10^{11}	
Total	191	3.568×10^{13}		

It was then concluded based on the above analyses that

1. There is relationship among the four products distributed (A).
2. There is relationship among the quarters of the distribution (B).
3. There is relationship among the years of the distribution (C).
4. There is relationship between the products and the quarters of the distribution (AB).
5. There is relationship between the products and the year of distribution (AC).
6. There is relationship between the quarters and years of distribution (BC).
7. There is relationship between the products, quarters, and years of distribution (ABC).

5. Conclusion

a. Based on Simple Regression Method: It was observed that there is relationship among the three variables examined (that is, petroleum products, quarters, and years).
b. Based on Experimental Design: It was observed that the four products (PMS, HHK, ATK, AGO) studied are significant, and there is relationship among the three variables (petroleum products, quarters, and years)

The research study recommends that Federal Government should focus and improve the rate of production and distribution of Petroleum Product in Nigeria. The study also recommends that Petroleum Product Distributor should use Statistics and its models effectively and efficiently because this will assist in improving the contribution which they (the Petroleum Product Distributor) make in the petroleum distribution.

6. References

1. Box GEP, Draper NR. Empirical Model Building and Response Surfaces, Wiley, Hoboken, NJ, 1987.
2. David M, Steinberg William G. Hunter Technical Summary Report, 2639-1984.
3. Murray RS, Larry JS. 2007; 4(2):205.
4. Johns HP. 2007; 27(3):225.
5. Akintunde OA. Design of Optimum Two-Level Multi-Factor Factorial Experiment, Journal of Nigerian Association of Mathematical Physics. 2017; 40:177-188.
6. William G. Cochran and Gertrude M. Cox Experimental Designs. A Wiley Interscience Publication. John Wiley & Sons Inc, 1992.
7. Box GEP, Behnken DW. Technometrics. 1960; 2(4):445-475.
8. Box GE, Hunter WG, Hunter JS. Statistics for Experimenters: Design, Innovation, and Discovery (2nd ed.). Wiley, 2005, ISBN 0-471-71813-0
9. Frank Y, Kenneth M. "Ronald A.F.". Biographical Memoirs of Fellow, 1963, 91-120.
10. Johns HP. The Asymmetric Effects of Changes in Price and Income on Energy and Oil Demands. The Energy Journal. 2007; 27(3):3.
11. Fisher RA. The Design of Experiments, Oliver Boyd, Edinburgh, Scotland, 1935.
12. Deming WE. Out of the Crisis, MIT Center for Advanced Engineering Study, Cambridge, MA, 1982.
13. Jaime JQ. Sadiq JZ. Experimental Design and Response Surface, 2014.
14. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Annals of Behavioral Medicine. 2005; 30(1):65-73.
15. Collins LM, Murphy SA, Strecher VJ. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. American Journal Preventive Medicine. 2007; 32(5):S112-8.
16. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. Psychol Methods. 2009; 14(3):202-24.
17. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. Psychol Methods. 2012; 17(2):153-75.
18. Apfel CC, Korttila K, Abdalla M, Kerger H, Turan A, Vedder I, *et al*. A factorial trial of six interventions for the prevention

of postoperative nausea and vomiting. N England Journal Of Medicine. 2004; 350:2441-245.
19. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM. Web-based smoking-cessation programs: results of a randomized trial. American Journal Of Preventive Medicine. 2008; 34(5):373-81.
20. Caldwell LL, Smith RA, Collins LM, Graham JW, Lai MH, Wegner L. Translational research in South Africa: evaluating implementation quality using a factorial design. Child Youth Care Forum. 2012; 41(2):11-36.
21. McClure JB, Derry H, Riggs KR, Westbrook EW, John JS, Shortreed SM, (2012): Questions about quitting (Q(2)): design and methods of a Multiphase Optimization Strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. Contemporary Clinical Trials. 2012; 33(5):1094-102.
22. Bolger N, Amarel D. Effects of social support visibility on adjustment to stress: Experimental evidence. Journal of Personality and Social Psychology. 2007; 92:458-475.
23. Wu C, Hamad AM. Experiments: Planning, analysis, and parameter design optimization. New York: Wiley, 2000.
24. Hays W L. Statistics. Orlando, Florida: Harcourt Brace & Company, 1994.