Semi-Latin squares and related "objects": *Statistics and Combinatorics aspects*

BY

PROF. POLYCARP .E. CHIGBU

Mr. Vice Chancellor Sir,
Distinguished Colleagues,
Ladies and Gentlemen,
Lions and Lionesses,

1. Introduction

1.1 Preamble

May I start my presentation today by thanking God Almighty and the Vice Chancellor for this opportunity of giving the 43rd inaugural lecture of our University on my 50th birthday, the first to be given in the year 2009 and not too long a time from the day my professorship at the University of Nigeria was formally announced.

Attaining the position of a Professor in a University of our kind would naturally require that one has contributed to knowledge in the various aspects of his/her discipline. Hence, it was not easy to choose what to embody and what not to embody in this lecture, especially as this is the first inaugural lecture from the department of Statistics, University of Nigeria and the third from our Faculty of Physical Sciences. Basically, the topic of my lecture is research-oriented. I have therefore considered the mode of its delivery to be more or less purely academic. However, I would most of the time endeavour to present it in simple Statistical and Mathematical language without distorting the real essence of the subject matter contained therein.

With regard to the topic of this lecture, the meanings of the main terminologies: Statistics, Combinatorics and semi-Latin square would be presented in sections 2 and Meanwhile, the word “object” as used in the topic and in this work simply means *configurations*. I shall have to give the implication of the term *configuration* more formally in section 2 but in the meantime, we bear in mind the following definition of *configuration* which pertains to this work and also given in The Concise Oxford Dictionary: “form, shape, or figure resulting from an arrangement of parts or elements in some manner” (Fowler and Fowler (1990)).

1.2 Basic Concepts

Before I get involved with discussing the main subject matter of this lecture, I crave your indulgence to take some time to briefly discuss its rudiments.

As you may be aware, the word *statistics* is usually used in two senses: first, it means numerical data relating to any field of endeavour, be it the Arts/Humanities, Sciences/Engineering or what have you; second, it refers to the scientific process for collecting, understanding, analyzing and interpreting numerical and non-numerical data. For the purpose of the second, let us consider that there exist two sets of people on earth: the "Scientists" and the Statisticians. While a "Scientist" is anybody who requires his/her statistical problem to be solved, the Statistician is anybody that has undergone the required standards of training in statistics for becoming and possesses the required skills of a Statistician; which also lends credence to statistical consulting.

Statistics is also an inductive science, which attempts to generalize concepts based on particular cases; deals with the whole based on information from a part; and draws inferences about populations on the basis of samples. A special ingredient for achieving these is *randomization* (a process whereby every member or item of a population of interest has equal chance of being observed as a member or item of an observable sample or part under investigation).

As also orchestrated by Nduka (2007), statistics as a subject is involved in almost all fields of endeavour, learning or study such that it is, for instance, studied as Biometrics in the Biological Sciences, Econometrics in Economics, Social Statistics in the general Social Sciences, Medical Statistics in the Medical Sciences, Environmental Statistics in Environmental Sciences, Official Statistics in Governmental and Non-governmental agencies, just to mention but a few. Statistics as a discipline studied at both undergraduate and postgraduate levels at the University of Nigeria and many tertiary level institutions in different parts of the world has among others the stress area known as Design and Analysis of Experiments often called Experimental Design or Design of Experiments (DOE).

While, the *analysis* phase of DOE involves the illustration of techniques, which enable the experimenter to analyze the experimental information in, say, a first- or second-order response (regression) model, the *design* phase involves the presentation and illustration of experimental layouts for the fitting of these models.

Historically, DOE and another stress area known as *Regression analysis* have developed separately. But it turns out that one of the best ways of appreciating the power of designed experiments is by first understanding regression analysis. Regression models are a statistical way of characterizing relationships between variables. A regression model can be defined in words as:

Y = function of X + random variation;
Symbolically written as:
$Y = f(X) + \varepsilon$

where Y represents a response variable called the dependent variable, X represents a predictor variable called an independent variable, $f(X)$ represents the systematic or repeatable part of the relationship between X and Y, ε represents the variation in Y which is not related to X or to any other measurable variable.

The main essence of DOE therefore is the designing of efficient and/or "best" experiments since life itself is full of experiments and experimentation on daily basis and whatever experiment that is being designed would need to be analyzed in some context by adopting certain appropriate techniques or criteria for interpretation. DOE entails the Statistics and Combinatorics aspects studied as Statistical and Combinatorial designs, respectively.

We all know that the term *statistical* is an adjective of the word *statistics*. However, the statistical design studies generally involve

(i) the determination of the relationships or otherwise between variables, attributes, criteria, properties or factors as might from time to time be represented by one regression (design) model or another;

(ii) the comparison of variables, attributes, etc, of an idealized representative model with the view of determining their effects in the representative model and/or the closeness of this model to a real life situation.

Contextually, it entails the hierarchical classification of a collection of configurations or design arrays, which possess a certain class of properties satisfying a particular design model, based on the comparison of variances of their treatments'/treatment contrasts' estimates or some other conditions which not only relate to the variances but on the nature of the incidence of treatments to the experimental units (plots).

The statistical design is usually presented from the standpoint of the general linear model, wherefrom least squares estimators are developed and discussed using possibly the notation of matrices. It also involves the use of designs (factorial and fractional factorial) to fit regression models with their attendant analysis-of-variance (ANOVA).

The term *combinatorial* is indeed a mathematical adjective pertaining to the word *combination*, which simply relates to the combination of items. The combinatorial design basically entails the exploitation of certain mathematical properties (or patterns) of configurations (designs), which are analogous to standard statistical basis of comparison or judgment (popularly called optimality criteria), in the making of good or "best"choices.

The combinatorial design involves studying the patterns of the application of treatments to plots in an experimental layout. In this regard, it can then be easily said that a particular pattern (graphically or otherwise or according to the level of treatment concurrences in plots) is better than another. To conduct meaningful research works in combinatorial design, one has to have full grip of the mathematical concepts of group and graph theory among others, which form part of the curriculum of a B.Sc. programme in mathematics.

I have taken all this time to make it clear that while the "Scientists" most times apply already-constructed designs (known in the literature as Classical designs) to their problems for analyses and interpretation, the Statisticians would most of the times be bothered with, among others, conceiving and constructing new designs (often called Optimal designs) for experimentation.

Thus, it is worthy to mention here that in many national and international conferences in statistics, topics on combinatorial designs or even general DOE are difficult to be accommodated appropriately for either Invited or Contributed presentation especially due to the fact that the right audience would hardly be found. Hence, there now exists a lot of combinatorial-based and DOE conferences where works on general DOE and especially combinatorial DOE gather the right calibre of attendees with the necessary skills to appreciate research outputs therefrom.

Indeed, the way and manner inputs (treatments) of an experiment are combined in application to plots determines the efficacy of the experiment and the fruitfulness of the output (yield).

There is no gainsaying the fact that the terms *treatment* and *plot* commonly used in DOE today originated from the interest of early British/European researchers like F. Yates and R.A. Fisher in this area whose interest was mainly on agricultural experimentation that could lead to bumper harvest amidst famine after the World wars of the 1930’s and 1940’s.

Mr. Vice Chancellor Sir, I have been conducting research in the area of statistical and combinatorial DOE for the past twenty two years and the subject matter treated here bothers on about 45% statistical DOE and 55% combinatorial DOE. The motivation for the higher proportion of my works on combinatorial DOE came from

(i) the need to construct new designs (forms of arranging/applying treatments to plots) to meet the need for solving new problems in this area; and (*ii* the supervisor of my research degree who is a properly-trained Abstract Algebraist-cum-Statistician.

From the foregoing, therefore, it is easy to see that, beyond the earlier senses I have attempted to describe Statistics, it is on the whole, a science of the "best" decision-making in experimentation. While a popular dictum states that Classical statistics dwells on making valid decisions under uncertainty, I will contextually state here that the aspect of statistics considered in this work specifically dwells on making optimum/optimal choices under many possibilities.

2. Some Relevant Terminologies and their Meanings

2.1 Experiment

Even though an experiment could simply be literally called a test: see Montgomery (1991), an experiment is a set of procedures which are carried out under a set of conditions, and which may occur repeatedly for a result: see, for example, Arua et al (2000).

In an experiment, one or more variables (or factors) are deliberately changed in order to observe the effect the changes have on one or more response variables.

2.2 Experimental Unit (Plot)

This is the smallest thing to which a treatment could be

applied. However, when a response is measured from this smallest unit, it is called an *observational unit*: see, for example, Bailey (2008).
In a real-life situation, the meaning of an experimental unit may vary from experiment to experiment. For example, while in an experiment involving the growing of varieties of a crop in genuine plots in a field, the *experimental units* are the *genuine plots*, the *experimental units* are the *sub-plots* in an experiment involving the growing of the same varieties of crop in *whole-plots* with fertilizers applied to *sub-plots*: see Bailey (2008).

2.3 Experimental Treatment

An *Experimental Treatment* is the entire description or totality of what is applicable to a plot at a given place or time, which gives rise to a measurable observation.

In the two real-life experiments of section 2.2, the *experimental treatments* are respectively, *variety of crop* and *variety-fertilizer combination*.

2.4 Design

Let W denote the set of plots of an experiment while T denotes a whole set of its treatments. A *design* can easily be said to be the assignment of treatments to plots, i.e. it is a function, f, which maps the elements of W to the elements of T (f: $\Omega \rightarrow T$). This implies that a plot ω E Ω, say, gets treatment, $f(\omega)$, during experimentation.

Bailey (1989) also defines *design* as a mapping or function, f, of the experimental units (plots), Ω, to treatments, T, in such a way that, treatments are allocated to plots. This definition seems to be circular in meaning, but it works well enough in practice.

In this regard, therefore, the usual aim of designing an experiment is to choose f such that certain combinatorial properties or patterns are satisfied or exploited, as the case might be.

2.5 Statistics and Statistical design

In section 1.2, some attempts have been made to define Statistics literally, technically and contextually. Here, we further state that in every statistical design, experimenters are usually interested in knowing about the effect of treatments applied and the plots on which they are applied. Usually, an experimenter has more control on the set of treatments applied than on the plots that receive these treatments.

Another usual issue of interest to experimenters is the allocation of treatments to plots: the randomization of this activity gives the statistical validity of the experiment, which forms the basis of any statistical design.

Thus, as highlighted in section 2.4, each design consists of two sets and a function between them. The sets are:

- a set that consists of treatments, denoted by *T*, say, and
- a set of experimental units, denoted by Ω.

These two sets are always finite.

On the whole, therefore, suppose Ω is a set of "things" (persons, animals, plants, machines, small portions of land, etc) upon which data are to be measured; a *factor* on Ω is a function, *f*, on Ω where we are interested not so much in the values of the function, *f*, as in which "things" have the same value of *f*.

Example: In a medical experiment involving a set, Ω, of persons, *f*(person) is the drug given to that person. To compare (the effect of) drugs, therefore, we need to know who and who had the same drug. The set of "things" with one given value of *f* is a subset of Ω. All such subsets form a *partition* of Ω and every"thing" is in one, and only one such subset: see, for example, Chigbu (1998a). In this regard, we shall always assume that the yield on plot ω, y_ω, is given by

$y_\omega = p_\omega + t_{f(\mathrm{w})}$, ……………………………………(1)

where $t_{f(\mathrm{w})}$ is a constant depending on the treatment, *f*(ω), applied to plot ω, p_ω is a random variable, depending on ω: see Chigbu (1995, 1998a).

Equation (1) is regression-based which leads to modeling statistical (regression-based) designs solved by the usual traditional regression analysis principles. This forms the basis of the Statistics aspect of our discussion here.

2.6 Combinatorics and Combinatorial design

The term *Combinatorics* was defined by Street and Street (1987), section 1.1 as “the branch of mathematics which deals with the problems of selecting and arranging objects in accordance with certain specified rules”. However, in studying Combinatorics, we always deal with Configurations.

Furthermore, Berge (1971, page 2) regarded Combinatorics as that which counts, enumerates (constructs and classifies), examines and investigates the existence of Configurations with certain specified properties or characteristics. In doing all these, the knowledge of the concept of isomorphism and isomorphism classes is quite essential.

It is not my idea to bore you with the details of technical mathematical terms, but briefly speaking and contextually, the *isomorphism* of two configurations implies their *sameness*. The definition of isomorphism as it pertains to semi-Latin squares would be given in section 3.

Hence, the Combinatorics aspects of the work reported in this lecture specifically involve the application of group and graph theory ideas in the construction and classification of semi-Latin squares and related "objects". These involve selecting and arranging the treatments of a semi-Latin square, say, based on the elements of the *Symmetric group*, S_n, and in accordance with the definition and constraints of a semi-Latin square. Eventually, different configurations and combinatorial patterns of the squares are grouped into two kinds of isomorphism classes based on their graphs ((treatment) variety-concurrence), family of the number of treatment-pairings which occur different number of times within the blocks of the squares known as Combinatorial parameters' method: see Preece and Freeman (1983), statistical design properties and the permutation sets associated with the arrangement of the treatments in the squares.

At this juncture, giving an illustrative example would help out in fixing the ideas of Combinatorial design fast. Thus, an experiment was conducted in an experimental area consisting of two fields, each divided into three strips of land, to compare three different varieties of corn: Red, White and Yellow, in combination with four amounts of phosphate fertilizer: 0, 70,140 and 210 (all in kg/hectare). Each strip was made up of four plots. The measured yield or response of this experiment was the total weight of the starch harvested from each plot.

The cultivation of the varieties of corn in the farms was mechanized such that varieties were sown on whole-strips and not in small areas due to practical possibilities. On the other hand, it was very convenient to apply the phosphate fertilizers to smaller areas of land within a whole-strip, here known as plots: see, for example, Bailey (2008). A typical layout of this experiment is given in Figure 1.

140	0	210
70	140	70
0	70	140
210	210	0
↑	↑	↑
Red	**Yellow**	**White**

0	70	140
70	140	0
210	0	210
140	210	70
↑	↑	↑
Yellow	**Red**	**White**

Figure 1: A typical layout of the Cultivation of Varieties of Corn

Considering Figure 1 for the above illustrative example, we notice the *existence of pattern* and the *lack of it*. The existence of pattern is depicted by the fact that each amount of fertilizer is applied to one plot per strip while each variety is applied to one strip per field. This pattern and studies pertaining to it manifests the combinatorial design investigation for this experiment. On the other hand, the existence of lack of pattern is depicted by, noting that there exists neither any systematic order in allocating the varieties to strips in each field nor any systematic order in the allocation of the amounts of fertilizer to plots in each strip. Having mentioned that Figure 1 is a typical layout of the experiment, it means that there could be other layouts, which derive their credence from some *randomization*. However, the lack of pattern of the above illustration, which is elicited by randomization subtly make for the statistical design discussed earlier in section 2.5.

Hence, Combinatorial design, as an integral aspect of the study of DOE (see, for example, Bailey (1991)), is "a way of choosing, from a given finite set, a collection of subsets with particular properties" (Street and Street (1987)).

The general algebraic connections between the analysis procedures of Statistical and Combinatorial designs are articulated in Chigbu (1998a).

3. What is a semi-Latin square?

3.1 Definition

With reference to the various ways the term semi-Latin square is defined in, for instance, Preece and Freeman (1983), Bailey (1988, 1990 and 1992), we hereby define a semi-Latin square as follows: An $(n \times n)/k$ semi-Latin square is an arrangement of nk symbols (treatments) in an $(n \times n)$ square array such that each row-column intersection contains k symbols and each symbol occurs once in each row and each column: see, for example, Chigbu (1995).

	1	…	*n*
1	*(1, …, k)*	…	*(1, …, k)*
. . .	. . .	… … …	. . .
n	*(1, …, k)*	…	*(1, …, k*

Figure 2: An array for an (n x n)/k semi-Latin square

In Figure 2, *(1, …, k)* indicates the size, k, of each block but not the actual symbols of a semi-Latin square. For convenience, a row-column intersection of a semi-Latin square is also called a *block* as long as no ambiguity of usage is introduced by doing this.

When n = 3 and k = 2, a typical semi-Latin square would be of the form,

1 2	3 4	5 6
3 4	5 6	1 2
5 6	1 2	3 4

Figure 3: A typical semi-Latin square for $n = 3$ and $k = 2$

where the symbols from 1 to 6 represent the six treatments of the square.

3.2 Historical Perspective

Here, I shall only give the historical background of semi-Latin squares up to 1992 when I first got involved with it.

Firstly, the origin of the Latin square could be traced back to the time experimenters in agricultural trials started considering two or more systems of blocking at once, which also involved the *Graeco-Latin squares*. However, the Graeco-Latin squares originated by the year 1782 when Euler invented the popular problem of thirty six army officers who were chosen, six from each of six different regiments, so that the selection from each regiment included one officer from each of six ranks, even though he later conjectured the impossibility of such arrangements: see Street and Street (1987), section 1.1. Long after 1782, R.A. Fisher called each of the two arrays that formed the Graeco-Latin squares, which Euler thought of adopting in solving his problem, the Latin square. Later on in 1926, Fisher gave information on the usage of Latin squares in experimental design problems.

General semi-Latin squares were so named by Yates (1935). Before this time, an example of this kind of arrangement for $n = 5$, $k = 2$ was given and called *equalized random blocks*: see “Student” (1931). Many agricultural and statistical writings of the 1930’s gave different examples of this kind of arrangement. One mention of semi-Latin squares in the 1940’s was by Ma and Harrington (1949) who studied the semi-Latin square design in agricultural field experiments. Preece and Freeman (1983) cited many German and Polish workers who, during the 1950’s and early 1960’s, used semi-Latin squares under the names *Lateinisches rechteck* (Latin rectangle) and *Cuadro latino modificado* (modified Latin square), respectively, in agricultural experiments. Harshbarger and Davis (1952) called the $(n \times n)/(n - 1)$ Trojan squares *Latinized rectangular lattices*; but Williams (1986) generalized their notion and called semi-Latin squares *Latinized incomplete-block designs*. Andersen and Hilton (1980) called semi-Latin squares *(1, 1, k) Latin rectangles*, while Rasch and Herrendorfer (1982, page 51) called them *pseudo latin squares*.

It can therefore be easily understood that semi-Latin squares have been used for agricultural experimentation since the inception of the last century.

3.3 Uses of semi-Latin squares

The statistical uses of semi-Latin squares are summarized in Preece and Freeman (1983) and Bailey (1992). One simple use of semi-Latin squares given by Bailey (1992) is in Consumer Testing where 8 new types of Vacuum Cleaners are tested by each of 4 housewives per week over a 4-week period. In this case, the plot is *one cleaning session per housewife* (two in each week). The treatments are the 8 Vacuum Cleaners, rows represent weeks and columns represent housewives.

A typical layout and square (configuration) for this use of semi-Latin square are shown in Figures 4 and 5, respectively.

Housewives

W	1	- -	- -	- -	- -
E	2	- -	- -	- -	- -
E	3	- -	- -	- -	- -
K	4	- -	- -	- -	- -

Figure 4: Layout of (4 x 4)/2 semi-Latin for the Consumer Testing Experiment

Housewives

W	1	1 2	3 4	5 6	7 8
E	2	3 4	1 2	7 8	5 6
E	3	5 6	7 8	1 2	3 4
K	4	7 8	5 6	3 4	1 2

Figure 5: A (4 x 4)/2 semi-Latin square

Each position marked with a dash in Figure 4 accommodates a particular Vacuum Cleaner while the symbols in each rowcolumn intersection of Figure 5 represent different Vacuum Cleaners. Also Figure 5 is a typical configuration as there are other possible configurations for the same description and this is why the configuration or arrangement of choice wouldpossess some inherent good statistical qualities, which depend on the nature of concurrences of the symbols in the row-column intersection of the array.

Apart from the above Consumer Testing example, which appears to be stereotyped against women especially as perceived by many in the western world, another way semi- Latin squares could be used which is analogous to the above is in Car Racing Championship competition where Housewives in the above example could be represented by Car Drivers while Vacuum Cleaners are represented by Cars. Weeks could remain the same or called Months. To appreciate this more, we consider the kind of monthly competition within a year associated with the (Formula One) Motor Grand Prix (GP) where the Championship competition involves determining the best driver and the best make of car/car constructors in a season. Such an "experiment" and the "yield" therefrom could be arranged in a semi-Latin square formation, where the "yield" here is indexed on the total number of points derivable from podium positioning all through a given season.

MONTHS	Drivers 1	2	3	4	5
1	1 2	3 4	5 6	7 8	9 0
2	9 0	1 2	3 4	5 6	7 8
3	7 8	9 0	1 2	3 4	5 6
4	5 6	7 8	9 0	1 2	3 4
5	3 4	5 6	7 8	9 0	1 2

Figure 6: A typical semi-Latin square for 10 Cars driven by 5 Drivers in 5 months in a Motor GP Competition

In Figure 6, the rows represent 5 months designated for the competition (one for each month), 5 columns represent 5 Drivers who are competing and the 10 symbols (0 to 9) in the row-column intersections are the treatments of the experiment, which are the Constructor's cars/types of car.

Another interesting illustrative example of a possible use of the semi-Latin square could be pitched on an arrangement or design formation where the rows of the semi-Latin square represent *Periods of Teaching*: morning, afternoon and night, say, or *School's Term of teaching*: first, second and third, columns represent *Teaching Method*, while treatments are people (students) who were taught by the use of the different Teaching Methods. From the earlier arrangements, I believe that the semi-Latin square formation for this

Teaching Method experiment can easily be deduced by anyone.

There has been an increased need for semi-Latin squares in analyzing agricultural and other types of experiments. This is mainly due to their usefulness, especially for small values of the number of treatments, *k*, per block of the square: see, for example, Bailey (1988, 1990 and 1992) for the practical uses of semi-Latin squares such as in Consumer Testing earlierdiscussed in this section, Glasshouse and Sugar beet trials, etc, Rojas and White (1957) and Darby and Gilbert (1958) for some other uses of semi-Latin squares.

3.4 Methods of Construction

The combinatorial possibilities of semi-Latin squares were first considered by Preece and Freeman (1983) who gave two methods of constructing the (4 x 4)/2 semi-Latin squares; *Trojan squares* and *Interleaving Latin squares*. The method of Trojan squares involves taking a set of *k* mutually orthogonal (*n* x *n*) Latin squares on *k* disjoint sets of symbols and superposing them. Each of the n^2 blocks of the square so formed contains the symbols, which occur in the corresponding blocks of the individual Latin squares.

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

α	β	γ	δ
γ	δ	a	β
δ	γ	β	α
β	α	δ	γ

Figure 7: Two mutually orthogonal (4 x 4) Latin squares

A α	B β	C γ	D δ
B γ	A δ	D α	C β
C δ	D γ	A β	B α
D β	C α	B δ	A γ

Figure 8: A (4 x 4)/2 Trojan square

By the superposition of the two (4 x 4) mutually orthogonalLatin squares in Figure 7, the (4 x 4)/2 Trojan square given in Figure 8 is formed. A Trojan square is a special sort of semi-Latin square.

The method of Interleaving squares is a generalization of the Trojan squares method. However, in this case, the *k* Latin squares used are not constrained to be mutually orthogonal. This method also involves the superposition of Latin squares. Bailey (1988) also gave some methods (inflation, explosion, superposition, etc), which could be used separately or in combination to construct semi-Latin squares.

3.5 Semi-Latin squares as Incomplete-block designs for analysis

In general, experimental designs need to be randomized in order to ensure that results of experiments are statistically unbiased and valid. In the literature, Nelder (1965), Preece and Freeman (1983) and Bailey (1992) recommend that the appropriate randomization of a semi-Latin square involve firstly, randomizing independently the rows and columns, and then randomizing independently the plots within each block. Based on the randomization procedure of these authors, the semi-Latin squares are recognized and statistically analyzed as a three-block-structured design, where the three types of blocks are the rows, columns and row-column intersections. Treatments are orthogonal to both rows and columns, which simply means that the treatments of a semi-Latin square are all accommodated in each row and each column.

Thus, a semi-Latin square is assessed for efficiency as abinary (0 or 1 observation of treatments in each plot) incomplete-block design where each row-column intersection is regarded as the block of its equivalent incomplete-block design and symbols its treatments: see Bailey (1988, 1992). It is a doubly-resolvable incomplete-block design and by ignoring its row and columns, it is called a quotient block design: see also Bailey (1988).

Based on some appropriate randomization described in the literature (Fisher (1935) and Nelder (1965)) the analysis of variance of the $(n \times n)/k$ semi-Latin square has four strata (sources of variation). The are rows, columns, blocks and plots with degrees of freedom $(n - 1)$, $(n - 1)$, $(n - 1)^2$ and $n^2(k - 1)$, respectively.

Bailey (1992) considered two types of models for the semi-Latin squares and highly recommended the one where the blocks having a random effect are included in the analysis. Thus, information about treatments is obtained from two strata: the blocks stratum and the plots stratum. The other model described by Bailey (1992), which ignores the blocks of a semi-Latin square in its analysis has no valid randomization procedure (Yates (1935)), in which case all semi-Latin squares are considered to be equally good with the same analysis.

An assessment of different designs of the same size can be made based on the amount of information, which can be recovered from both within and between blocks. Therefore, a basis for comparing different non-orthogonal block designs would need to be defined always and used. This basis is the *efficiency factors*, obtained only from the intra-block analysis. The optimality criteria, which in turn are based on the precision with which the estimates of treatment comparisons are made in the intra-block analysis, can be calculated from the efficiency factors.

Since for a semi-Latin square, treatments are orthogonal to (completely exhaust) rows and columns, semi-Latin squares whose incomplete-block designs possess high efficiency factors are always desirable.

The Statistical design analysis for the semi-Latin square emphasizes more on situations when there exist response data while its Combinatorial design analysis deals more on situations when there do not exist response data.

3.6 Semi-Latin squares as Graphs

The graphical representations of semi-Latin squares are quite vital in their classification especially as it is known in Paterson (1983) that graphical concepts were adopted to check the efficiency and optimality of incomplete-block designs.

Definition [Wilson (1990), section 1.1]: a *simple graph*, *G*, is a pair (V(*G*), E(*G*)), where V(*G*) (sometimes called the vertex-set) is a non-empty finite set of elements called vertices (nodes or points) and E(*G*) (also sometimes called the edgeset) is a finite set of unordered pairs of distinct elements of V(*G*) called edges (or lines).

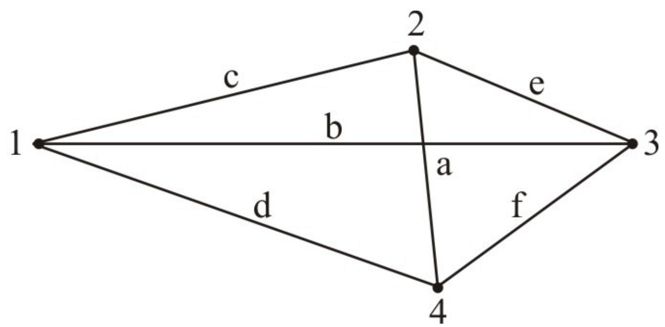

Figure 9: A simple graph, *G*

In Figure 9, V(*G*) = {1, 2, 3, 4} and E(*G*) = {a, b, c, d, e, f}. Most times, and as shown in Figure 9, graphs are defined with the restriction that any edge must join two distinct vertices. However, if this restriction is excluded such that we allow for edges to join vertices to themselves and multiple edges between vertices, then we refer to a *general graph* or simply a *graph*: see also Wilson (1990), section 2.2.

Definition [Wilson (1990), section 2.3]: Suppose that the vertex-set of a graph, *G*, can be split into two distinct sets V_1 and V_2 in such a way that every edge of G joins a vertex of V_2 to a vertex of V_2, then *G* is a *bipartite graph*. If *G* can be split into *k* disjoint sets such that no two vertices in a set are joined by an edge, then *G* is a *k*-partite graph.

When graphs with the same number of vertices and edges are examined an interesting problem, which often arises is to establish whether they are the same or not. It is this notion of sameness that necessitates the definition of the isomorphism of graphs, which has also been exploited in my works.

Definition: A variety-concurrence graph, $G(\Lambda)$, of a semi-Latin square, Λ, has treatments as vertices and the number of edges between any two treatments, s_1 and s_2, say, equals the number of blocks containing s_1 and s_2 in Λ.

When the variety (treatment)-concurrence graphs of semi-Latin squares are given (see, for example, Bailey (1992)), the general notion of relating graph theory to these squares is indeed implied. It is, at least, known that an inflated semi-Latin square could easily be differentiated from a Trojan square of equivalent size just by observing their variety-concurrence graphs.

Thus, for the three (4 x 4)/2 semi-Latin squares given by Figures 10, 11 and 12, their corresponding variety-concurrence graphs are given as Figures 13, 14 and 15, respectively.

12	34	56	78
34	12	78	56
56	78	12	34
78	56	34	12

Figure 10

12	34	56	78
34	12	78	56
78	56	12	34
56	78	34	12

Figure 11

12	34	56	78
34	12	78	56
56	78	14	32
78	56	32	14

Figure 12

(Semi-Latin squares)

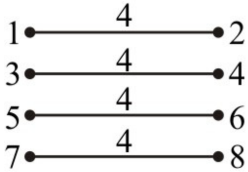

Figure 13

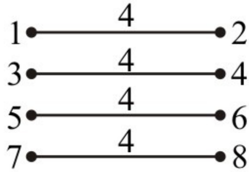

Figure 14

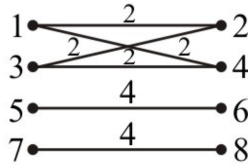

Figure 15

(Variety-Concurrence graphs of Figures 10, 11 & 12)

Therefore, it can easily be seen that the graphs of Figures 13 and 14 are quite different from the graph of Figure 15.

4. Related "Objects"

The word "objects" as used in quotes here simply implies array (arrangement of symbols or letters or things in rows and columns).

In defining the semi-Latin square, it is conventional to disregard the order in which k symbols are written in any row-column intersection. This is simply because in adopting the semi-Latin square for the purpose of experimental design, individual "little" columns have always been assumed to have no statistical role. A semi-Latin square can therefore be thought of as not so much as an (n x nk) (n rows and nk columns) Latin rectangle in one combinatorial sense, but rather as an (n x n) square array with each row-column intersection containing k unordered symbols: see Preece and Freeman (1983). On the other hand, the individual "little" columns might be judged to have some statistical role to play in the design of experiments. Though some extensions are possible, much depends on the roles attached to the rows and sets of "little" columns of the squares.

From the foregoing reasoning, therefore, different structures of designs, which are not semi-Latin squares per se, may arise. Each structure necessitates a particular way of viewing its isomorphism. These structures have been defined as *quasisemi-Latin square* but called *related "objects"* in this work.

A *quasi-semi-Latin square* is here defined as a combinatorial object whose entries are *ab initio* arranged as in the semi-Latin square formation without any regard to any other block structure apart from the one associated with the usual semi-Latin square but which actually has a peculiar blocking system as might eventually be defined as the case may be: see Chigbu and Oladugba (2008).

5. My Research Works/Contributions on semi-Latin squares and related "objects"

Mr. Vice Chancellor Sir, Ladies and Gentlemen, I have taken some reasonable time allocated to this lecture to attempt clearing some background due to the seemingly technical nature of my research activities, in general and my chosen topic, in particular.

In this section and subsequently, I shall specifically highlight the initial research problems/questions that existed in this area before I got involved in 1992, present my research and other academic contributions as precisely as possible and discuss specific issues concerning statistical consulting, mishandling of statistical data/methods and panacea.

5.1 Initial Research Problems/Questions

Before I delved into semi-Latin squares and related "objects", I started my research work with Response Surfaces Methodology (RSM) involving Spline functions. Then, I met my Ph.D. supervisor – Professor Rosemary A. Bailey who introduced semi-Latin squares to me. As highlighted earlier, she is an Algebraist, a Combinatorialist/DiscreteMathematician, a Statistician, and one of the world's foremost experts on experimental design. You could then imagine the uphill task associated with starting to work in this area especially with my B.Sc. and M.Sc. Statistics degrees' background from University of Nigeria where few mathematics courses formed part of the curricula then. Of course, with some re-training and training in more Mathematics courses and by dint of hard work, I was able to cope.

The initial research problems could be summarized as follows.

(i) Semi-Latin squares with $n = k = 4$ were needed for experiments involving Banana trees in Jamaica, as indicated by Bailey (1992). Trojan squares are known to be A-, D- and E-optimal: see Cheng and Bailey (1991); but no Trojan square exists if $k \geq n$ and in particular there is no Trojan square with $k = n = 4$. Bailey (1992) then gave two semi-Latin squares of this size with different sets of concurrence parameters and the same good canonical efficiency factors but neither knew if they were optimal nor relate canonical efficiency factors to known optimality theorems to cover this type of design: see, for example, John and Mitchell (1977), John and Williams (1992), Paterson (1983) and Bailey (1992).

(ii) No method of construction (inflation, superposition, etc) of semi-Latin squares in existence then or even a combination of them could be used to construct all possible semi-Latin squares of a given size. Thus, one of the interests then was to give a method which generates systematically all possible semi-Latin squares for given values of n and k and especially for $n = 4$ with the hope that the method could be extended for slightly bigger values of n and k.

(iii) Even after constructing all possible semi-Latin squares of a given size, the existing methods of classifying them (Combinatorial parameters' and Variety-concurrence graphs' methods) into different equivalence classes (species and transformation sets of Fisher and Yates (1938) and Preece and Freeman (1983)) were cumbersome, especially when many possibilities were the case.

Let $\boldsymbol{a}_i$ be the number of symbol-pairings that occur i times (i = 0, 1, …, n) within the blocks of a semi-Latin squares. The family $(a_1)_{i=1}^{n}$ of any semi-Latin square is known as the Combinatorial Parameters of that square: see Preece and Freeman (1983).

There is no gainsaying the fact that the initial research problems/questions led to a lot more which needed to be solved then and now.

5.2 My Contributions

5.2.1 Statistics aspects:

(i) Three A-, D- and E-optimal (4 x 4)/4 semi-Latin squares including the two good ones by Bailey (1992) have been ascertained: see Chigbu (1995 and 1999a). The squares have the same canonical efficiency factors. The third one, which was originally found by me, is given in Figure 16.

(ii) By considering and analyzing each of the semi-Latin squares of a given size as an incomplete-block design without response data and comparing them, a lemma about the equality of the harmonic mean of the canonical efficiency factors and the overall efficiency factor based on the average variance of the pairwise treatment differences of a connected design is given and proved: see Chigbu (1998a & b).

(iii) The Graeco-Latin square and the Trojan square designs were considered with the view of discriminating one kind of design from the other as well as highlight their relationships, where possible: see Chigbu (2001a).

(iv) A program in *Genstat*, which is amenable to computing in both UNIX and DOS environments using the standard *Genstat* commands for calculating the canonical efficiency factors and the A-, D- and E-optimality criteria of incomplete-block designs (IBD's) with no response data and l^2 number of blocks, where l^2 is any positive integer for which an incomplete-block design exists has been given: see Chigbu (2001b).

(v) The "best" of the three (4 x 4)/4 semi-Latin squares (Figure 17) have also been ascertained by, analytically finding and comparing the variances of elementary contrasts of treatments for the squares: see Chigbu (2003).

A α a 1	B β b 2	C γ c 3	D δ d 4
B γ d 2	A δ c 1	D α b 4	C β a 3
C δ b 4	D γ a 3	A β d 1	B α c 2
D β c 3	C α d 4	B δ a 2	A γ b 1

Figure 16: The (4 x 4)/4 semi-Latin square, found by me

A α a 1	B β b 2	C γ c 3	D δ d 4
B β c 4	A α d 3	D δ a 2	C γ b 1
C γ d 2	D δ c 1	A α b 4	B β a 3
D δ b 3	C γ a 4	B β d 1	A α c 2

Figure 17: The "best" (4 x 4)/4 semi-Latin square

(vi) A numerical approach, which basically involves the computation of the generalized inverses of the information matrices of the three optimal semi-Latin squares for sixteen treatments in blocks of size four, has been adopted to further classify them: see Chigbu (2004).

5.2.2. Combinatorics aspects:

- An (n x n)/k semi-Latin square is regarded as a collection of k permutations in S_n or a family of *nk* permutations of *n* objects subject to certain definitional restrictions so that it can be constructed: see Bailey and Chigbu (1997) as well as for the following description of the systematically group-theoretic method of constructing semi-Latin squares.

Thus, let L be an $(n \times n)/k$ semi-Latin square where we always assume that the rows and columns of L are labeled 1, …, *n*. Let $X^\wedge$ be the set of letters in L where the superscript would always be omitted if there is no ambiguity. Then, for *i.,j* E {1, …, *n*}, let Λ_{ij} be the set of letters in $X^\wedge$ which occur in row *I* and column *j* of L, each letter *x* in *X* determines a permutation $\boldsymbol{\pi}^\wedge_x$ in S_n by

$i\,\boldsymbol{\pi}^\wedge_x = j \Leftrightarrow x \in \Lambda_{ij}\, for 1 \leq i, j \leq n.$

On the other hand, each permutation σ in S_n determines a subset $Y^\wedge_\sigma$ of *X* by

$Y^\wedge_\sigma = \{\, x \in X \colon \boldsymbol{\pi}^\wedge_x = \sigma \,\},$

We write $N^\wedge_\sigma$ for $|\, Y^\wedge_\sigma|$ and therefore can clearly see that

$\sum_{\sigma \in S_n : i\sigma = j} N^\wedge_\sigma = k$ for *i.,j* E {1, …, *n*} ………………(2)

Moreover, if the semi-Latin square *M* is obtained from Λ simply by relabelling the letters then $N^\wedge_\sigma = N^M_\sigma$ for all σ in S_n. Furthermore, given any family of non-negative integers $(N_\sigma : \sigma \in S_n)$

satisfying equation (2), then there exists a semi-Latin square, Λ, such that $\hat{N}_{\sigma} = N_{\sigma}$ for all σ in S_n.

To illustrate the above algebraic derivation so that you do not get put off by one of the major contribution of my research endeavour,
let $n = k = 3$ and put

$\Lambda =$

	1	2	3
1	*a b c*	*d e f*	*g h i*
2	*f g h*	*a b i*	*c d e*
3	*d e i*	*c d h*	*a b f*

then, $\pi_f = (1\ 2)$, $\pi_g = \pi_h\ (1\ 3\ 2)$,
$Y_{(1\ 2)} = \{f\}$, $Y_{(1\ 3\ 2)} = \{g, h\}$,
$N_{(1\ 2)} = 1$, $N_{(1\ 3\ 2)} = 2$.

Semi-Latin squares may therefore be identified with families of non-negative integers (N_{σ} : σ E S_n) satisfying equation (2). Since N_{σ} can take only $k + 1$ values, the family can be succinctly represented by the partition of S_n into subsets
$\Lambda_0, \Lambda_1, \ldots, \Lambda k$, where
$\Lambda_r = \{ \sigma$ E $S_n : \hat{N}_{\sigma} = r \}$.

- Semi-Latin squares of a given size were made to fall into *strong isomorphism classes* (interchange of rows and columns not permitted), which are grouped into *weak isomorphism classes* (interchange of rows and columns permitted). Hence, group theory, graph theory, design theory and computing (via the *Nauty* package (McKay (1990)) and regarding the semi-Latin squares as graphs) were used to find all weak and strong isomorphism classes of (4 x 4)/*k* semi-Latin squares for $k = 2$ or 3 or 4: see Chigbu (1995) and Bailey and Chigbu (1997).

While considering semi-Latin squares as graphs for isomorphism classification via the *Nauty* package, we identified five types of vertex for each square, namely: the row-type, the column-type, the treatment-type, the position-type and the extra-type vertices. The number n of rows, the number k of treatments per block and the number nk of treatments are predetermined. Then, the number n^2k of positions in a square and the total number υ of its vertices are calculated. Semi-Latin squares with different values of υ cannot be compared for isomorphism.

In general, the five types of vertices are labeled from 0 to υ - 1, and with respect to Figure 11, for instance, the relevant parameters are: $n = 4$,
$k = 2$, $nk = 8$, $\upsilon = 50$ and the vertex-labels are:
- o row-type: 0, …, 3,
- o column-type: 4, …, 7,

- o treatment-type (for treatments 1, 2, 3, 4, 5, 6, 7,8): 8, …, 15,
- o position-type (for all the 32 entries of the square): 16, …, 47,
- o extra-type: 48, 49.

An example of the adjacencies given for the first position (vertex 16) is {0, 4, 8}, which represents the first row, first column and first letter (symbol); other adjacencies are determined accordingly: see Chigbu (1995) and, for example, Bailey and Chigbu (1997).

In Bailey and Chigbu (1997) also, five theorems: two on the strong and weak isomorphism of the general $(n \times n)/k$ semi-Latin squares were given and proved analytically; while three on the number of the strong and weak isomorphism classes for the $(4 \times 4)/k = 2, 3$ & 4 semi-Latin squares were made evident and justified by some complete search.

- Chigbu (1999b) explains what isomorphism of semi-Latin squares means and distinguishes semi-Latin squares from somewhat similar designs (already defined as quasi-semi-Latin squares in section 4) with different block structures.

Nelder (1965) defines the *simple block structure* as any arrangement involving nesting (usually denoted by →), crossing (usually denoted by x) or both with suitable brackets to indicate the order of combination. The arrangement of an (n x n)/k semi-Latin square is the same as the simple block structure and designs with this type of arrangement fall into the category of row-column designs with nested experimental units: see Bailey (1993).

Thus, the way the block structure of a particular design is defined necessitates a particular way of viewing its isomorphism.

At present, a number of quasi-semi-Latin squares are being examined with the purpose of determining their null and full analysis-of-variance models and strata which should be different from those of semi-Latin squares.

- Group-theoretic lemmas which would enable one find the non-isomorphism classes of the (n x n)/k semi-Latin squares when the sizes and number of squares under consideration are large, and the computing method via *Nauty* is not readily available, were given with proofs: see Chigbu (2001c).

- A program in *Qbasic* language that enables one generate and construct the $(n \text{ x } n)/k$ semi-Latin squares which were originally generated and constructed via the systematic group-theoretic algorithm of Bailey and Chigbu (1997) has been given and discussed: see Chigbu and Eze (2001).

6. My Other Academic Contributions

Since my occupation is essentially Teaching and Research, I think that in an occasion of this sort, it is worthy to highlight the outcome of my research works in some other areas other than semi-Latin squares, my postgraduate supervision, conference attendance and teaching activities.

6.1. Some Research Works on other areas of DOE and Statistics

Aside from the contributions of section 5 above, my research has also involved the Polynomial Spline functions (piecewise polynomials of degree m, say, $(m > 0)$ whose functional values and first $(m - 1)$ derivatives coincide at the points where they join called knots (Smith (1979)) where: (i) the optimal positions of their knots were determined by varying certain optimality criteria, which were also explored. Here, the first- and second-order polynomial spline regressors were established to be optimal if their knots are placed at equal intervals along each axis: see

Chigbu (1991); (ii) the performance of the design matrices obtained from polynomial classical and spline regression functions for a given experimental design was compared using the popular optimal design criteria. The polynomial classical regressors with observations on equally spaced levels or support points of the experimental design were found to be better than the polynomial spline regressors with respect to the optimal design criteria considered: see Chigbu and Nduka (2005).

I have also done some other research works on Response Surfaces Methodology (RSM), which is popularly seen as a bridging link between the subject of DOE and the subject of Unconstrained Optimization, especially via the Super Convergent Line Series (SCLS) (SCLS is a line series for optimization in mathematical programming, with unique nice properties for convergence not possessed by other line series): see Onukogu and Chigbu (2002), chapter 5, Chigbu and Ugbe (2002) and Chigbu and Ukaegbu (2007).

In the area of the optimization of Linear Programming Transportation Problems, the following papers exposes my contributions: Chigbu and Udoh (2002, 2006) and Udoh and Chigbu (2008).

6.2. Postgraduate Supervision and Conference Attendance

I have been involved in Postgraduate supervision in mainly the broad areas of Design of Experiments and General Optimization techniques at M.Sc. level for over ten years and Ph.D. level for about five years. To date, I have successfully supervised ten M.Sc. projects most of which have either been published or are publishable and currently supervising seven M.Sc. and six Ph.D. students. A Ph.D. thesis I co-supervised was recently examined successfully.

With regard to subjecting the outcome of my research activities to regular scrutiny by professionals and colleagues elsewhere in the world, I regularly attend and present papers at relevant national and international conferences as could be found documented on my curriculum vitae. Recently, I attended the Royal Statistical Society Conference (September 1 – 5, 2008), University of Nottingham, U.K. and a picture of mine taken by the organizers during my presentation at the conference appeared on the cover page of the Royal Statistical Society (RSS) News of November 2008 (see Appendix).

6.3. Teaching

My teaching experience at this University spans over twenty-three years excluding the period of my study leave in United Kingdom (1992 – 1996) (during which I occasionally taught Statistics courses at undergraduate level). For the rest of the time, I started my teaching career at our University with the teaching

of mostly Statistics service courses to Social Sciences, Biological Sciences and Engineering students. Subsequently, I have over the years taught almost all undergraduate Statistics courses at one time or the other and Design and Analysis of Experiments at postgraduate level in my department. I have also supervised many B.Sc. projects in Statistics.
For a substantial number of years from the late 1990's, I also taught both undergraduate and postgraduate Quantitative Economics courses: Mathematical Economics, Econometrics, etc, at the department of Economics of this University.

7. Desiderata

Mr. Vice Chancellor Sir, please permit me to digress a bit to dwell on a number of things enormously lacking but greatly desired in the attainment of fulfilled Statistical career especially as we are celebrating my inauguration to the Chair of Statistics of this University.

7.1. Basics of Statistical Practice and Consulting

In attempting to elucidate the proper procedures of Statistical Practice and Consulting especially for self-employment, one needs to have an overview of the subject matter – DOE vis-àvis the various stages of designing efficient experiments. Indeed, there are a number of things that should be put into consideration but however, often neglected, if a proper statistically designed experiment is to be achieved.

Every practicing Statistician who designs experiments should realize that a good design considers *plots* and *treatments* first, and then allocates treatments to plots. A choice of a design for experimentation should not just be made from an already existing list of named designs. Emphasis should always be laid on choosing a design, which answers the research question under consideration instead of just resorting to analysis involving a given design purported to be adaptable for the research question.

Having declared earlier that the world is just that of the Statisticians and the "Scientists", the first stage of designing an efficient experiment involves Consultation where both the "Scientist" and the Statistician need to collaborate, arrive at a consensus well ahead of the time of embarking on the experiment. At this stage, the Statistician should have at the back of his/her mind that the "Scientist" would hardly present his/her problems with the required statistical precision.

However, it is expected that the experience of the Statistician count so much here since with careful cross-examination and persuasion of the "Scientist" the limited knowledge of the "Scientist" on the right way forward will be circumvented.

The second stage involves the determination of a suitable statistical design for a given problem especially if the "Scientist" comes forth with some recommendation of a design from a shortlist of the ones he/she knows.

Third stage is about the collection of data where the Statistician, in collaboration with the "Scientist", prepares a data-collection format for the collection of all relevant information/data as long as it is available with no modifications of any sort.

The fourth stage demands that the Statistician scrutinizes the data collected by the "Scientist" with the view of detecting obvious anomalies and malpractices. In this regard, the Statistician needs to promptly query dubious data while the
activities of the "Scientist" are still fresh in his/her memory.

The fifth stage is on Analysis proper. Some traditional and necessary calculations mapped out at the Consultation stage with the data obtained would then be made. ANOVA tables, p-values, means and standard errors, etc should be computed. However, some modification on earlier proposed things to calculate could be made in the light of further development during experimentation. One way of appreciating what one does here is by first of all knowing how to calculate things by hand/using pocket calculators. A reliable statistical computing package is inevitable especially for large data set.

The final stage involves Conclusions and Interpretation. The presentation of the outcome of standard calculations means little or nothing to the "Scientists". However, proper interpretation of the results by a skilled Statistician by the use of terms the "Scientist" would easily understand would suffice. The above procedures were also discussed in Bailey (2008).

7.2. General Mishandling of Statistical data/methods and Panacea

In Nigeria today, Statistics has not been accepted as a professional course as the case in the developed world like United Kingdom. As a result of this, few people take to studying Statistics at University level. This has led to the existence of very few properly skilled Statisticians. Currently, in Nigeria, the number of Professors of Statistics is not more than twenty-five but there exist about forty degree programmes in Statistics run by various Universities. The dearth of Professors and other cadres of lecturers of Statistics in Nigeria have led to the production of a few properly trained and qualified Statisticians. Consequently, a lot of people without the required statistical skills handle and analyze essential statistical data wrongly or even secure employment to teach Statistics at the various levels of education. Even the basics of appreciating the four (nominal, ordinal, interval and ratio) scales of measurement of data before analysis are thrown to the wind. Once statistical data are handled wrongly without this appreciation, a lot of distorted and unreliable analysis, conclusions and interpretation occur. Based on the foregoing, it is pertinent to, at this juncture, assert as follows.

(i) There should be no over-ambitiousness in the adoption or application of statistical methods in analysis and problem solving. It does not really matter what statistical method is adopted/applied, whether pictorial (using bar charts, pie charts, histograms, etc), descriptive (using means, variances, etc) or even some sophisticated methods of analysis, as long as the method used conforms with the type of data under consideration.

(ii) A great proportion of the populace of a nation is exposed to high risks or danger when wrong decisions arising from wrong statistical approaches to problems by unskilled Statisticians are made.

Some analogy of the kind of danger involved here could be drawn from the danger associated with graduating a half-baked, unskilled medical doctor who opens a clinic in an environment of uninformed citizens, who before they know it, people would have been dying in droves in his clinic. Well, of course, as the news of the casualties spread more people would desist from going to the clinic since their lives are in serious danger. For an incompetent Statistician, in this regard, more people would have been "killed" without their knowing it since applying statistical methods wrongly would lead to a whole lot of wrong decisions and planning affecting humanity, for instance, by governmental and non-governmental organizations which would subsequently lead to even higher rate of mortality.

(iii) Younger people who have good quantitative aptitude and have shown reasonable interest in studying Statistics at tertiary level education should be encouraged and motivated by their being awarded grants and scholarships to undergo their courses by the Federal Ministries of National Planning and Education and/or their agencies as well as multinational organizations.

(iv) The National Bureau of Statistics (NBS) need to be properly financed by the Federal Government of Nigeria so that it could perform its statutory duties which among others include reliable surveys and the collection of genuine data for analysis and planning. Towards the end of last year, it was reported in the news media that most of the functions of NBS were not to be performed during the last financial year but for the financial assistance of some international agencies. If so, it is a matter that should be seriously looked into by the government.

8. Conclusions

Mr. Vice Chancellor Sir, within the last one hour or thereabout, I have made attempts to define and discuss Statistics, Combinatorics and semi-Latin squares and related "objects", in all their ramifications. In the process of my presentation, I have also highlighted issues pertaining to the Statistical and Combinatorial design of experiments and especially the aspects of Statistics and Combinatorics studies associated with research works on the combinatorial "object" known as semi-Latin square. Above all, my research contributions over the years have been exposed and some key issues associated with proper Statistical practice and consulting discussed.

At this juncture, I think it is worthwhile to put on record that a number of persons and bodies have given me various kinds of support in my academic pursuit to this level and in making life worth living. However, as this afternoon's assignment of mine is mainly academic, my expression of appreciation would mostly be in relation to academic issues.

In this regard, therefore Mr. Vice Chancellor Sir, I, once more, crave your indulgence to thank you again, and express my profound gratitude to the following:

- My parents – Peter (Late) and Sussana Chigbu for insisting on and providing for my early education;
- Sir Walter Atufunwa for all his advice when I first arrived Nsukka for my B.Sc. programme, which enabled me settle down to studies fast;
- Professor I.B. Onukogu, my M.Sc. supervisor and mentor at the University of Nigeria, and Professor G.E.O. Ogum of Nnamdi Azikiwe University, Awka for all his encouragement;
- Professor O.V. Ekechukwu (now at National Universities Commission, Abuja) for his wonderful counsel all through my academic career from 1984 to this stage, and special friendliness and brotherliness;
- Professor P.I. Uche (my undergraduate Academic Adviser and Lecturer), my Lecturer – Chief C.N. Onuoha and other Lecturers of mine, distinguished colleagues and friends at the University of Nigeria, Nsukka for making it possible for me to learn a lot academically and otherwise;
- My dear wife, Baptista, and our children: Ifeoma, Chinanu, Ozioma, Okwuchi and Chinwoke, for giving me the day to day support

required for my works and for understanding even when they should not;

- Professors D.C. Onyejekwe, J.U. Oguakwa, D.I. Nwoga (Late), M. Ikeme (Late) for their advice and assistance before my proceeding for my Ph.D. degree programme in the United Kingdom;
- Professor J.N. Adichie for discovering to believe in me and my research capabilities and all his encouragement;
- The University of Nigeria, Commonwealth Scholarship Commission in the U.K. and The British Council for making it possible for me to go to Goldsmiths' University of London through the Commonwealth Academic Staff Scholarship award;
- Professor Rosemary A. Bailey, my Ph.D. supervisor, for introducing semi-Latin squares to me, inspiring me in my research endeavour and meticulously supervising my work with unrelenting enthusiasm;
- Mr. and Mrs. Jude O. Oparah, who in a variety of ways helped in making the period of my Ph.D. programme in London fruitful, for their consistent hospitality during my academic visits to U.K.
- Professor F.A. Amuchie for his assistance especially during my undergraduate days and special advice on the right steps to take to attain

my present position when I newly returned from U.K. in 1996;

- Professor C.E. Chidume (now at The Abdus Salam ICTP, Trieste, Italy) for his encouragement academically and otherwise, and especially during the period of my Regular Associateship of the ICTP;
- Dr. De Gaulle Chigbu (my cousin in the U.S.) who among others, insists on making sure that all the fees for the professional associations abroad, which I belong to, are promptly paid;
- My sister in-law, Emilia (in the U.S.), for especially funding the publication of the first book I authored;
- Professors C.C. Osuagwu, M.O. Ikejiani-Clark, B.N. Okolo and M.C. Madukwe for vouching for me especially during the era of doldrums of my last assessment process;
- Professor Obi Njoku and other members of the Senate Ceremonials Committee of this University, for granting this special opportunity of giving an inaugural lecture this day and their cooperation.

I must also acknowledge the goodwill wishes from many of you to me on hearing of my promotion to the pinnacle of my profession especially those documented by Prof. (Tr.) R.U. Okafor, Mr. Gozie Arazu, Prof. & Mrs. J.N. Adichie and Prof. C.A. Nwadinigwe.

There are surely many other individuals and organizations, too numerous to mention here, that have created a lot of positive and memorable impact to my life, believing in me and being there for me during the course of the race to attain my present position at our University. I appreciate you all.

Finally, I hereby acknowledge the fact that the Senior Staff Club, University of Nigeria, Nsukka is one forum that I have benefited a lot from especially in “learning the ropes” of the socio-political environment/terrain of our great University.

Thank you for listening and God bless.

References

[1] Arua, A.I., Chigbu, P.E., Chukwu, W.I.E, Ezekwem, C.C. and Okafor, F.C. (2000), Advanced Statistics for Higher Education (Vol. 1), *The Academic Publishers, Nsukka*.

[2] Andersen, L.D. and Hilton, A.J.W. (1980), Generalized Latin rectangles I: Constructions and decomposition, *Discrete Maths., 31*, 125 – 152.

[3] Bailey, R.A. (1988), Semi-Latin squares, *Journal of Statist. Planning and Inference 18*, 299 – 312.

[4] Bailey, R.A. (1989), Designs: mappings between structured sets, *Surveys in Combinatorics, 1989 (ed. J. Siemons), London Mathematical Society Lecture Notes Series 141, Cambridge University Press, Cambridge*,22 – 51.

[5] Bailey, R.A. (1990), An efficient semi-Latin square for twelve treatments in blocks of size two, *Journal of Statist. Planning and Inference 26*, 263 – 266.

[6] Bailey, R.A. (1991), Strata for randomized experiments, *J. Roy. Statist. Soc. Ser. B*, 27 – 66.

[7] Bailey, R.A. (1992), Efficient semi-Latin squares, *Statistica Sinica 2*, 413 – 437.

[8] Bailey, R.A. (1993), Recent advances in experimental design in agriculture, *Bull. Of the International Statistical Institute, 55*, 179 – 193.

[9] Bailey, R.A. (2008), Design of Comparative Experiments, *Cambridge Series in Statistical and Probabilistic Mathematics (No. 25).*

[10] Bailey, R.A. and Chigbu, P.E. (1997), Enumeration of semi-Latin squares, *Discrete Maths., 167/168*, 73 – 84.

[11] Berge, C. (1971), Principles of Combinatorics, *Academic Press Inc., New York and London*.

[12] Chigbu, P.E. (1991), On optimal placement of the knots of a polynomial spline function under variations of optimality criteria, *Bull. Of the International Statistical Institute, 54(4B)*, 515 – 529.

[13] Chigbu, P.E. (1995), Semi-Latin squares: methods for enumeration and comparison, *Ph.D. Thesis, University of London*.

[14] Chigbu, P.E. (1998a), Block Designs: efficiency factors and optimality criteria for comparison, *Linco Press Ltd., Enugu*.

[15] Chigbu, P.E. (1998b), Efficiency, efficiency factors and optimality criteria for comparing incomplete-block designs, *Journ. Nig. Stat. Assoc., 12*, 39 – 55.

[16] Chigbu, P.E. (1999a), Optimal semi-Latin squares for sixteen treatments in blocks of size four, *Journ. Nig.Stat. Assoc., 13*, 11 – 25.

[17] Chigbu, P.E. (1999b), The block structures and isomorphism of semi-Latin squares and related designs, *Bull. Of the Institute of Combinatorics and its Appl., 12*, 53 – 65.

[18] Chigbu, P.E. (2001a), Graeco-Latin squares versus Trojan squares, *Journ. Nig. Stat. Assoc., 14*, 11 – 17.

[19] Chigbu, P.E. (2001b), A program for identifying optimal incomplete-block designs, *Res. Comm. Stat., 1(2)*, 1 –

[20] Chigbu, P.E. (2001c), On finding the non-isomorphism classes of the (n x n)/k semi-Latin

squares, *Global Journ. Pure & Appl. Scs., 7(1)*, 189 – 197.

[21] Chigbu, P.E. (2003), The "best" of the three optimal (4 x 4)/4 semi-Latin squares, *Sankhya: The Indian Journal of Statistics, 65(3)*, 641 – 648.

[22] Chigbu, P.E. (2004), A note on discriminating equally optimal semi-Latin squares for sixteen treatments in blocks of size four, *Journ. Nig. Stat. Assoc., 17*, 1 – 7.

[23] Chigbu, P.E. and Eze, B.C. (2001), Automating the group-theoretic-based construction procedure for the (n x n)/k semi-Latin square, *Utilitas Mathematica, 60*, 107 – 123.

[24] Chigbu, P.E. and Nduka, U.C. (2005), On the comparison of the polynomial classical and spline regression functions under variations of optimality criteria, *Global Journ. Math. Scs., 5(2)*, 121 – 132.

[25] Chigbu, P.E. and Oladugba, A.V. (2008), On block structures, null analyses of variance and expectations of mean squares of some quasi-semi-Latin squares, *preprint*.

[26] Chigbu, P.E. and Udoh, N.S. (2002), An optimal

transportation algorithm (OTA), Journ. Nig. Stat. Assoc., 15, 20 – 35.

[27] Chigbu, P.E. and Udoh, N.S. (2006), An application of optimal transportation algorithm (OTA) to the distribution of petroleum products in Nigeria, *Global Journ. Math. Scs., 7(1)*, 41 – 51.

[28] Chigbu, P.E. and Ugbe, T.A. (2002), On the segmentation of the response surfaces for super convergent line series' optimal solutions of constrained linear and quadratic programming problems, *Global Journ. Math. Scs., 1(1&2)*, 27 – 34.

[29] Chigbu, P.E. and Ukaegbu, E.C. (2007), On the precision and mean square error matrices approaches in obtaining the average information matrix via the super convergent line series, *Journ. Nig. Stat. Assoc., 19*, 4 – 18.

[30] Cheng, C.-S. and Bailey, R.A. (1991), Optimality of some two-associate-class partially balanaced incomplete-block designs, *Ann. Statist., 19*, 1667 –1671.

[31] Darby, L.A. and Gilbert, N. (1958), The Trojan square, *Euphytica 7*, 183 – 188.

[32] Fisher, R.A. (1926), The arrangement of field experiments, *J. Minist. Agric. (GB), 33*, 503 – 513.

[33] Fisher, R.A. (1935), Contribution to a discussion of F. Yates' paper on Complex experiments, *J. Roy. Statist. Soc. Suppl., 2*, 229 – 231.

[34] Fisher, R.A. and Yates, F. (1938), Statistical tables for biological, agricultural and medical research, *Oliver and Boyd, Edinburgh*.

[35] Fowler, H.W. and Fowler, F.G. (1990), The Concise Oxford Dictionary of Current English, *Oxford University Press*.

[36] Harshbarger, B. and Davis, L.L. (1952), Latinized rectangular lattices, *Biometrics, 8*, 73 – 84.

[37] John, J.A. and Mitchell, T.J. (1977), Optimal incomplete block designs, J. *Roy. Statist. Soc. Ser. B, 39*, 19 – 43.

[38] John, J.A. and William, E.R. (1982), Conjectures for optimal block designs, *J. Roy. Statist. Soc. Ser. B, 44*, 221 – 225.

[39] Ma, R.H. and Harrington, J.B. (1949), A study on field experiments of semi-Latin square design, *Scientific Agriculture, 29*, 241 – 251.

[40] McKay, B.D. (1990), Nauty user's guide (version 1.5), *Technical report Tr – CS – 90 – 02, Computer Science department, Australian National University, Canberra, G.P.O. Box 4, ACT 2601 Australia*.

[41] Montgomery, D.C. (1991), Design and analysis of experiments, *John Wiley & Sons*.

[42] Nduka, E.C. (2007), Statistics has it that …, *Inaugural Lecture Series No. 57, University of Port Harcourt*.

[43] Nelder, J.A. (1965), The analysis of randomized experiments with orthogonal block structure. I: Block Structure and the null analysis of variance, *Proc. Roy. Soc. Ser. A, 283*, 147 – 162.

[44] Onukogu, I.B. and Chigbu, P.E. (eds.) (2002), Super Convergent Line Series (in optimal design of experiments and mathematical programming), *AP Express Publishers, Nsukka*.

[45] Paterson, L. (1983), Circuits and efficiency in incomplete block designs, *Biometrika, 70*, 215 – 225.

[46] Preece, D.A. and Freeman, G.H. (1983), Semi-Latin squares and related designs, *J. Roy. Statist. Soc. Ser. B, 45*, 267 – 277.

[47] Rasch, D. and Herrendorfer, G. (1982), Statistische Versuchs planung, *VEB Deutscher Verlag der Wissenschaften, DDR-1080 Berlin*.

[48] Rojas, B. and White, R.F. (1957), The modified Latin square, *J. Roy. Statist. Soc. B, 19*, 305 – 317.

[49] Smith, P.L. (1979), Splines as a useful and convenient statistical tool, *The American Statistician, 33(2)*, 57 – 62.

[50] Street, A.P. and Street, D.J. (1987), Combinatorics of experimental design, *Oxford University Press*.

[51] "Student" [Gosset, W.S.] (1931), Yield trials, *reprinted in Student's collected papers (eds. E.S. Pearson and J. Wishart), London: Biometrika Office, University College, 1942*, 150 – 168.

[52] Udoh, N.S. and Chigbu, P.E. (2008), An optimal transportation algorithm (OTA) with sensitivity analysis, *Ultra Science, 20(2)M*, 377 – 384.

[53] Williams, E.R. (1986), Row and column designs with contiguous replicates, *Australian J. Statist., 28*, 154 –163.

[54] Wilson, R.J. (1990), Introduction to graph theory, *Longman Scientific and Technical, Longman Group UK Ltd*.

[55] Yates, F. (1935), Complex experiments, *J. Roy. Statist. Soc. Suppl., 2*, 181 – 247.