A POTPOURI OF RESULTS CONCERNING THE DESIGN AND ANALYSIS OF EXPERIMENTS*

by

Walter T. Federer

BU-715-M** April 1980

1. Population and Sampling Structures and Response Model Equations

Statistical literature contains an empty set when it comes to discussing populations to which experiment designs pertain. The sampling of sampling units (the elements making up a population) to form experimental units (the smallest unit to which a single treatment is applied) likewise receives no discussion. The statistician appears to assume that this is in the province of the investigator and the investigator feels that the statistician should tell him what it is all about. As a consequence, these topics are not discussed. To make a start on providing a literature to fill this void, three papers have been published. They are:

Federer, W. T. (1976). Sampling, blocking, and model considerations for the completely randomized, randomized complete block, and incomplete block designs. Biometrische Zeitschrift 18(7):511-525.

Federer, W. T. (1976). Sampling, blocking, and model considerations for r-row by c-column experiment designs. Biometrische Zeitschrift 18(8):595-607.

Federer, W. T. (1977). Sampling, blocking, and model considerations for split plot and split block designs. Biometrische Zeitschrift (Biometrical Journal) 19(3):181-200.

Sample survey design is the only segment of statistical literature which discusses population and sampling structures. There is a large intersection of survey design and experiment design theory and application, but again there are few papers in the area. A serious discussion of the intersection of survey design and experiment design theory is needed. Some papers in this area are:

---

* Outline of lecture presented at Kansas State University on April 14, 1980 at "Conference on Experimental Designs and Their Applications".

** In Mimeo Series of The Biometrics Unit, Cornell University.

Chakrabarti, M. C. (1963). On the use of incidence matrices of designs in sampling from finite populations. Journal of the Indian Statistical Association 1:78-85.

Homeyer, P. G. and C. A. Black. (1946). Sampling replicated field experiments on oats for yield determinations. Proceedings, Soil Science Society of America 11:341-344.

Raghavarao, D. and R. Singh. (1975). Applications of linked block designs in successive sampling. Applied Statistics (R. P. Gupta, Editor), pages 301-309.

Singh, R., D. Raghavarao and W. T. Federer. (1976). Applications of higher associate class PBIB designs in multidimensional cluster sampling. Estadistica 30:202-209.

The problem of selecting a response model equation does not arise in statistical literature. It is obtained by definition and is "the linear model". The one published paper which addresses the problem of model selection from a class of models is

Box, G. E. P. and D. R. Cox. (1964). An analysis of transformations. Journal of the Royal Statistical Society, B, 26:211-252.

There are a number of papers which discuss certain inadequacies of a given linear model, but are not of the nature of the above paper. Two such recent papers are:

Atkinson, A. C. and V. V. Federov. (1976). The design of experiments for discriminating between two rival models. Biometrika 62:57-70.

Jones, E. R. and T. J. Mitchell. (1978). Design criteria for detecting model inadequacy. Biometrika 65:541-551.

The problem of constructing a class of possible models for an investigation should involve the investigator and statistician jointly. Then, the statistician devises methods for selecting one or more models from the class which describe the results best in some sense. For example, given that the experiment design was a randomized complete block, that μ is an effect common to all observations, τ_i is an effect due to the i^{th} treatment, β_j is an effect due to the j^{th} block and ϵ_{ij} is an error term related to the ij^{th} observation, then a class of response model equations could be:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad \text{(classical)}$$

$$Y_{ij} = \mu_{i.}\mu_{.j}/\mu + \epsilon_{ij} = \mu + (\mu_{i.}-\mu) + (\mu_{.j}-\mu) + \frac{1}{\mu}(\mu_{i.}-\mu)(\mu_{.j}-\mu) + \epsilon_{ij}$$

$$= \mu + \tau_i + \beta_j + \frac{1}{\mu}\tau_i\beta_j + \epsilon_{ij} \qquad \text{(J. W. Tukey)}$$

$$Y_{ij} = (\mu_{i.}+\mu_{.j}-\mu)^p + \epsilon_{ij} \qquad \text{(D. S. Robson)}$$

$$Y_{ij} = (\mu+\beta_j+\epsilon_{ij})\tau_i \qquad \text{(R. C. Nair)}$$

$$Y_{ij} = (\mu+\epsilon_{ij})\beta_j\tau_i \qquad \text{(R. C. Nair)}$$

and Y_{ij} replaced by $f(Y_{ij})$, some function of the observations. The distributions of the ϵ_{ij} and β_j need to be stated to complete the formulation of the model for fixed treatment effects.

Two methods have been suggested for determining which of the above set of models fit a set of data from a randomized complete block design. The first is an extension of results in the following two papers:

Tukey, J. W. (1949). One degree of freedom for non-additivity. Biometrics 5:232-242.

Mandel, J. (1961). Non-additivity in two-way analysis of variance. Journal of the American Statistical Association 56:878-888.

In particular, individual regressions for both treatments and blocks are considered, whereas the second paper above considers regressions in only one of the categories, treatments vs blocks. The method is discussed and illustrated in two forthcoming papers to be published in the Proceedings of the Xth International Biometrics Conference held in Guarujá, Brazil, August 6-10, 1979. The papers are:

Federer, W. T. (1979). Response equations and data analysis for some simple experiment designs. (also BU-676-M in the Mimeo Series of the Biometrics Unit, Cornell University)

Kirton, H. C. (1979). Catching mice and TA's using non-additivity.

These papers use the response model equation:

$$Y_{ij} = \mu + \tau_i + \beta_j + \alpha\tau_i\beta_j + \delta_j\tau_i + \gamma_i\beta_j + \epsilon_{ij}$$

where ϵ_{ij} are NIID$(0,\sigma^2_\epsilon)$. An analysis of variance key-out of degrees of freedom for this response model equation is:

Source of variation	Degrees of freedom
Total	rv
Correction for mean	1
Blocks	r-1
Treatments	v-1
Blocks × treatments	(r-1)(v-1)
One degree of freedom for non-additivity	1
Block deviations for non-additivity	r-2
Treatment deviation for non-additivity	v-2
Remainder (error?)	(r-2)(v-2)

If any of the non-additivity components, i.e.:

One degree of freedom for non-addivity,
Block deviations for non-additivity,
Treatment deviations for non-additivity,

are relatively large (significant as some prescribed level), this is an indication of non-additivity of the residuals, of the residuals in one or more of the blocks, or of the residuals in one or more of the treatments. While J. W. Tukey's data analysis methods concentrate on outliers in the data, the above method looks for outlying blocks and/or treatments. For example, by leaving out one block (or a treatment), the classical linear model for Y_{ij} or $f(Y_{ij})$ may hold. When one applies these methods to examples in statistical methods books (e.g., in the best of all statistical methods book, i.e., Snedecor and Cochran), one finds that many of the examples contain discrepant blocks and/or treatments. The two examples described in H. C. Kirton's paper are also of this nature, and a real reason was found for the discrepancy once the investigator was told what the data indicated. A large colony of mice had moved into one of the blocks in a

chicken nutrition experiment and a technician assistant (TA) had been in charge on the days when the investigator was on leave from the investigation.

The second method involves selecting one or more particular response models and computing the residuals, $\hat{e}_{ij}$. Then, take the absolute values of the residuals and perform the same statistical analyses that were performed on the Y_{ij} or $f(Y_{ij})$. The null hypothesis for all effects should be true if the response model equation is correct. The unsolved problem with this procedure is that the distributions of quantities like

$$\sum_{i=1}^{v} \frac{\left(\sum_{j=1}^{r} |\hat{e}_{ij}|\right)^2}{r} - \frac{\left(\sum_{i=1}^{v}\sum_{j=1}^{r} |\hat{e}_{ij}|\right)^2}{rv} \tag{1.1}$$

and

$$\sum_{i=1}^{v}\sum_{j=1}^{r}\left(|\hat{e}_{ij}| - \frac{\sum_{i=1}^{v} |\hat{e}_{ij}|}{v} - \frac{\sum_{j=1}^{r} |\hat{e}_{ij}|}{r} + \frac{\sum_{i=1}^{v}\sum_{j=1}^{r} |\hat{e}_{ij}|}{rv}\right)^2 \tag{1.2}$$

are unknown. These are being investigated at the present time, but so far, little is known of their properties. We do know that $\Sigma_{i=1}^{v}\Sigma_{j=1}^{r}|\hat{e}_{ij}|^2/\sigma_\epsilon^2$ is distributed as a chi-square with $(r-1)(v-1)$ degrees of freedom when the $\hat{e}_{ij}$ are $NIID(0,\sigma_\epsilon^2)$, but we do not know how the partitioned sums of squares are distributed.

If any of the above "mean squares" are "significantly larger" than the "residuals mean square" from (1.2) this would indicate that the residuals $\hat{e}_{ij}$ contain sources of variation other than random error. Using a multiple comparisons procedure, one could determine outlying treatments or blocks. Then, one could delete the outliers and determine if the response model equation would fit the remaining blocks and treatments.

This procedure is extendible for any experiment design. For example, in a medical experiment involving four methods of treating an asthma attack, 17 patients and six visits of each patient, and in using a simple three factor main effect model, it was found that one of the patients was super-reactive and should not have been included in the experiment, four more patients were irrationally

reactive and reacted as strongly to the placebo as to the best treatment, and the other 12 patients were uniform in discriminating among the four treatments. Findings of this nature are important for medical treatment, as a different procedure of treatment would be used for each of the three groups.

2. Some Classes of Experiment Designs Useful in Experimentation

In many types of investigation, the early stages involve screening or allocating large numbers of new treatments or subjects. In some cases, only a small amount of new material is available and few replications are possible by necessity; also, it is desired to obtain some information on the new treatments during the period of producing more of the new materials. In other cases, one is unwilling to allocate more than a small number of replications to the new material because of costs involved in screening large numbers of new treatments. An example of the former is in plant breeding and production of new commercial varieties. An example of the second case is in the screening of the numerous new fungicides, herbicides, or soil fumigants. Also, it is desired to have a number of standards in the experiment with more replication than on the new treatments. In fact, in some cases it may be desirable or necessary to have the new treatments included only once or twice in the experiment and to have the standard or check treatments included r times. A class of experiment designs which accomplish this objective is augmented designs and the related reinforced and staircase designs. For augmented block designs with new treatments in once and the standards included r times, one simply sets up any block design for the standards, then one enlarges the blocks to accommodate the standards and k of the v = kr new treatments. The treatments in each block are completely randomized within the block. The statistical analysis is performed on the standard yields only to obtain solutions for block effects; these are then used to adjust the yields of the new treatments for the blocks in which they appear. Discussion of these designs may be found in the following papers:

Das, M. N. (1958). On reinforced incomplete block designs. Journal of the Indian Institute of Agricultural Statistics 10:73-77.

Federer, W. T. (1956). Augmented (or hoonuiaku) designs. Hawaiian Planters' Record 55:191-208.

Federer, W. T. (1960). Augmented designs with two-, three-, and higher-way elimination of heterogeneity. BU-329-M in the Mimeo Series of the Biometrics Unit, Cornell University.

Federer, W. T. (1961). Augmented designs with one-way elimination of heterogeneity. Biometrics 17:447-473.

Federer, W. T. (1963). Procedures and designs useful for screening material in selection and allocation, with a bibliography. Biometrics 19:553-587.

Federer, W. T., R. C. Nair and D. Raghavarao. (1975). Some augmented row-column designs. Biometrics 31:361-374.

Federer, W. T. and D. Raghavarao. (1975). On augmented designs. Biometrics 31:29-35.

Graybill, F. A. and W. E. Pruitt. (1958). The staircase design: theory. Annals of Mathematical Statistics 29:523-533.

Searle, S. R. (1965). Computing formulae for analyzing augmented random complete block designs. BU-207-M in the Mimeo Series of the Biometrics Unit, Cornell University.

Incomplete block designs have been in use by experimenters since their introduction in the thirties by Frank Yates. They are widely used in many areas of research. The number of incomplete block designs to cover situations for any number of treatments has been severely limiting. Attempts have been made to create a file or catalogue of all incomplete block designs, but this endeavor has failed because of its considerable gaps and availability to an experimenter. Since it is not feasible to construct a file of incomplete block designs for all situations, and since it is not feasible to make such a file available and usable to investigators, some simple construction procedures usable by an experimenter would be desirable. Three such procedures have recently been made available. They are described in:

Jarrett, R. G. and W. B. Hall. (1978). Generalized cyclic incomplete block designs. Biometrika 65:397-401.

Khare, M. and W. T. Federer. (1980). A simple construction procedure for resolvable incomplete block designs for any number of treatments. Biometrical Journal (accepted for publication).

Patterson, H. D. and E. R. Williams. (1978). A new class of resolvable incomplete block designs. Biometrika 63:83-92.

The first paper above requires a set of cyclic generators, and the last paper above requires a set of α-generators; but once these are available it is a simple procedure to construct designs. The second paper above does not require a set of generators, and contains algorithms for constructing incomplete block designs for any number of treatments; the incomplete block designs have maximum possible efficiency and are resolvable block designs.

A simple example is used to illustrate the procedure and the reader is directed to Khare and Federer (1980) for further details. Suppose that an experimenter wants incomplete block designs for v = 15, for v = 17, and for v = 20. The block size may be 3 or 4. The first step is to set up a balanced incomplete block for v = 25 in blocks of size k = 5 using the diagonalization method. First write square 1 in serial order; second form square 2 from square 1 by interchanging rows and columns of square 1; for square 3, take main right diagonal of square 2 as first row of square 3 and then write columns serially; do likewise for square 4 from square 3, square 5 from 4, and square 6 from 5. Doing the same thing to square 6, one obtains square 2 again as a check. To illustrate for $v^2 = 25$:

square 1

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

square 2

1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

square 3

1	7	13	19	25
2	8	14	20	21
3	9	15	16	22
4	10	11	17	23
5	6	12	18	24

square 4

1	8	15	17	24
2	9	11	18	25
3	10	12	19	21
4	6	13	20	22
5	7	14	16	23

square 5

1	9	12	20	23
2	10	13	16	24
3	6	14	17	25
4	7	15	18	21
5	8	11	19	22

square 6

1	10	14	18	22
2	6	15	19	23
3	7	11	20	24
4	8	12	16	25
5	9	13	17	21

For v = 15, 17, 20 use squares 2 to 6 and drop all numbers from 16 to 25, from 18 to 25, and from 21 to 25, respectively to obtain the three experiment designs. The block sizes will be 3 for v = 15, 4 for v = 20, and 3 and 4 for v = 17.

A third class of experiment designs is the one known as generalized block designs. The majority of statistical literature on block experiment designs pertains to the situation wherein a treatment either does not occur, $n_{ij} = 0$, or it does occur $n_{ij} = 1$. These zero-one occurrence designs appear to dominate the thinking of statisticians and investigators. Few people think of or use zero-one-two occurrence designs or even zero-one-····-n-1 occurrence designs (n-ary designs) as discussed in

Tocher, K. D. (1952). The design and analysis of block experiments (with discussion). Journal of the Royal Statistical Society B, 14:45-100.

Since n-ary designs are not contemplated, one should not expect the more generalized occurrence type designs to be considered. Designs in which the occurrences of treatments in blocks is $m_0, m_1, m_2, \cdots, m_{n-1}$ have been studied in the following papers:

Shafiq, M. and W. T. Federer. (1979). Generalized N-ary balanced block designs. Biometrika 66:115-123.

Shafiq, M. and W. T. Federer. (1980). General binary partially balanced block designs. (In the process of publication.)

Two examples are given to illustrate some properties of generalized block designs which do not hold for zero-one designs. Consider the following three experiment designs for four treatments (A,B,C,D) in blocks of size k = 4, and six replicates:

Design 1 blocks (randomized complete block design)

1	2	3	4	5	6
A	A	A	A	A	A
B	B	B	B	B	B
C	C	C	C	C	C
D	D	D	D	D	D

Design 2 blocks (balanced incomplete block design)

1	2	3	4	5	6
A	A	A	B	B	C
A	A	A	B	B	C
B	C	D	C	D	D
B	C	D	C	D	D

Design 3 blocks (partially balanced incomplete block design)

1	2	3	4	5	6
A	A	A	A	A	C
B	B	B	B	A	C
C	C	C	C	B	D
D	D	D	D	B	D

Design 1 has a variance of a difference between adjusted two treatment means of $\sigma^2/3$, design 2 of $\sigma^2/2$, and design 3 has a variance of a difference of $7\sigma^2/18$. Here we see that design 3, a partially balanced design, has a lower average variance than design 2, which is a balanced incomplete block design. The theorem stating that "among all incomplete block designs, the balanced incomplete block design is optimal" is false and holds only for $n_{ij} = 0,1$. We can see that optimality considerations need to be re-examined when one considers generalized block designs. Also, one now has a class of designs greater than one, and it is necessary to develop criteria for selecting among members of the class. A number of criteria have been developed by Shafiq and Federer (1979, 1980).

Generalized block designs are useful for:

(i) testing block x treatment interaction,

(ii) providing for minimum blocking, and

(iii) maximum utilization of nonheterogeneous material.

To illustrate, consider that an experimenter has an oven holding 8 pies or cakes, that he wishes to compare 3 baking treatments (A, B, C), and that he wishes to have 8 replications of each treatment. Three possible experiment designs are:

Design 1

Use a randomized complete block design with 8 bakings of treatments A, B, and C. This means that only 3 pies are baked at one time and energy is required for 8 bakings.

Design 2

Use a randomized complete block design with 4 bakings of treatments A, B, and C included twice in one baking. This means that 6 pies are baked at one time and energy is required for 4 bakings.

Design 3

Use a generalized balanced block design in which a treatment occurs either 2 or 3 times in each of the 3 blocks. This means that 8 pies are baked at one time and energy is required for only 3 bakings.

An analysis of variance table for the three designs is:

	Degrees of freedom (d.f.)		
Source of variation	Design 1	Design 2	Design 3
Total	24	24	24
Correction for mean	1	1	1
Blocks	7	3	2
Treatments	2	2	2
Error	14	18	19
Efficiency factor	1.000	1.000	0.984
Efficiency factor corrected for d.f.	1.000	1.025	1.014
Efficiency factor corrected for d.f.	0.975	1.000	0.989

Here we see that a generalized balanced block design is more efficient than a competing orthogonal design with 8 blocks of three and is almost as efficient, 0.989 versus 1.000, as a competing orthogonal design with 4 blocks of six. If one takes account of the cost of additional blocks, the generalized balanced block design is considerably better than either orthogonal design. Here, again, is a counter example of statistical theory which states that an orthogonal design is more efficient than a nonorthogonal design. This theory only holds for zero-

one occurrence designs.

In the area of repeated measures designs, there appears to be a considerable amount of confusion over what can be done, what should be done, and what is designed. For example, in the following book

Winer, B. J. (1962). Statistical Principles in Experimental Design. McGraw-Hill Book Company, New York.

there are algebraic errors and conceptual errors related to repeated measures designs, as well as complete omission of the concept of residual treatment effects. This book has wide usage in the social sciences and in the pharmaceutical industry. Likewise, some of the leading "authorities" on repeated measures are making misleading statements. In order to set the record straight, Ronald P. Kershner (now with Arnar-Stone, Inc., Waukeegan, Illinois) has recently completed a Ph.D. thesis wherein he provides remedies for the supposed problems of repeated design, considers efficiencies of a wide array of designs, gives a literature review of papers on the topic, considers estimability of parameters for a wide variety of response models, gives an in-depth discussion of the two-treatment p-period repeated measures designs, and gives recommendations for p-period designs for 2, 3, and 4 treatments in order to obtain solutions for the more general response model equations. A paper on the two-treatment p-period designs is being readied for publication. In addition, a paper on some unusual repeated measures designs and response models is in the process of preparation. One of the designs considered involves four diets (A, B, C, D) on 24 boys over varying numbers of periods. Two different designs were used; design 1 had maximal variance efficiency, whereas design 2 had minimal variance efficiency but maximal nutritional efficiency. The designs were:

Design 1

		Boy number											
Period	Week	1	2	3	4	5	6	7	8	9	10	11	12
1	2	A	B	C	D	A	B	C	D	A	B	C	D
2	3	B	A	D	C	D	C	B	A	C	D	A	B
	4	B	A	D	C	D	C	B	A	C	D	A	B
3	5	C	D	A	B	B	A	D	C	D	C	B	A
	6	C	D	A	B	B	A	D	C	D	C	B	A
4	7	D	C	B	A	C	D	A	B	B	A	D	C
	8	D	C	B	A	C	D	A	B	B	A	D	C
5	9	E	E	E	E	E	E	E	E	E	E	E	E
	10	E	E	E	E	E	E	E	E	E	E	E	E
	11	E	E	E	E	E	E	E	E	E	E	E	E

Treatment E was a standard diet.

Design 2

Period	Week	Boy number											
		13	14	15	16	17	18	19	20	21	22	23	24
1	2	B	C	D	A	C	D	A	B	D	A	B	C
2	3	A	A	A	B	B	B	C	C	C	D	D	D
	4	A	A	A	B	B	B	C	C	C	D	D	D
3	5	A	A	A	B	B	B	C	C	C	D	D	D
	6	A	A	A	B	B	B	C	C	C	D	D	D
4	7	A	A	A	B	B	B	C	C	C	D	D	D
	8	A	A	A	B	B	B	C	C	C	D	D	D
5	9	A	A	A	B	B	B	C	C	C	D	D	D
	10	A	A	A	B	B	B	C	C	C	D	D	D
	11	A	A	A	B	B	B	C	C	C	D	D	D

Many characteristics were measured and many variations in the conduct of the experiments were made. As a consequence, many statistical analyses are required. For example, at the end of six weeks a change was made in the amount of food the boys were allowed to eat. This was dictated by the amount of weight the boys were losing. The change in amount of food consumed could cause a boy by period interaction for certain characteristics. Also, diet C contained a wood cellulose additive (approved by Food and Drug) which killed the bacteria in the intestine and hence affected digestibility for a considerable period after the application of treatment C. Thus, treatment C has a continuing effect in all periods following the period of application. There appeared to be no residual effects for the remaining treatments which involved course bran, fine bran, and cabbage extract additives to bread. This experiment is one of the studies investigating causes of cancer of the colon in humans.

Some Cornell papers on crossover designs with residual effects and various types of interactions are listed below:

Federer, W. T. and R. P. Kershner. (1979). On the design and analysis of repeated measures experiments. BU-681-M in the Mimeo Series of the Biometrics Unit, Cornell University (now being readied for publication).

Kershner, R. P. (1980). On the theory of crossover designs with residual effects. Ph.D. Thesis, Cornell University, May.

Kershner, R. P. and W. T. Federer. (1980). Two-treatment crossover designs for estimating a variety of effects. BU-675-M (revised) in the Mimeo Series of the Biometrics Unit, Cornell University (being readied for publication in the Journal of the American Statistical Association).

Another area of research on experiment design which has received our attention is that of orthogonality in latin square and F-square designs. An F-square in which each treatment occurs once in a row and once in a column of an n × n square is a latin square. When the i^{th} treatment occurs λ_i times in each row and in each column of an n × n square for $\lambda_i \geq 1$, for $\Sigma_{i=1}^{t}\lambda_i = n$, and for $t \leq n$ treatments, we denote this as an F-square. Thus, we see that F-squares are generalizations of latin squares and that a latin square is a special case of an F-square. In practice, one would use an F-square for the same type of investigation as one would use a latin square, but for which there are $t \leq n$ treatments and $n\lambda_i$ replications for treatment i. For example, one could use 2 or 3 treatments in an F-square of order n = 4 as follows:

2 treatments

	Column			
Row	1	2	3	4
1	A	A	B	B
2	A	A	B	B
3	B	B	A	A
4	B	B	A	A

$\lambda_A = \lambda_B = 2$

$r_A = r_B = 8$

3 treatments

	Column			
Row	1	2	3	4
1	A	B	C	A
2	A	A	B	C
3	C	A	A	B
4	B	C	A	A

$\lambda_A = 2,\ \lambda_B = 1 = \lambda_C$

$r_A = 8,\ r_B = 4 = r_C$

F-squares are orthogonal designs and hence have a simple statistical analysis making use of arithmetic means.

The problem of constructing two or more latin squares which are orthogonal to each other, i.e., each treatment of one square appears once with all treatments of the second square, has received the attention of mathematicians and

statisticians since the 18th century. However, there are many unsolved and unresolved problems in this area; it appears that these problems will remain with us for a long time. To illustrate orthogonal latin squares for n = 3 and 4 treatments, the following squares and hypothetical experiments are presented:

	Store		
Week	1	2	3
1	A α	B β	C γ
2	C β	A γ	B α
3	B γ	C α	A β

A, B, C are treatments on packaging apples

α, β, γ are treatments on packaging carrots

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

A, B, C, D are treatments on brands of table napkins

α, β, γ, δ are treatments on brands of toilet tissue

a, b, c, d are treatments on brands of paper towels

Despite the fact that the theory of constructing sets of orthogonal latin squares has been around for about 200 years, statistical analyses for sets of orthogonal latin squares has not been available in published form until recently. However, it is believed that what is available is incomplete and/or inappropriate. For example, for two sets of orthogonal latin square experiments on, say, paper napkins and paper towels, an analysis of variance for sales of the two products could be:

Paper napkin sales:

Source of variation	Degrees of freedom	
Total	16	
Correction for mean	1	
Stratification variables	9	
Stores		3
Weeks		3
Paper towel treatments		3
Paper napkin treatments	3	
Remainder	3	

Paper towel sales:

Source of variation	Degrees of freedom	
Total	16	
Correction for mean	1	
Stratification variables	9	
Stores		3
Weeks		3
Paper napkin treatments		3
Paper towel treatments	3	
Remainder	3	

Perhaps a more appropriate statistical analysis would be a multivariate approach, in this case, a bivariate analysis.

Some papers on construction of sets of orthogonal latin square, F-square, latin cube, F-cube and latin and F higher dimensions are listed below:

Federer, W. T. (1977). On the existence and construction of a complete set of orthogonal F(4t;2t,2t)-squares design. Annals of Statistics 5:561-564.

Federer, W. T., et al. (1971). Some techniques for constructing mutually orthogonal latin squares. Proceedings of the 15th Conference, Design of Experiments in Army Research Development and Testing, ARO-D Report 70-2, pp. 673-796.

Finney, D. J. (1945). Some orthogonal properties of the 4 x 4 and 6 x 6 latin squares. Annals of Eugenics 12:213-219.

Finney, D. J. (1946). Orthogonal partitions of the 5 x 5 latin squares. Annals of Eugenics 13:1-3.

Finney, D. J. (1947). Orthogonal partitions of the 6 x 6 latin squares. Annals of Eugenics 13:184-196.

Hedayat, A. (1969). On the theory of the existence, non-existence, and the construction of mutually orthogonal F-squares and latin squares. Ph.D. Thesis, Cornell University, June.

Hedayat, A., D. Raghavarao, and E. Seiden. (1975). Further contributions to the theory of F-squares design. Annals of Statistics 3:712-716.

Hedayat, A. and E. Seiden. (1970). F-square and orthogonal F-square design: A generalization of latin squares and orthogonal latin squares design. Annals of Mathematical Statistics 41:2035-2044.

Mandeli, J. P. (1975). Complete sets of orthogonal F-squares. M.S. Thesis, Cornell University, August.

Mandeli, J. P. (1978). Contributions to the theory of F-square and F-cube designs. Ph.D. Thesis, Cornell University, August.

Mandeli, J. P. and W. T. Federer. (1979). Complete sets of orthogonal F-squares of prime power order with differing numbers of symbols. BU-679-M in the Mimeo Series of the Biometrics Unit, Cornell University (in the process of publication).

Mandeli, J. P., F.-C. H. Lee and W. T. Federer. (1977). On the existence and construction of orthogonal F-squares of order $n = 2s^p$, s a prime number. BU-622-M in the Mimeo Series of the Biometrics Unit, Cornell University (in the process of publication).

Also, Fen-Chen Helen Lee is writing a Ph.D. dissertation in this area. She has a number of new results which have not yet been written up.

3. Treatment Designs

Considerable research on fractional replication of factorial treatment designs has been carried on over the last 25 years at Cornell University, as well as at several other places. A book is being published on the theory of fractional replication. It is rather mathematical in nature, and hence will not be too useful for researchers using fractional replicates. Much of this research is joint with B. Leo Raktoe, University of Guelph, Donald A. Anderson, University of Wyoming, and others. Since the list of publications is lengthy in this area, the reader is referred to Annual Reports of the Biometrics Unit for a listing of technical reports, theses, and published papers. One area of research being pursued at the present time is the development of a class of fractional replicates to achieve minimum number of runs or combinations and to allow estimation of certain two-factor interactions as well as the main effects. The use of orthogonal arrays to form fractions results in too many combinations for the

investigator. For example, suppose that one has four factors, one at 4 levels, and three factors each at 2 levels, and that one wishes to obtain information on a two-factor interaction for 2 of the factors at two levels. The use of an orthogonal array would require 16 combinations or runs, whereas use of the following nonorthogonal array requires only 8 combinations.

Factor 1 at 4 levels	1	2	3	4	1	2	3	4
Factors 2 and 3 each at 2 levels	00	01	10	11	11	00	01	10
Factor at 2 leve	1	1	1	1	2	2	2	2

Thus a combination is given by the levels in columns, the first combination being 1001, the second being 2011, etc. These 8 combinations form the fractional replicate for this situation.

A large class of these nonorthogonal arrays (fractional replicates) can be formed using experiment design theory on balanced row-column designs, and orthogonal latin squares and F-squares theory. The results in the following paper on the minimal and maximal values of the determinant of the X'X matrix can take are useful in assessing how good these fractions are:

Anderson, D. A. and W. T. Federer. (1974). Representation and construction of main effect plans in terms of (0,1)-matrices. BU-499-M in the Mimeo Series of the Biometrics Unit, Cornell University. (In the process of publication.)

Another area of treatment design receiving our current attention relates to mixtures of levels of factors wherein the ratio of the factors in a mixture is fixed by the investigator, by the nature of the process, or by the nature of the material. In a genetic cross or a tournament each parent or player enters in a 1:1 ratio. In mixtures of crops, the farmer may want $\frac{1}{2}$ maize to $\frac{1}{4}$ soybeans to $\frac{1}{4}$ cowpeas for a 2:1:1 ratio. In a chemical compound involving radicals on a carbon chain, the ratio of radicals is fixed. In taking courses in high school, each course is one hour in duration and a student takes four courses, giving a fixed ratio of 1:1:1:1. Thus, we can see that there is a wide variety of situations where the ratio of quantities in a mixture is fixed. The large literature on response surface designs is not useful, since the primary goal here is to

estimate the ratio, whereas in fixed-ratio mixture designs the ratio is given. This means that response model equations, statistical designs and analyses, and statistical inferences are required for this type of treatment design. Some progress has been made in this area as indicated in the following selected references:

Federer, W. T. (1979). Statistical designs and response models for mixtures of cultivars. Agronomy Journal 71:701-706.

Federer, W. T. and A. Wijesinha. (1979). Statistical definitions, designs, response equations and analyses for experiments on fixed-ratio mixtures in agriculture. BU-677-M (revised 3/80) in the Mimeo Series of the Biometrics Unit, Cornell University. (In the process of publication.)

Free, Jr., S. M. and J. W. Wilson. (1964). A mathematical contribution to structure-activity studies. Journal of Medical Chemistry 7:395-399.

Raghavarao, D. and W. T. Federer. (1979). Block total response as an alternative to the randomized response method in surveys. Journal of the Royal Statistical Society, B, 41:40-45.

4. Statistical Analyses

The subject of covariance analysis is rather poorly understood by both statisticians and investigators using a covariance analysis. Some references on covariance are:

Federer, W. T. (1955). Experiment Design - Theory and Application, Macmillan, New York (republished by the Oxford and IBH Publishing Company, Calcutta, India, in 1964 and reprinted in 1974 and 1979). Chapter XVI.

Federer, W. T. (editor). (1979). Special Issue on Covariance, Communications In Statistics A8(8).

Federer, W. T. and H. V. Henderson. (1979). Covariance analysis of designed experiments x statistical packages: An update. Proceedings, Annual Symposium on the Interface of Computer Science and Statistics, Waterloo, Ontario.

As most statisticians know, there is a literature on the problem of comparing means of two random samples when the two populations have unequal

variances. This is known as the Behrens-Fisher problem. The literature did not discuss the problem for more than two means; consequently, this resulted in the following papers:

Grimes, B. (1979). Cochran-like and Welch-like approximate solutions to the problem of comparison of means from two or more populations with unequal variances. M.S. Thesis, Cornell University, August.

Grimes, B. and W. T. Federer. (1980). Cochran-like and Welch-like approximate solutions to the problem of comparison of means from two or more populations with unequal variances. Social Statistics Section Proceedings, American Statistical Association.

Several other types of analyses have been developed and are described in the Annual Reports of the Biometrics Unit.