RESPONSE EQUATIONS AND DATA ANALYSIS FOR SOME SIMPLE EXPERIMENT DESIGNS

Walter T. Federer

BU-676-M*

June 1979

SUMMARY

For any given experiment design, many response model equations are possible despite the universal textbook practice of writing, "the model for this experiment is . . .". A number of experimentally plausible models are suggested for consideration when the experiment design is a completely randomized, a randomized complete block, or a latin square. These forms of a response model equation may hold for any experiment design. Data analytic techniques are used to select a model from this class which best describes the data from the experiment. A post mortem analysis should be conducted on the assumed experiment design, on the selected response equation and model assumptions, on the data obtained, and on hypotheses tested. Nunerical examples are presented to illustrate the procedure.

1. INTRODUCTION

Statistical literature abounds with such phrases as "we shall assume . . ." and "let us define . . ." with little or no effort being expended on checking the adequacy, the assumptions or definitions, or of considering alternatives. Within the confines of the chalkboard world of statisticians, we talk freely about optimality, independence, BLUE estimation, etc., etc. One should ask how grey, or even murky brown, are the BLUE's in a real world situation or, if there are several types of optimality some of which are nonstatistical. Many of our hallowed statistical terms are in difficulty in a real world situation; we should have much more literature along the lines of Tukey's (1975) paper where he states seven assumptions required for the Gauss theorem and discusses what happens when the assumptions are removed. Statistical texts use the phrase

---

* In the Mimeo Series of the Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, New York, 14853.

"the linear model is . . .". At best, they should say "a linear model is . . .", or even better, this linear model is regarded as an approximation of the true unknown model". Then, we frequently hear the term "GLM theory" (general linear model theory) and often in awed tones by the speaker. One should ask what really is "general" about GLM used in practice. Does it, for example, consider the simultaneous minimization of a sum of weighted sums of squares? What about variance heteroscedasticity? What about various degrees of correlated residuals in the various sums of squares? Are the error structures obtained by definition, e.g., $\underline{\epsilon} = \underline{Y} - E(\underline{Y})$, where $\underline{\epsilon}$ is an error vector and $\underline{Y}$ is an observation vector? Are the populations, sampling and experimental procedures, and response model equations obtained by definition, which are then recommended for use in a real world situation? Does any statistical text discuss the population, the sampling procedure, alternative response model equations, measurement errors, etc., for the common experiment designs? How can one make inferences if the population about which the inferences are made is not precisely defined and described?

In selecting a response model equation for data from an investigation the following questions require consideration:

(i) What are the parameters of the model and are they explicitly and precisely defined in the context of the investigation?

(ii) How adequate is the selected model?

(iii) What is the sampling and experimental procedure and error structure for the investigation?

(iv) Is the main object to estimate, or at least obtain unique solutions for, the parameters of the model?

(v) Are there simple and efficient statistical estimators of the parameters?

(vi) What conditions are necessary for the selected estimators and how are they affected when the conditions are not met?

If it is not possible to obtain satisfactory answers to all of the above questions prior to the start of this investigation, one should at least use available techniques to obtain some answers from the data. One such available technique is "data analysis" as put forward by J. W. Tukey and co-workers over the past few years. There are three basic questions for a serious data analyst to consider. These are:

(i) Are the data contaminated with discrepant (rogue, far-out, ···) observations and what action should be taken when data are contaminated?

(ii) How to check the adequacy of the response model used to analyze the data?

(iii) How to check for a discrepant block, treatment, etc. when (ii) is satisfied for the remainder of the data?

There is a considerable literature on (i) above, and some literature (e.g., Box and Cox (1964), Atkinson and Fedorov (1976), Jones and Mitchell (1978), and some of the references contained in the preceding) on checking adequacy of a response model. There appears to be relatively little on (iii); one technique was suggested to the author by H. C. Kirton, N.S.W. Department of Agriculture, Sydney, Australia in Spring 1977. Some properties of the technique are unknown, but its application is straightforward and considered useful; it is applied to two examples in section four.

In the next section, we put forth a procedure which is to be applied to the planning and analysis of any investigation. Then, we demonstrate the procedure for data from a completely randomized design in section three, a randomized complete block design in section four, and a latin square design in section five.

2. A PROPOSED PROCEDURE FOR ALL INVESTIGATIONS

Whether the investigation be an observational one, a survey, or an experiment, the following steps are helpful (and necessary) in making interpretations about the results from the investigation and for more efficient planning of future investigations:

(i) The variables measured and the method of measuring them must be clearly understood and precisely stated.

(ii) The population from which observations are made must be precisely described and understood.

(iii) The sampling procedure used to obtain an observational unit (the smallest unit on which an observation is made), a sampling unit (an element of the population), and an experimental unit (the smallest unit of material to which one treatment is applied) must be understood fully if correct statistical analyses are to be made. The number of replications for a given treatment will be the number of experimental units, and not necessarily the number of observational units or sampling units, for a given treatment.

(iv) Additional sources of variation can affect the results of an investigation such as measurement errors and environmental effects during the course of an investigation. These need to be clearly and precisely described.

(v) In the event it is not possible to select "the model", a class of plausible models needs to be delineated.

(vi) From the selected class of response models, one needs to find a model(s) which adequately describes the data using statistical criteria. Then, one needs to consider whether the model selected provides an adequate description of the data for the particular field of investigation. Analysis of data as the thirteenth root of the observations may be incomprehensible in the particular field of investigation. Analyzing data on a scale satisfying statistical criteria can considerably alter the ranking of treatment means over that obtained using the original observations.

(vii) A post mortem should be conducted on an investigation in order to check on the inferences made and to aid in the planning and analysis of future investigations.

3. COMPLETELY RANDOMIZED DESIGN

The population structure of a completely randomized design should be a single population when the v treatments are all identical (see Federer (1976a)). This means that the sampling units used to form the experimental units form a simple random sample from the population. (Note that if one knows the correct response model, it is unimportant how the sample from the population is obtained. However, if the correct model is unknown, then it is essential that a simple random sample be obtained from the population.) In addition, the measurement errors, investigational errors, and/or environmental errors that occur during the course of the investigation must constitute a single population so that there is one population consisting of a mixture of two populations, i.e., the original population plus the population of other nontreatment sources of variation which occurred during the course of the experiment. This point was omitted by Federer (1976a, 1976b) in discussing the population structure of experiment designs.

Although the population may be unstratified for one variable, it could be stratified for a second variable. Thus, for each variable to be measured, it is necessary to verify that the original population is an unstratified (unblocked) one. Also, the sampling units obtained to form the experimental units may have come from a single population, but the method of conducting the investigation may be such as to stratify or block the observations. Hence, it is always useful and instructive to think first of an analysis of an experiment for the case when all treatments are identical; then, if the response for the ij^{th} experimental

unit can be expressed as Y_{ij}, $i=1,2,\cdots,v$, $j=1,2,\cdots,r_i$, equals a mean, μ, plus an error deviation ϵ_{ij}, the conditions for a completely randomized design are satisfied. Secondly, one should consider how the addition of the i^{th} treatment effect, τ_i, influences the value of $Y_{ij} = \mu + \epsilon_{ij}$. In the absence of knowledge concerning the correct response model equation, one should consider a class of plausible models. One class of response model equations suggested experimental unit yields from a completely randomized design is:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij} \quad \text{(classical additive model)}, \tag{3.1}$$

$$f(Y_{ij}) = \mu + \tau_i + \epsilon_{ij} \quad \text{(sometimes used)}, \tag{3.2}$$

$$Y_{ij} = \mu\tau_i\epsilon_{ij} \quad \text{(multiplicative model)}, \tag{3.3}$$

$$Y_{ij} = \mu\tau_i + \epsilon_{ij}, \tag{3.4}$$

$$Y_{ij} = (\mu+\tau_i)^p + \epsilon_{ij}, \quad \text{and} \tag{3.5}$$

$$Y_{ij} = (\mu+\epsilon_{ij})\tau_i . \tag{3.6}$$

The above response equations can be considered for both equal and unequal treatment error variances, for both uncorrelated and correlated observations, for known or unknown distributions, and for discrete or continuous variables. For positive μ, τ_i, and ϵ_{ij}, equation (3.3) can be converted into (3.2). Also, this means that the error deviations $\epsilon_{ij} \geq 0$ and that $E(\epsilon_{ij}) = 1$. Equations (3.3), (3.4), and (3.6) suggest using ratios of means, e.g., "in percent of check or standard treatment", rather than differences between pairs of treatment means.

D. S. Robson, Cornell University, has done considerable work, most of which is unpublished, on equation (3.5), and Nair (1976) has considered equations of the

form of (3.6) for a number of experiment designs. The latter author has considered various error structures to some extent. For example, the error term could enter equation (3.5) in various ways such as

$$Y_{ij} = (\mu+\tau_i)^p\epsilon_{ij} \tag{3.7}$$

and

$$Y_{ij} = (\mu+\tau_i+\epsilon_{ij})^p \quad , \tag{3.8}$$

as well as the form in (3.5).

Selection of response equations (3.1) or (3.2) offer little or no statistical difficulties under variance homoscedasticity but may otherwise such as, for example, in making linear contrasts among v treatments under a "Behrens-Fisher" situation (see Grimes (1979)). If the elements of (3.3) are all positive a logarithmic transformation converts it into (3.2), but are there situations where this is not true, and which would allow negative Y_{ij}? Equation (3.4) is the Tukey one-degree-of-freedom for nonaddivity analogue for a completely randomized design (see Federer (1955), pages 50-1). The above equations (3.1) to (3.8), or some subset thereof, with definitions of parameters and distributional descriptions of the elements of an equation form a class of response models for consideration by an experimenter.

Two examples will be considered. These are examples IV.1 and IV.2 in Federer (1955). The first example appears to be a standard textbook example of a completely randomized design at first site in that 20 chicks were randomly divided into four lots of five each and a different "tropical feedstuff" was given to each of the lots. There was no description available for the conduct of the experiment, but it appears highly likely that each lot was in a different pen and location. Hence, the design is really one replicate of a

randomized complete block design with five observational units per experimental unit. This situation is the most frequent mistake occurring in practice that the author has encountered. The second most frequent incorrect analysis is for one replicate of a split plot design. A second problem encountered in this example is that there is one low yield in each of the four treatments. Did the experimenter distribute the chicks randomly to the four treatments, or did he "balance" the chicks between the treatments?

In example IV-2 of Federer (1955), a number of questions arise, some of which are:

(i) Since approximately 10 (variance) = mean for the treatments, does this indicate a square root transformation to stabilize variances?

(ii) Why should the square root of dry shrub weight be used instead of log of dry shrub weight?

(iii) What should one do about the unusually low observation in the "off-type" group and the unusually high one in the "aberrant" group?

(iv) Since the sample sizes are random variables, how should one proceed?

(v) If one does not transform the data, how should one proceed in making contrasts among the $v = 3$ treatments?

4. RANDOMIZED COMPLETE BLOCK DESIGN

The population structure for one-way blocked designs (Federer (1976a)), which include the randomized complete and incomplete block experiment designs, consists of a population made up of subpopulations, the blocks. To obtain the sampling units for an experiment, one selects a random sample (random sample not necessary if true model is known) of subpopulations (blocks) and then a simple random sample within each subpopulation. These sampling units are used to form the experimental units for an experiment. The original variation among

subpopulations and sampling units within subpopulations plus the variation during the conduct of the experiment represents the total variation. Thus, in the absence of treatment effects, a response, or some function of the response, is expressible as the j^{th} subpopulation mean $\mu_{.j}$ plus an error deviation ϵ_{ij}. That is, Y_{ij} = i^{th} observation from the j^{th} subpopulation $= \mu_{.j} + \epsilon_{ij}$, $i=1,2,\cdots,v$ and $j=1,2,\cdots,r$ in the case of a randomized complete block design. Given that $\mu_{.j}$ is made up of two components, the overall population mean μ and a subpopulation (block) effect, say β_j, the three components μ, β_j, and ϵ_{ij} could have one of the relationships described by equations (3.1) to (3.8) prior to the addition of treatment effects. Some randomized complete block design response model equations for consideration in the presence of treatment effects are:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \text{(classical additive model)}, \tag{4.1}$$

$$f(Y_{ij}) = \mu + \tau_i + \beta_j + \epsilon_{ij} \quad \text{(sometimes used)}, \tag{4.2}$$

$$Y_{ij} = \mu\tau_i\beta_j\epsilon_{ij} \quad \text{(multiplicative model)}, \tag{4.3}$$

$$\begin{aligned} Y_{ij} &= \mu\tau_i\beta_j + \epsilon_{ij} = \mu_{i.}\mu_{.j}/\mu + \epsilon_{ij} \\ &= \mu + \tau_i + \beta_j + \tau_i\beta_j/\mu + \epsilon_{ij}, \end{aligned} \tag{4.4}$$

where $(\mu_{i.}-\mu) = \tau_i$ and $(\mu_{.j}-\mu) = \beta_j$, $\mu_{i.}$ = i^{th} treatment mean (Tukey non-additive response model, where Mandel (1961), e.g., writes $\tau_i\beta_j/\mu$ as π_{ij}),

$$Y_{ij} = (\mu+\tau_i+\beta_j)^p + \epsilon_{ij}, \tag{4.5}$$

$$Y_{ij} = (\mu+\beta_j+\epsilon_{ij})\tau_i, \tag{4.6}$$

and

$$Y_{ij} = (\mu+\epsilon_{ij})\beta_j\tau_i. \tag{4.7}$$

The above response model equations can be considered for both equal and unequal variances, for both correlated and uncorrelated observations, for known or unknown distributions, and for discrete or continuous variables. For positive μ, τ_i, β_j, and ϵ_{ij}, equation (4.3) can be converted to (4.2). Again, equations (4.3), (4.4), (4.6), and (4.7) imply that interest is in ratio of means rather than differences. Equation (4.5) has been studied by D. S. Robson, Cornell University, unpublished results. Nair (1976) has obtained maximum likelihood solutions for the parameters in (4.6), given that the ϵ_{ij} are $N(\mu,\sigma^2_\epsilon)$; he has also considered equations of the form of (4.7). Although some response equations such as (4.1) and (4.2) under variance homoscedasticity provide a straightforward statistical analysis, solutions for the parameters in equations (4.6) and (4.7) are obtained by iterative methods. Federer (1951) considered a more complicated form of equation (4.6), and using this method solutions for $\log \tau_i$ and $\log(\mu+\beta_j+\epsilon_{ij})$ would be obtained; then from the $Y_{ij}/\hat{\tau}_i$ values, solutions would be obtained for μ, β_j and σ^2_ϵ. Using the same procedure one could obtain solutions for $\log \tau_i$, $\log \beta_j$, and $\log(\mu+\epsilon_{ij})$ in (4.7), given that τ_i, β_j, and $\mu+\epsilon_{ij}$ are positive values, and then from the $Y_{ij}/\hat{\tau}_i\hat{\beta}_j$ values, solutions for μ and σ^2_ϵ may be obtained.

The following two examples indicate ways in which apparently standard examples are fraught with difficulties. Mandel (1961) gave a procedure to check for various types of nonadditivity. H. C. Kirton, Department of Agriculture, New South Wales, Sydney, Australia, uses an extended version of Mandel's (1961) procedure. This is given in Table 4.1; then, after an apparent appropriate response model equation has been selected, the same analysis of variance is performed on the absolute values of the residuals. For this case, we do not know the properties of the F-tests performed, but the null hypothesis should be true if one has the correct response model.

A response model equation for Table 4.1 is

$$Y_{ij} = \mu + \tau_i + \beta_j + \tau_i\beta_j/\mu + (\rho_j - 1)\tau_i + (\rho_i - 1)\beta_j + \epsilon_{ij} \quad , \tag{4.8}$$

where ρ_i is the linear regression coefficient for treatment i, ρ_j is the linear regression coefficient for block j, b_i in Table 4.1 is the least squares solution for ρ_i, b_j is the least squares solution for ρ_j, and $\Sigma_1^r \rho_j/r = \Sigma_1^v \rho_i/v = 1$. Also, $\Sigma_1^v b_i/v = \Sigma_1^r b_j/r = 1$. Tukey's one degree of freedom for nonadditivity, TNA, is an over-all check on nonadditivity. The blocks (deviations) sum of squares can be used to test for uniformity of nonadditivity over all blocks. One or more discrepant blocks could occur. Likewise, the treatments (deviations) sum of squares can be used to test for uniformity of treatment slopes with respect to this type of nonadditivity. Thus, while the majority of attention in data analysis has been on discrepant observations, the above provides a method of detecting discrepant levels of one or more factors. This procedure can be helpful in grouping levels of a factor for analysis. For example, for patients as blocks, it may be necessary to stratify the patients and to run separate analyses for each group.

Example 4.1. Data from a randomized complete block design with five treatments composed of a low, 20 ppm, or a high, 100 ppm, application of nitrogen at three different dates. The treatment was applied to soybean plants in pots containing a sand culture in a greenhouse (Federer (1955), example V.1, pages 117-9). At first sight, this example appears to be of the standard textbook variety using response equation (4.1). However, if one computes a Spearman's rank order correlation on $\Sigma_{j=1}^r \hat{e}_{ij}^2$ and $\bar{y}_{i\cdot}$, the correlation coefficient in one (D. S. Robson, Cornell University, proved that this is a Spearman's rank order correlation coefficient.). Likewise, the correlation between $\Sigma_{i=1}^r |\hat{e}_{ij}|/r$ and $\bar{y}_{i\cdot}$ is also unity.

These relationships do not hold for block means. Since there appears to be a type of nonadditivity, one could consider the analysis given in Table 4.1. Doing this, we note that TNA is near zero; block (deviations) and treatment (deviations) have smaller mean squares than the remainder term. Also, p was computed as p = .9, which indicates use of $Y_{ij}^{p} = Y_{ij}^{.9}$ or approximately the untransformed Y_{ij} values in an analysis of variance. In light of the relationship, between treatment means and "variances", it appears that although we might have additivity on the Y_{ij} scale, we do not have variance homoscedasticity. Thus, we would need to break out single degree of freedom contrasts and obtain an error variance with 5 degrees of freedom for each contrast. Or, we perhaps should use another response model, e.g., (4.6). Note that (4.7) appears inappropriate as the relation between block means and variances does not appear to be present. In order to use (4.6) and obtain maximum likelihood solutions under normality of error deviations, one would need to obtain solutions minimizing (Nair (1976)):

$$\sum_{i=1}^{v} \sum_{j=1}^{r} (Y_{ij}/\hat{\tau}_{ij} - \hat{\mu} - \hat{\beta}_j)^2 \quad .$$

These would need to be obtained iteratively. As a first solution, one could:

(i) transform Y_{ij} to ln Y_{ij},

(ii) obtain mean ln Y_{ij} for treatments, which are 14.43, 22.16, 26.29, 34.54, and 49.01, respectively, and

(iii) since $\pi_{i=1}^{v} \tau_i^{+} = 1$, one needs to divide each of the means of ln Y_{ij} by the geometric mean, 26.9554, of these means to obtain $\tau_1^{+} = 0.535$, $\tau_2^{+} = 0.822$, $\tau_3^{+} = .975$, $\tau_4^{+} = 1.281$, and $\tau_5^{+} = 1.818$,

(iv) obtain $Y_{ij}^{+} = Y_{ij}/\tau_i^{+}$ values (Table 4.2),

(v) note relationship of block means and variances for Y_{ij}^{+} values (slight if any in Table 4.2),

Table 4.1. Analysis of variance to check for additivity and uniformity of levels of factors for a randomized complete block design.

Source of variation	Degrees of freedom	Sum of squares*
Total	rv	$\Sigma_1^v \Sigma_1^r Y_{ij}^2$
Correction for mean	1	$Y_{..}^2/rv = C$
Blocks	r-1	$\Sigma_1^r Y_{.j}^2/v - C$
Treatments	v-1	$\Sigma_1^v Y_{i.}/r - C$
Blocks × treatments	(r-1)(v-1)	$\Sigma_1^v \Sigma_1^r (Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$
Nonadditivity	1	$TNA = \frac{[\Sigma_1^v \Sigma_1^r Y_{ij}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})]^2}{\Sigma_1^v (\bar{y}_{i.} - \bar{y}_{..})^2 \Sigma_1^r (\bar{y}_{.j} - \bar{y}_{..})^2}$
Blocks (deviations)	r-2	$\Sigma_1^r (b_j - 1)^2 \Sigma_1^v (\bar{y}_{i.} - \bar{y}_{..})^2 - TNA$
Treatments (deviations)	v-2	$\Sigma_1^v (b_i - 1)^2 \Sigma_1^r (\bar{y}_{.j} - \bar{y}_{..})^2 - TNA$
Remainder	(r-2)(v-2)	by subtraction

*$Y_{i.}$ and $\bar{y}_{i.}$ = i^{th} treatment total and mean, respectively,

$Y_{.j}$ and $\bar{y}_{.j}$ = j^{th} block total and mean, respectively,

$Y_{..}$ and $\bar{y}_{..}$ = grand total and mean, respectively,

$b_j = \Sigma_{i=1}^v Y_{ij}(\bar{y}_{i.} - \bar{y}_{..})/\Sigma_1^v (\bar{y}_{i.} - \bar{y}_{..})^2$, and

$b_i = \Sigma_{j=1}^r Y_{ij}(\bar{y}_{.j} - \bar{y}_{..})/\Sigma_1^r (\bar{y}_{.j} - \bar{y}_{..})^2$.

Table 4.2. Five treatments consisting of levels and times of nitrogen applications to soybean plants arranged in a randomized complete block design with r = 6 replicates (Example V.1, Federer (1955)).

Blocks	Treatments (gms/pot) 1	2	3	4	5	$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$	b_j	$b_j - 1 - \alpha\hat{\beta}_j$
1	8.8	23.5	41.2	28.4	67.4	3.030	1.48696	0.47454
2	12.9	26.3	22.5	48.4	33.2	-2.170	0.65763	-0.33348
3	11.7	21.6	21.8	16.4	59.5	-4.630	1.20208	0.22106
4	31.2	15.6	46.3	44.5	49.8	6.650	0.74906	-0.27820
5	22.0	24.4	15.6	38.8	57.1	0.750	1.09607	0.09300
6	9.9	23.3	22.6	43.6	36.6	-3.630	0.80818	-0.17694
$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$	-14.747	-8.380	-2.496	5.853	19.770	0	6	0
b_i	1.42079	-0.51896	2.27931	0.77018	1.04868	5	$\bar{y}_{..} = 30.8300$	
$b_i - 1 - \alpha\hat{\tau}_i$	0.48124	-1.48461	1.28954	-0.25381	-0.03236	0		

$\alpha = \Sigma_1^r \hat{\beta}_j b_j / \Sigma_1^r \hat{\beta}_j^2 = \Sigma_1^v \hat{\beta}_i b_i / \Sigma_1^v \hat{\beta}_i^2 = 0.382409/93.2886 = 2.947604/719.0389 = .004099;$

$1 - \alpha\bar{y}_{..} = p = 0.87.$

Analysis of variance and F-ratios

Source of variation	d.f.	Mean square	F-ratio
Blocks	5	93.29	0.78
Treatments	4	1,078.55	9.06
Blocks x treatments	20	118.986	-
TNA	1	1.131	0.01
Blocks (deviations)	4	90.631	0.67
Treatments (deviations)	3	129.862	0.96
Remainder	12	135.539	-

Table 4.2. (Cont'd)

Residuals $\hat{e}_{ij} = Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$

Blocks	Treatments					$\lvert\hat{e}_{ij}\rvert$	
	1	2	3	4	5	Total	Mean
1	-10.31	- 1.98	9.84	-11.31	13.77	47.21	9.44
2	- 1.01	6.02	- 3.66	13.88	-15.23	39.80	7.96
3	0.25	3.78	- 1.90	-15.65	13.53	35.11	7.02
4	8.47	-13.50	11.32	1.17	- 7.45	41.91	8.38
5	5.17	1.20	-13.48	1.37	5.75	26.97	5.39
6	- 2.55	4.48	- 2.10	10.55	-10.37	30.05	6.01
Total $\lvert\hat{e}_{ij}\rvert$	27.76	30.96	42.30	53.93	66.10	221.05	-
Mean $\lvert\hat{e}_{ij}\rvert$	4.63	5.16	7.05	8.99	11.02	-	7.37

Analysis of variance on $\lvert\hat{e}_{ij}\rvert$ and "F-ratios"

Source of variation	"d.f."*	"Mean square"	"F-ratio"
Total	20	118.986	
Correction for mean	2/3	2,443.155	
Blocks	10/3	17.311	0.44
Treatments	8/3	63.966	1.63
Blocks x treatments	40/3	39.200	

*"d.f." = d.f. in ANOVA table times (r-1)(v-1)/rv in order to make "d.f." add to (r-1)(v-1), the degrees of freedom for the total sum of squares for the $\lvert\hat{e}_{ij}\rvert$.

Table 4.2. (Cont'd)

$Y^{+}_{ij} = Y_{ij}/\tau^{+}_{i}$ values

Block	Treatments 1	2	3	4	5	Mean $y^{+}_{.j}$	Variance s^2_j
1	16.4	28.6	42.3	22.2	37.1	29.3	88.96
2	24.1	32.0	23.1	37.8	18.3	27.0	48.26
3	21.9	26.3	22.4	12.8	32.7	23.2	42.19
4	58.3	19.0	47.5	34.7	27.4	37.4	197.21
5	41.1	29.7	16.0	30.4	31.4	29.7	64.30
6	18.5	28.4	23.2	34.0	20.1	24.8	32.42

$Y^{+}_{ij} - y^{+}_{.j} = e^{+}_{ij}$ values

Block	Treatments 1	2	3	4	5	Sum
1	-12.9	- 0.7	13.0	- 7.1	7.8	.1
2	- 2.9	5.0	- 3.9	10.8	- 8.7	.3
3	- 1.3	3.1	- 0.8	-10.4	9.5	.1
4	20.9	-18.4	10.1	- 2.7	-10.0	-.1
5	11.4	0.0	-13.7	0.7	1.7	.1
6	- 6.3	3.6	- 1.6	9.2	- 4.7	.2
Sum	8.9	- 7.4	3.1	0.5	- 4.4	0.7
"s^2_i"	128	63	79	61	58	

Analysis of variance on $|e^{+}_{ij}|$ values and "F-ratios"

Source of variation	"d.f."	"Mean square"	"F-ratio"
Blocks	10/3	62.11	1.45
Treatments	8/3	19.62	0.46
Blocks x treatments	40/3	42.76	

(vi) compute residuals for Y^{+}_{ij} values for a one-way analysis (note that the within block sum of squares will have (r-1)(v-1) and not r(v-1) degrees of freedom because the Y^{+}_{ij} values are corrected for treatment effects),

(vii) compute "variances" s^{2}_{i} for treatments (in Table 4.2, we note that the relationship between treatment means and variances now appears to be nonexistent and the variances are stabilized, except for perhaps the block 4 variance), and

(viii) run a two-way ANOVA on the $|e^{+}_{ij}|$ values (last part of Table 4.2).

Performing the above on the data given in the first part of Table 4.2, we note that the variances appear to be stabilized, except for perhaps block 4. Here we note two relatively large residuals, 20.9 and -18.4. The next step would be to question the experimenter about these two values to determine whether or not they are outliers and the reasons for the discrepancies. Otherwise, equation (4.6) appears to fit the data as well as any other.

Example 4.2. Data on number of poppy plants in oats are given in Table 11.15.1 of Snedecor and Cochran (1967), page 326. The experiment was designed as a randomized complete block experiment design of four blocks and five treatments, A, B, C, D, and E, where A and B were untreated experimental units and C, D, and E were treatments for controlling weeds (including poppies) in oats plots. The data are given in Table 4.3 where analyses of variance of the form of Table 4.1 are performed on Y_{ij} = number of poppies per 3.75 square feet, $\sqrt{Y_{ij}}$, and $Y^{0.8}_{ij}$, and on the data from blocks 1, 2, and 4 for Y_{ij} and log Y_{ij}. Snedecor and Cochran (1967) used the square root transformation, but did not note the fact that block 3 yields are very different from the remaining three blocks. This becomes immediately apparent when one observes the $b_j - 1 - \alpha\hat{\beta}_j$ values, where block 3 has a large negative value and the remaining ones are all positive.

Table 4.3. Five treatments to control weed infestation in cereals arranged in a randomized complete block design with r = 4 replicates. Number of poppy plants per 3.75 square feet (Table 11.15.1, Snedecor and Cochran (1967)).

Blocks	Treatments (number of poppy plants) A	B	C	D	E	$\hat{\beta}_j=\bar{y}_{.j}-\bar{y}_{..}$	b_j	$b_j-1-\alpha\hat{\beta}_j$
1	438	538	77	17	18	24.1	1.289	0.195
2	442	422	61	31	26	2.9	1.112	0.101
3	319	377	157	87	77	9.9	0.699	-0.339
4	380	315	52	16	20	-36.9	0.900	0.044
$\tau_i=\bar{y}_{i.}-\bar{y}_{..}$	201.25	219.50	-106.75	-155.75	-158.25			
b_i	0.475	3.074	0.814	0.376	0.260	$\bar{y}_{..}$ = 193.5		
$b_i-1-\alpha\hat{\tau}_i$	-1.311	1.216	0.231	-0.015	-0.121			

$\alpha = \Sigma_1^r \hat{\beta}_j b_j / \Sigma_1^r \hat{\beta}_j^2 = 0.0039$; $1 - \alpha\bar{y}_{..} = p = 0.244$.

Analysis of variance

Source of variation	d.f.	Mean square Y_{ij}	$Y_{ij}^{0.5}$	$Y_{ij}^{0.8}$
Blocks	3	3,414.73	7.538	275.26
Treatments	4	149,378.50	216.739	11,811.20
Blocks x treatments	12	3,254.56	4.058	230.84
TNA	1	4,673.40	11.816	0.12
Blocks (deviations)	2	12,337.80	14.536	1,080.50
Treatments (deviations)	3	2,231.34	1.590	144.04
Remainder	6	501.96	0.523	29.47

Table 4.3. (Cont'd)

Analysis of variance omitting block 3

Source of variation	d.f.	Y_{ij}		log Y_{ij}	
		Mean square	F-ratio	Mean square	F-ratio
Blocks	2	4,795.40	2.11	0.121,146	2.91
Treatments	4	135,800.00	59.72	6.856,240	164.83
Blocks x treatments	8	2,273.91	-	0.041,596	-
TNA	1	10,754.20	23.05	0.006,991	0.17
Blocks (deviations)	1	138.03	0.30	0.148,606	3.61
Treatments (deviations)	3	1,966.37	4.21	0.017,870	0.43
Remainder	3	466.63	-	0.041,188	-

Also, the blocks (deviations) have large F-values for both Y_{ij} and $\sqrt{Y_{ij}}$; the TNA F-values are large for Y_{ij} and $\sqrt{Y_{ij}}$ but the F is small for $Y_{ij}^{0.8}$ values. From the analysis on $\sqrt{Y_{ij}}$, $1 - \alpha\bar{y}_{..} = 2.33$, indicating $(\sqrt{Y_{ij}})^{2.33} = Y_{ij}^{1.17}$, whereas the p value from the Y_{ij} analysis is $p = .24 \doteq 1/4$ or $Y_{ij}^{0.25}$. Thus, there is a considerable discrepancy on what Y_{ij}^{p} values to use, possibly caused by the inconsistent results in block 3 where the differences between treatments A and B and C, D, and E are greatly reduced.

Leaving out block 3, and performing the analysis of Table 4.1 on the remaining data for Y_{ij}, we find that TNA has a large F-ratio and blocks (deviations) and treatments (deviations) have relatively small F-ratios. The analysis on the Y_{ij} values indicates that $p = 1 - \alpha\bar{y}_{..} = -0.06$. This means that log Y_{ij} or log (Y_{ij} - constant) values would yield additivity. When log Y_{ij} are analyzed, all types of nonadditivity have disappeared. Unless there is biological evidence to the contrary, one should analyze these data using a logarithmic transformation.

5. LATIN SQUARE DESIGN

Two population structures for a latin square experiment design have been discussed by Federer (1976b), where he should have mentioned that these are the population structures resulting after, as well as before, the experiment has been conducted. The first of two population structures discussed is to have subpopulations arranged in rows and columns (generic terms for two-way variation) and to randomly select r of R rows of subpopulations and c of C columns of subpopulations. The row-column intersection results in the subpopulation selected. Then, from the rc selected subpopulations, enough sampling units to form one experimental unit are randomly selected from each of the rc subpopulations. The

measurement and other variation encountered during the course of the experiment should be completely random over the experiment or should be completely confounded with the original row and column effects. If, during the conduct of an experiment, a third source of variation, e.g., invasion by insects, enters and is not completely accounted for by either rows or columns, this variation should be taken into account in the statistical analysis, perhaps by covariance.

The second type of population and sampling structure discussed in Federer (1976b) was that the population consisted of a single population of sampling units, where each sampling unit contained two sources of variation. Thus, the sampling units could be fields, which have gradients in two directions. This situation corresponds to that for some of the agronomic experiments conducted in a single field, and where a latin square experiment design is used.

A third population and sampling structure would be to have an overall population composed of subpopulations with one sampling unit selected from each subpopulation; then a sequence of treatments would be applied to the sampling unit through time. An experimental unit would be the period of time one treatment is applied to the sampling unit. The time periods would be the rows and the sampling units would be the columns of a latin square experiment design. Many types of experiments fall in this category, e.g., patients, animals, students, machines, plots of land for long term agricultural experiments, stores, hospitals, factories, etc.; the class of repeated measures experiment designs would have this population structure.

Some response model equations for consideration are:

$$Y_{hij} = \mu + \rho_h + \gamma_i + \tau_j + \epsilon_{hij} \quad \text{(classical additive equation)}, \tag{5.1}$$

where μ is an overall effect, ρ_h is the h^{th} subpopulation row effect, γ_i is the i^{th} subpopulation column effect, and τ_j is the j^{th} treatment direct effect,

$$f(Y_{hij}) = \mu + \rho_h + \gamma_i + \tau_j + \epsilon_{hij} \quad \text{(sometimes used)} , \tag{5.2}$$

$$Y_{hij} = \mu\rho_h\gamma_i\tau_j\epsilon_{hij} \quad \text{(multiplicative model)} , \tag{5.3}$$

$$Y_{hij} = \mu_{h..}\mu_{.i.}\mu_{..j}/\mu^2 + \epsilon_{hij} = \mu + (\mu_{h..}-\mu=\rho_h) + (\mu_{.i.}-\mu=\gamma_i)$$

$$+ (\mu_{..j}-\mu=\tau_j) + (\rho_h\gamma_i+\rho_h\tau_j+\gamma_i\tau_j)/\mu + \rho_h\gamma_i\tau_j/\mu^2 + \epsilon_{hij} , \tag{5.4}$$

$$Y_{hij} = (\mu+\rho_h+\gamma_i+\tau_j)^p + \epsilon_{hij} , \tag{5.5}$$

$$Y_{hij} = (\mu+\rho_h+\gamma_i+\epsilon_{hij})\tau_j , \tag{5.6}$$

and

$$Y_{hij} = (\mu+\epsilon_{hij})\rho_h\gamma_i\tau_j . \tag{5.7}$$

The above response model equations can be considered for both equal and unequal variances, for both correlated and uncorrelated observations, for known or unknown distributions, and for discrete or continuous variables. When all effects are positive, equation (5.3) can be converted into (5.2). Here, as before, equations (5.3), (5.4), (5.6), and (5.7) imply a ratio of treatment effects rather than differences of effects as implied by equations (5.1), (5.2), and (5.5). Equation (5.3) is of the form for computing Tukey's one degree of freedom for nonadditivity, (5.5) is of the form considered by D. S. Robson, unpublished results, for nonadditivity, (5.6) was considered by Nair (1976), and (5.7) is another form of (5.6) where row and column effects come in a response model equation in the same manner as the treatment effects. Nair (1976) has obtained maximum likelihood solutions for μ, ρ_h, γ_i, and τ_i in (5.6) under normality, independence, and variance homoscedasticity of residuals; he obtains the solutions by minimizing the following sum of squares:

$$\sum_h \sum_i (Y_{hij}/\tau_j-\mu-\rho_h-\gamma_i)^2 . \tag{5.8}$$

The procedure is iterative.

A one-degree-of-freedom test for nonadditivity for a randomized complete block experiment design is to regress the residuals from (4.1), $\hat{e}_{ij}$, on the residuals from (4.4), $e^{+}_{ij} = Y_{ij} - \bar{y}_{i.}\bar{y}_{.j}/\bar{y}_{..}$; then, the sum of squares is computed as $(\Sigma_i\Sigma_j\hat{e}_{ij}e^{+}_{ij})^2/\Sigma_i\Sigma_j(e^{+}_{ij})^2$. Likewise, for a one-degree-of-freedom sum of squares in a latin square design, we regress the residuals, $\hat{e}_{hij}$ from (5.1) on the residuals from (5.4), i.e., $e^{+}_{hij} = Y_{hij} - \bar{y}_{h..}\bar{y}_{.i.}\bar{y}_{..j}/\bar{y}^2_{...}$; then, the sum of squares for nonadditivity is computed as $\Sigma_h\Sigma_i(\hat{e}_{hij}e^{+}_{hij})^2/\Sigma_h\Sigma_i(e^{+}_{hij})^2$. This test is different from that put forth by J. W. Tukey (see Snedecor and Cochran (1967), pages 334-6) in that he puts the term $\hat{\rho}_h\hat{\gamma}_i\hat{\tau}_j/\bar{y}^2_{...}$ in with e^{+}_{hij} whereas we include it in the nonadditivity sum of squares. The nonadditivity here can be from rows and columns, rows and treatments, columns and treatments, or/and from all three. These different types are not separable in a latin square design (see Nair (1976)).

Viewing computation of sum of squares for nonadditivity for equations (4.1) and (4.4) or (5.1) and (5.4), as we have done, allows one to compute this sum of squares for an n-way classification which may have the effects orthogonal or nonorthogonal. Likewise, it is simpler pedagogically than the presentation in Snedecor and Cochran (1967). One may also compare equations (4.5) and (4.4) rather than (4.1) versus (4.4) or (4.5), in the same manner as we have used here. That is, any equation may be taken as the null and any other as the alternative.

6. DISCUSSION

We have considered classes of response model equations for the completely randomized, the randomized complete block, and the latin square experiment designs. These results are directly extendible to any other experiment design such as incomplete blocks, double-change-over, latin cube, F-square and cube, split plot,

split block, etc. The results are directly extendible to any n-way classification, orthogonal or not, and may be used in connection with the statistical analysis of factorial treatment designs.

The present status of statistical model formulation is primitive. We approach response equations in a deterministic manner and then add stochastic variation. This may not be the manner in which responses are formed in nature. The status of statistical model selection is also primitive. Perhaps the best we can do is to select a class of response models on subject matter grounds and then select one "best fitting the data" model from this class. For example, this is what Box and Cox (1964) do. When the writers of statistical textbooks and the teachers of statistics courses realize that something more than hypothesis testing and assuming a linear model is involved in the real world of statistical applications, perhaps progress will be made on these two very important topics of model formulation and selection of a model from a class of models.

REFERENCES

Atkinson, A. C. and Fedorov, V. V. (1976). The design of experiments for discriminating between two rival models. Biometrika 26(1):57-70.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. J. Royal Statistical Society, B, 26(2):211-252.

Federer, W. T. (1951). Evaluation of variance components from a group of experiments with multiple classifications. Iowa Agric. Expt. Sta. Bull. 380:24-310.

Federer, W. T. (1955). Experimental Design: Theory and Application. Macmillan, New York (republished by Oxford and IBH Publishing Co., Calcutta, India, 1967, 1974, 1977).

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

Federer, W. T. (1976b). Sampling, blocking, and model considerations for the r-row by c-column experiment designs. Biometrische Zeitschrift 18(8): 601-613.

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.

Jones, E. R. and Mitchell, T. J. (1978). Design criteria for detecting model inadequacy. Biometrika 65(3):541-551.

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

Nair, R. C. (1976). Some contributions to the theory of nonadditivity in data analysis. Ph.D. Thesis, Cornell University, January.

Snedecor, G. W. and Cochran, W. G. (1967). Statistical Methods, 6th edition. Iowa State University Press, Ames, Iowa, U.S.A.

Tukey, J. W. (1975). Instead of Gauss-Markov least squares, what? Applied Statistics (Editor, R. P. Gupta), North Holland Publishing Co., Amsterdam, Oxford, pp. 351-372.