

# DIAGNOSTIC PROCEDURES FOR ANALYSES OF VARIANCE

by

Walter T. Federer

BU-1161- M\*

June 1992

## Abstract

The analysis of variance (ANOVA) developed by Sir Ronald A. Fisher around 1920 is a partitioning of the total variance into its component parts. Several uses may be made of the results from an analysis of variance such as, e.g., calculating efficiency of stratification or blocking, estimation of variance components, making significance or hypothesis tests, and obtaining an error mean square for constructing simultaneous confidence intervals and making multiple comparisons. Each use carries with it a set of assumptions about the statistical design, the response model, and the distribution of the random elements of the model. Diagnostic procedures to detect deviation from the assumptions have been developed for several types of departures. The statistical design must be such that it is representative of the population for which inferences are being made. An appropriate response model must be validated for an investigation and not merely obtained by definition. Some response models and methods for determining an appropriate response model are presented. Many response models are based upon additivity of effects. Several tests for non-additivity are discussed. Many uses of the results in an ANOVA require homoscedascity, and several procedures for detecting variance heterogeneity are available, some of which are discussed. Most uses require independent observations but this requirement is violated in certain investigations. Four situations leading to non-independence of responses are discussed. Statistical designs and analyses for removing the effect causing non-independence are presented. Patterns, trends, and discrepant observations of residual effects need to be studied. An example of a single outlier is used to demonstrate how the interpretation of results can be changed if the outlier is ignored. In modeling responses from two-factor factorials, it is desirable to do this in a parsimonious manner and retain as few parameters as possible. Bi-plot and AMMI procedures are useful in this context. Finally, after an investigator has summarized the results of an investigation, a critical examination, a post mortem, should be made for all aspects of the investigation. In particular, the statistical design, the analyses, the results, and the interpretations should be scrutinized carefully to ascertain that nothing has gone wrong and that everything makes sense.

**Key words:** Model selection, non-additivity, variance heterogeneity, statistical design, independence of observations, exploratory data analysis, residuals

\*In the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, N.Y.

## 1. Introduction

An analysis of variance (ANOVA) is a partitioning of the total variance into its component parts. The ideas in the development of ANOVA had their beginnings when the total sum of squares from a regression analysis was partitioned into that due to regression and that due to deviations from regression and when J. Arthur Harris (1913) introduced the intraclass correlation coefficient. Sir Ronald A. Fisher formalized these ideas and called this the analysis of variance in a series of publications around 1920 (1918, 1923, 1925, etc.). In the published statistical literature, an ANOVA may mean different things to different writers and users of an analysis of variance. Perhaps the most frequent use associated with an analysis of variance is the computation of an F or z statistic. Many users appear to denote this use as an ANOVA. Many include the uses being made of an ANOVA as part of the definition. For the ensuing discussion, we use the definition of an ANOVA as given in the beginning of this paragraph devoid of the use to which it is put, i.e., a partitioning of the total variance into its component parts. Rather than using the name analysis of variance, perhaps the name for this procedure should have been a partitioning of variance.

To illustrate the ideas, we make use of two simple examples--a balanced one-way array such as a completely randomized experiment design (CRED) or a stratified simple random sample survey design and a balanced two-way array such as a randomized complete block experiment design (RCBED) or a two-way factorial treatment design. For a balanced one-way array the total sum of squares for  $rv$  observations from  $v$  groups of  $r$  items each is partitioned as:

$$\begin{aligned} \sum_{i=1}^v \sum_{j=1}^r Y_{ij}^2 &= \sum_{i=1}^v \sum_{j=1}^r [(Y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}_{..}) + \bar{y}_{..}]^2 \\ &= \sum_{i=1}^v \sum_{j=1}^r (Y_{ij} - \bar{y}_i)^2 + r \sum_{i=1}^v (\bar{y}_i - \bar{y}_{..})^2 + r v \bar{y}_{..}^2, \end{aligned} \quad (1.1)$$

where  $Y_{ij}$  is the  $j$ th observation in the  $i$ th group,  $\bar{y}_i$  is the arithmetic mean for group  $i$ , and  $\bar{y}_{..}$  is the arithmetic mean of the  $rv$  observations. The results from this algebraic partitioning of the total sum of squares may be put in what is known as an analysis of variance table as in Table 1.1.

For a balanced two-way array such as a randomized complete block experiment design with  $r$  blocks and  $v$  treatments, the total sum of squares may be partitioned as follows:

$$\begin{aligned} \sum_{i=1}^v \sum_{j=1}^r Y_{ij}^2 &= \sum_{i=1}^v \sum_{j=1}^r [(Y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..}) + (\bar{y}_i - \bar{y}_{..}) + (\bar{y}_j - \bar{y}_{..}) + \bar{y}_{..}]^2 \\ &= \sum_{i=1}^v \sum_{j=1}^r (Y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}_{..})^2 + r \sum_{i=1}^v (\bar{y}_i - \bar{y}_{..})^2 + v \sum_{j=1}^r (\bar{y}_j - \bar{y}_{..})^2 + r v \bar{y}_{..}^2, \end{aligned} \quad (1.2)$$

where  $\bar{y}_j$  is the arithmetic mean of the  $j$ th block and the other symbols are as defined above with groups being synonymous with treatments. We may put these sum of squares in an ANOVA as in Table 1.2.

Table 1.1 ANOVA for a one-way classification.

Source of variation	degrees of freedom	sum of squares	mean square (variance)
Total	$rv$	$\sum_{i=1}^v \sum_{j=1}^r Y_{ij}^2$	
Correction for mean	1	$r v \bar{y}_{..}^2$	
Groups	$v - 1$	$r \sum_{i=1}^v (\bar{y}_{i.} - \bar{y}_{..})^2 = G$	$G/(v - 1)$
Within groups	$v (r - 1)$	$\sum_{i=1}^v \sum_{j=1}^r (Y_{ij} - \bar{y}_{i.})^2 = W$	$W/v(r - 1)$

Table 1.2 ANOVA for a RCBED.

Source of variation	degrees of freedom	sum of squares	mean square (variance)
Total	$rv$	$\sum_{i=1}^v \sum_{j=1}^r Y_{ij}^2$	
Correction for mean	1	$r v \bar{y}_{..}^2$	
Blocks	$r - 1$	$v \sum_{j=1}^r (\bar{y}_{.j} - \bar{y}_{..})^2 = B$	$B/(r - 1)$
Treatments	$v - 1$	$r \sum_{i=1}^v (\bar{y}_{i.} - \bar{y}_{..})^2 = T$	$T/(v - 1)$
Remainder	$(r - 1)(v - 1)$	$\sum_{i=1}^v \sum_{j=1}^r (Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = R$	$R/(r - 1)(v - 1) = E$

Note that the above results make no assumptions about a response model, about effects, about independence of effects, about additivity of effects, about homoscedasticity, or about the statistical distribution of any of the elements presented. The above is purely algebraic manipulation. Requirements for making practical use of the above will vary with the application made of the results. The first requirement is that the items listed under "Source of variation" must have practical meaning for the experimenter who obtained these results. Once it has been established that the items have meaning, the above partitioning may be used in several ways. For the partitioning in equation (1.1) and where the groups are strata, the use requiring the fewest assumptions is the computation of the efficiency of a stratified survey design relative to a simple random sample. The following form of a response model is implied:

$$Y_{ij} = f(\mu, \alpha_i) + \epsilon_{ij} \tag{1.3}$$

where  $f(\mu, \alpha_i)$  is usually taken to be  $\mu + \alpha_i$  with  $\mu$  being an overall mean of a population and  $\mu + \alpha_i$  being the true mean for the  $i$ th stratum, and with  $\epsilon_{ij}$  being a random error effect associated with the  $ij$ th observation. A measure of efficiency of stratification versus no stratification is:

$$\left[ \sum_{i=1}^v \sum_{j=1}^r (Y_{ij} - \bar{y}_{..})^2 / (rv - 1) \right] / [W / v(r - 1)] \tag{1.4}$$

For a RCBED, a measure of efficiency of an RCBED relative to a CRED is:

$$[B + R + R / (r - 1)] / E (rv - 1) . \tag{1.5}$$

For this use, it is assumed that the effects are additive in the response model and that  $W/v(r - 1)$  is an unbiased estimate of the within stratum variance and that  $R/(v - 1)(r - 1) = E$  is an unbiased estimate of the variance for the RCBED. This implies that the  $\epsilon_{ij}$  are independently distributed with mean zero and variance parameter  $\sigma_\epsilon^2$ . Randomization was used to obtain the unbiasedness and independence.

For a variance component analysis, a linear model of the form

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \text{ for a CRED} \tag{1.6}$$

or

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \text{ for a RCBED} \tag{1.7}$$

is assumed. In addition, it is assumed that the  $\alpha_i$  are identically and independently distributed (I I D) with mean zero and variance  $\sigma_\alpha^2$ , that the  $\tau_i$  are I I D(0,  $\sigma_\tau^2$ ), the  $\beta_j$  are I I D(0,  $\sigma_\beta^2$ ), and that the

$\epsilon_{ij}$  are I I D( $0, \sigma_\epsilon^2$ ). In some cases, the various variances could be a linear combination of other variance components and this would affect the use of a variance component. Note that the form of the distribution is unspecified and the effects are random.

For many situations, the effects are fixed effects and the expected value,  $E(\cdot)$ , of an effect is the effect parameter and not necessarily zero. For most CREds and RCBEDs,  $E(\alpha_i) = \alpha_i$  and  $E(\tau_i) = \tau_i$ , and it is desired to have interval estimates of these effects and perhaps to perform significance or hypothesis tests. For this, it is necessary to specify the form of the distribution for the  $\epsilon_{ij}$ , which is mostly taken to be a normal distribution, i. e., N I I D. Note that it only makes experimental sense to consider the  $\beta_j$  as random effects even though the majority of statistical methods books imply that they are fixed effects.

In the following, we shall look at techniques designed to determine if one or more of the requirements for a particular use has been violated. *Assuming* that all the requirements for a statistical procedure are satisfied does not mean that *they are* for a particular experiment. Conclusions about results may be considerably altered if the requirements for a procedure are not satisfied.

## 2. Selection of a Response Model Equation

It is common practice in statistical literature to state “‘the’ linear model is ...”. This is an incorrect statement in that the best anyone can do is to say “‘a’ linear model is ...”. Proving that a linear model is unique, or even that the model is linear, is next to impossible in the majority of cases. It can, however, be a good first approximation to an appropriate response model for an investigation. In the “chalkboard world” of the classroom, a linear model is obtained by definition and not from the actual situation in the investigation. However, an investigator does not have this luxury, but must make a decision about which response model to use for each investigation being conducted. A good discussion of model selection may be found in Box (1980).

Box and Cox (1964) in an excellent paper, work with a parametric family of transformations of data from  $Y$  to  $Y^\lambda$  such that  $Y^\lambda$  for a specified value of  $\lambda$  is a normal homoscedastic linear model. They present procedures for separating the contributions of normality, homoscedasticity, and additivity. The two important examples they consider are

$$Y^{(\lambda)} = \begin{cases} (Y^\lambda - 1) / \lambda \text{ or simply } Y^\lambda & \text{for } \lambda \neq 0 \\ \log Y & \text{for } \lambda = 0, Y > 0 \end{cases} \quad (2.1)$$

and

$$Y^{(\lambda)} = \begin{cases} [(Y + \lambda_2)^{\lambda_1} - 1] / \lambda_1 & \text{for } \lambda_1 \neq 0, \lambda_2 > 0 \\ \log (Y + \lambda_2) & \text{for } \lambda_1 = 0 \text{ and } Y > -\lambda_2 \end{cases} \quad (2.2)$$

They state the procedure is simpler if a normalized transformation is used for (2.1) and (2.2), respectively, as follows which for the simple power transformation is

$$Z^{(\lambda)} = (Y^\lambda - 1) / \lambda \tilde{Y}^{\lambda-1}, \quad (2.3)$$

and for the power transformation with a shifted location is

$$Z^{(\lambda)} = [(Y + \lambda_2)^{\lambda_1} - 1] / \lambda_1 [\text{gm}(Y + \lambda_2)]^{\lambda_1 - 1}, \quad (2.4)$$

where  $\tilde{Y}$  is the geometric mean of the  $Y$ s and  $\text{gm}(Y + \lambda_2)$  is the geometric mean of the  $(Y + \lambda_2)$ s. Let  $S(Z, \lambda)$  denote the residual sum of squares in an ANOVA.  $S(Z, \lambda)$  is computed for a series of values of  $\lambda$ , and the  $\lambda$  producing the minimum residual sum of squares is the value to use for the power transformation. They show how to compute an approximate  $(100 - \alpha)$  per cent confidence interval for  $\lambda$ . It should be noted that there may be no value of  $\lambda$  which fits the data set in that the correct response model for the data is not in this family of power transformations.

Another model that is rather widely used is the one for a diallel crossing experiment in genetics. This model has been found to have uses in a variety of situations, such as, e.g., psychological and personnel rating investigations. A response model of this type was introduced by W. G. Cochran in a paper by Sprague and Tatum (1941). Griffing (1956) presents models for several diallel crossing situations. One of these is discussed in some detail in Federer (1955). The concepts of general combining ability (how a line combines on the average with the other  $v - 1$  lines with which it is crossed) and specific combining ability (how a line performs with a particular line) were used in constructing the response models.

Federer (1979, 1992) has constructed a number of response models for dealing with cropping systems for multiple cropping situations with and without changes in densities of the crops in a mixture. The concepts of general mixing ability, general competing ability, bi-mixing ability, bi-competing ability, tri-mixing ability, etc. as well as the effect of a second crop's density on the yield of a given crop, have been constructed and applied to specific examples. Procedures for combining the yields of the components of a mixture system were also devised. The ideas associated with diallel crossing models and those of Martin (1980) were useful in constructing these models.

For  $r$ -row by  $c$ -column designs, several variations have appeared in the literature, but the following is mostly used by textbook writers whether it is appropriate or not for the examples selected to illustrate the computations:

$$Y_{hij} = \mu + \rho_h + \gamma_i + \tau_j + \epsilon_{hij}, \quad (2.5)$$

where  $Y_{hij}$  is the response for treatment  $j$  in the  $h$ th row and  $i$ th column,  $\mu$  is a general mean effect,  $\rho_h$

is an effect for the  $h$ th row,  $\gamma_i$  is an effect for the  $i$ th column,  $\tau_j$  is an effect for the  $j$ th treatment, and  $\epsilon_{hij}$  are random error effects which are I I D  $(0, \sigma_\epsilon^2)$ . Cox (1958) proposed a response model for the situation wherein the gradients within each column (or row) differ. The form of the response model is

$$Y_{ij} = \mu + \gamma_i + \beta_i a_{ij} + \tau_j + \epsilon_{ij} , \quad (2.6)$$

where  $Y_{ij}$  is the response for treatment  $j$  in column  $i$ ,  $\beta_i$  is the linear regression of  $Y_{ij}$  on  $a_{ij}$  to depict the linear regression in column  $i$ ,  $a_{ij}$  are row constants for the  $j$ th treatment in the  $i$ th column, measured from a zero mean at the center of the row with  $\sum_{i=1}^c a_{ij}^2 = \sum_{j=1}^v a_{ij}^2 = 1$  and  $v = r$ , and the other effects are defined as in (2.5). The  $\gamma_i$  are the location parameters for the linear regressions. As Cox (1958) points out, other polynomial regression coefficients may be added to (2.6). He illustrates this by adding quadratic after linear constants and demonstrates the computations on a numerical example. Federer and Schlotfeldt (1954) and Outhwaite and Rutherford (1955) used a model with constant regressions in each row but a more appropriate response model and analysis for their example would have been (2.6). It should also be noted that fitting a polynomial of degree  $v - 1$  as used by Outhwaite and Rutherford (1955) is the same as using model (2.5).

For  $r$ -row (period) by  $c$ -column designs where the treatments and rows (periods) are added sequentially to the same sampling unit and which are known as rotation and as repeated measures change-over designs, a variety of response models may be utilized depending upon the nature of the responses. There may be direct effects of the treatment in the period in which it is applied, first-, second-, etc. period carry-over effects of the previous treatment, continuing effects of treatments, and permanent effects (see e.g., Kershner and Federer, 1981, who present designs for obtaining estimates of parameters for a variety of response models).

The population structures associated with an experiment and/or treatment design has been ignored in statistical literature except for Fisher (1935) and Federer (1976a, 1976b, 1977, 1991). The latter author discusses population structures and response models for block, row-column, split-plot, and split-block designs. Kempthorne (1952), e.g., discusses randomization models over the particular set of experimental units for a particular experiment without regard as to how the sample was obtained from the population. Assuming that any sample obtained by an experimenter is representative of the target population can be grossly incorrect and often is. More attention needs to be given to the planning stages of an investigation in order to accurately make inferences about the target population.

Several response models have been proposed to take account of various forms of non-additivity in experiments. Some of these are described in the following section.

### 3. Tests for Non-Additivity

In a classic paper, Tukey (1949) provided a one degree of freedom test for non-additivity in a two-way array such as an RCBD. The test statistic is

$$\begin{aligned} & \left[ \sum_{i=1}^V \sum_{j=1}^F (Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..}) \right]^2 / \sum_{i=1}^V (\bar{y}_{i.} - \bar{y}_{..})^2 \sum_{j=1}^F (\bar{y}_{.j} - \bar{y}_{..})^2 \\ & = \left[ \sum_{i=1}^V \sum_{j=1}^F \hat{\epsilon}_{ij} (\hat{\delta}_{ij} - \hat{\epsilon}_{ij}) \right]^2 / \sum_{i=1}^V \sum_{j=1}^F (\hat{\delta}_{ij} - \hat{\epsilon}_{ij})^2, \end{aligned} \quad (3.1)$$

which is the sum of squares due to regression of  $\hat{\epsilon}_{ij}$  on  $(\hat{\delta}_{ij} - \hat{\epsilon}_{ij})$ . In the above,

$$\hat{\epsilon}_{ij} = Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \quad (3.2)$$

and

$$\hat{\delta}_{ij} = Y_{ij} - \bar{y}_{i.} \bar{y}_{.j} / \bar{y}_{..} \quad (3.3)$$

For a three way classification, it is suggested that the residual for the alternate multiplicative model be computed as

$$\hat{\delta}_{hij} = Y_{hij} - \bar{y}_{h..} \bar{y}_{.i.} \bar{y}_{.j} / \bar{y}_{...}^2, \quad (3.4)$$

for a four way classification, that the residual be computed as

$$\hat{\delta}_{ghij} = Y_{ghij} - \bar{y}_{g...} \bar{y}_{.h..} \bar{y}_{.i.} \bar{y}_{.j} / \bar{y}_{....}^3, \quad (3.5)$$

and so forth for higher way classifications. Then, use may be made of a form similar to equation (3.1) to compute the sum of squares for non-additivity. Note that  $\hat{\delta}_{hij}$  in (3.4) is different from the residual used by Tukey for a latin square design (see e.g., Snedecor and Cochran, 1980, section 15.14). The above forms consider all interactions and not only two factor interactions as suggested by Tukey.

Robson (1970) studied the family of transformations

$$E[Y_{ij}] = (\mu + \rho_i + \gamma_j)^p \quad (3.6)$$

and reported on a test for non-additivity for  $p = 2$ . The residual is computed as

$$\hat{\pi}_{ij} = Y_{ij} - (\hat{\mu} + \hat{\rho}_i + \hat{\gamma}_j)^2, \quad (3.7)$$

where

$$(\hat{\gamma}_j - \bar{\gamma}_.) = \sqrt{(\bar{y}_{.j} - \sigma_\rho^2)} - \frac{r}{\sum_{j=1}^r \sqrt{(\bar{y}_{.j} - \sigma_\rho^2)}} / r, \quad (3.8)$$

$$(\hat{\rho}_i - \bar{\rho}_.) = \sqrt{(\bar{y}_{i.} - \sigma_\gamma^2)} - \frac{v}{\sum_{i=1}^v \sqrt{(\bar{y}_{i.} - \sigma_\gamma^2)}} / v, \quad (3.9)$$

$$\sigma_\rho^2 = \frac{v}{\sum_{i=1}^v (\hat{\rho}_i - \bar{\rho}_.)^2} / v, \quad (3.10)$$

$$\sigma_\gamma^2 = \frac{r}{\sum_{j=1}^r (\hat{\gamma}_j - \bar{\gamma}_.)^2} / r, \quad (3.11)$$

and

$$\frac{r}{\sum_{j=1}^r \sqrt{(\bar{y}_{.j} - \sigma_\rho^2)}} / r = \frac{v}{\sum_{i=1}^v \sqrt{(\bar{y}_{i.} - \sigma_\gamma^2)}} / v. \quad (3.12)$$

Then,  $\hat{\pi}_{ij}$  is substituted for  $\hat{\delta}_{ij}$  in equation (3.1) to obtain Robson's test for non-additivity for the above model. He notes that a similar test may be constructed for  $p = -1$ , the reciprocal transformation. Robson (1970) states that Tukey's test compares the additive model

$$E[ Y_{ij} ] = \mu + \tau_i + \beta_j, \quad (3.13)$$

against the multiplicative alternative model

$$E[ Y_{ij} ] = (\mu + \tau_i) (\mu + \beta_j) / \mu, \quad (3.14)$$

whereas his test against the additive model is for the alternative hypothesis

$$E[ Y_{ij} ] = (\mu + \tau_i + \beta_j)^2. \quad (3.15)$$

A numerical comparison of Robson's and Tukey's test along with a third one was made by Federer (1970) on a numerical example involving counts of an insect and treatments for controlling the insects in an RCBD. Despite the fact that a square root transformation might be appropriate, Tukey's test recovered the largest sum of squares.

Mandel (1961, 1971) extended Tukey's test for non-additivity by regressing the residual from (3.2) on the difference between the residuals in (3.2) and (3.3) for each treatment (and/or for each block) in an RCBD and then computing a sum of squares to compare the  $v$  treatment regressions. Kirton (1984) extended Mandel's procedure to include simultaneously both categories of a two way

classification, and to compute an ANOVA using the absolute values of the residuals both for an additive and for a non-additive model. The procedure is presented here, and the response model when both blocks and treatments non-additivity effects are included, is

$$Y_{ij} = \mu + \tau_i + \beta_j + \tau_i\beta_j / \mu + (\rho_j - 1)\tau_i + (\rho_i - 1)\beta_j + \epsilon_{ij} , \quad (3.16)$$

where  $\rho_i$  is the linear regression coefficient for treatment i, i.e.,

$$\hat{\rho}_i = \frac{\sum_{j=1}^r Y_{ij} (\bar{y}_{.j} - \bar{y}_{..})}{\sum_{j=1}^r (\bar{y}_{.j} - \bar{y}_{..})^2} , \quad (3.17)$$

$\rho_j$  is the linear regression coefficient for block j, i.e.,

$$\hat{\rho}_j = \frac{\sum_{i=1}^r Y_{ij} (\bar{y}_{i.} - \bar{y}_{..})}{\sum_{i=1}^r (\bar{y}_{i.} - \bar{y}_{..})^2} , \quad (3.18)$$

and the remaining symbols are as defined above. The formulae for the sums of squares are given in Table 3.1. In order to determine which p to use for the transformation  $Y^P$ , we compute

$$\alpha = \frac{\sum_{j=1}^r \hat{\beta}_j \hat{\rho}_j}{\sum_{j=1}^r \hat{\beta}_j^2} = \frac{\sum_{i=1}^r \hat{\tau}_i \hat{\rho}_i}{\sum_{i=1}^r \hat{\tau}_i^2} . \quad (3.19)$$

The value p is equal to  $1 - \alpha \bar{y}_{..}$ . Then to compare the various blocks, use  $\hat{\rho}_j - 1 - \alpha \hat{\beta}_j$ , and to compare the various treatments, use  $\hat{\rho}_i - 1 - \alpha \hat{\tau}_i$ .

There is a considerable literature on non-additivity and references may be found in the Current Index to Statistics. Also, since a considerable amount of work by D. S. Robson and his students has not been published, it would be wise for researchers in this area to check the Annual Reports of the Biometrics Unit at Cornell University for Technical Reports and Theses on this topic. The Annual Reports have been widely distributed. Also, the same type of search should be made at Princeton University, where J. W. Tukey and his students have done considerable work in this area.

#### 4. Homoscedasticity

Bartlett (1937) presented a test for homogeneity of a set of variances for a test of the null hypothesis  $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_v^2$  against the alternative  $H_1: \text{not } H_0$ . This test, sometimes called Bartlett's M test, is

$$M = [ N \ln ( \sum_{j=1}^v f_j s_j^2 / N ) - \sum_{j=1}^v f_j \ln s_j^2 ] / [ 1 + ( \sum_{j=1}^v 1 / f_j - 1/N ) / 3 (v - 1) ] , \quad (4.1)$$

where  $N = \sum_{j=1}^v f_j$ ,  $s_j^2$  is the jth variance in the set,  $\ln$  is the natural logarithm, and M is approximately distributed as chi square with  $v - 1$  degrees of freedom.

Table 3.1. 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	$rv$	$\Sigma_1^Y \Sigma_1^r Y_{ij}^2$
Correction for mean	1	$Y_{..}^2 / rv = C$
Blocks	$r - 1$	$\Sigma_1^r Y_{.j}^2 / v - C$
Treatments	$v - 1$	$\Sigma_1^Y Y_{i.}^2 / r - C$
Blocks $\times$ treatments	$(r - 1)(v - 1)$	$\Sigma_1^Y \Sigma_1^r (Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$
Nonadditivity	1	$TNA = \frac{[\Sigma_1^Y \Sigma_1^r Y_{ij} (\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})]^2}{\Sigma_1^Y (\bar{y}_{i.} - \bar{y}_{..})^2 \Sigma_1^r (\bar{y}_{.j} - \bar{y}_{..})^2}$
Blocks (deviations)	$r - 2$	$\Sigma_1^r (b_j - 1)^2 \Sigma_1^Y (\bar{y}_{i.} - \bar{y}_{..})^2 - TNA$
Treatments (deviations)	$v - 2$	$\Sigma_1^Y (b_i - 1)^2 \Sigma_1^r (\bar{y}_{.j} - \bar{y}_{..})^2 - TNA$
Remainder	$(r - 2)(v - 2)$	by subtraction

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

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

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

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

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

For the following tests, let  $f_j = f$  degrees of freedom for all  $j$ . Then, Cochran's  $W$  test is computed as

$$W = s_{\max}^2 / \sum_{j=1}^v s_j^2, \quad (4.2)$$

where  $s_{\max}^2 = \max \{s_1^2, s_2^2, \dots, s_v^2\}$ . Cochran (1941) gives upper 5% points for  $W$  when  $v = 3, 4, \dots, 10$  and  $f = 1, 2, 3, 4, 5, 6, 8, 10$ . Hartley (1950) proposed the following test for variance heterogeneity for a set of  $v > 2$  estimated variances  $s_j^2$ :

$$F_{\max} = s_{\max}^2 / s_{\min}^2, \quad (4.3)$$

where  $s_{\max}^2$  is defined above and  $s_{\min}^2 = \min \{s_1^2, s_2^2, \dots, s_v^2\}$ . Tables for the  $F_{\max}$  statistic may be found in Pearson and Hartley (1962). The tests  $W$  and  $F_{\max}$  provide for quick tests of variance heterogeneity. All three tests,  $M$ ,  $W$ , and  $F_{\max}$ , are non-robust in that the actual confidence or significance levels are sensitive to the form of the underlying distributions. They depend heavily upon the normality assumption. To remedy this sensitivity to the underlying distribution, Box (1953) proposed the following procedure:

- (i) divide each sample into  $c$  subsamples of size  $m$ ,
- (ii) compute the subsample variances  $s_{jk}^2, j = 1, 2, \dots, v, k = 1, 2, \dots, m$ ,
- (iii) set  $X_{jk} = \ln s_{jk}^2$ , and then  $E[X_{jk}] = \ln \sigma_j^2$  with variance  $\text{Var}(X_{jk}) = 2/(m-1) + \gamma/m$ , where  $\gamma$  stands for kurtosis,
- (iv) compute a one-way analysis of variance on the  $X_{jk}$ , and
- (v) use the  $F$  statistic to test  $H_0$ .

For the Box procedure, there is no firm rule for selecting the values of  $c$  and  $m$ , and the investigator is left to rely upon his own judgement. Scheffé (1959) gives more detail on this procedure.

For a two-way classification such as a RCBED or a two-way factorial, one may use Spearman's rank order correlation to test for a relationship between treatment (factor) means and residual sums of squares by computing the ranks of the  $\bar{y}_{.j}$  and  $\sum_{i=1}^r \hat{\epsilon}_{ij}^2$ ; then, use Spearman's rank order correlation (see Federer, 1979; D. S. Robson and C. L. Wood, personal communication, proved that the above test procedure is indeed distributed as Spearman's rank order correlation.)

Other procedures for comparing a set of variances and covariances are discussed in Federer (1955) and Votaw (1948). Several other test procedures for comparing a set of variances may be found in published literature. Oftentimes when variance heterogeneity is suspected or is present, it may be possible to select a transformation of the data to reduce or eliminate the heterogeneity (See, e.g., Federer, 1955 chapter II, Snedecor and Cochran, 1980, chapter 15, and Box and Cox, 1964.) Also, it may be desirable in certain situations to deal with unequal variances and use such procedures as a Behrens-Fisher method (See, e.g., Grimes and Federer, 1984.).

### 5. Independence of e.u.s

Non-independence of responses in an investigation may take several forms. Four such forms are

- (i) carry-over or residual effects from treatments in previous periods on the e.u.,
- (ii) competition between treatments in adjacent e.u.s,
- (iii) competition among treatments (intercropping mixtures, mixtures of drugs, mixtures of programs, etc.) which occur in the same e.u., and
- (iv) gradients from s.u. to s.u. or e.u. to e.u. within each stratum or block of an investigation.

Under (i), numerous experiment designs and statistical analyses have been developed to take into account the effects of treatments in previous periods on the response of a treatment in the present period. Some experiment designs for obtaining estimates of carry-over effects are variously known as repeated measures and double change-over or reversal designs. For three and four treatments (letters), particular designs are:

3 treatments							4 treatments				
column							column				
period	1	2	3	4	5	6	period	1	2	3	4
1	A	B	C	A	B	C	1	A	B	C	D
2	B	C	A	C	A	B	2	D	A	B	C
3	C	A	B	B	C	A	3	B	C	D	A
							4	C	D	A	B

Note that each letter is preceded by and is followed by every other letter, but not itself, an equal number of times. These designs are balanced for residual or carry-over effects. For four treatments, an alternate experiment design would be to use a set of three orthogonal latin squares of order four to form 12 columns. This results in an ED which is balanced for carry-over effects. Many forms of EDs and various response models have appeared in published literature (See, e.g., Kershner and Federer, 1981.). These designs allow estimation of the various direct and carryover effects of treatments when the same material is used in several periods.

Competition in experiments can take two forms, i.e., intra-experimental unit and inter-experimental unit competition. In many types of investigations, intra-e.u. competition is not a concern except as it may relate to the density in an e.u. In other types of investigation, such as growing a mixture of cultivars in the same area (one form of intercropping), using a mixture of various procedures, a mixture of drugs, or compounds to treat patients, using a mixture of various educational or recreational programs, etc., intra-e.u. competition effects and their estimation are a main concern of the investigation. To obtain estimates of the various effects, a major problem is to select appropriate

treatment designs and response models (See Federer, 1979, 1992, and Federer and Raghavarao, 1987.), whereas the selection of an appropriate ED is usually straight-forward. The appropriate treatment combinations for inclusion in a design depends upon the types of effects to be estimated and included in the selected response model.

Competition between adjacent e.u.s under (ii) will make for dependence among e.u.s in an experiment or investigation. Some treatments make be good competitors relative to their neighbors while others may be unaffected or poor competitors. Experimenters usually try to conduct experiments in such a manner as to eliminate any effect of competition by using space or border material, by discarding edge material of an e.u., by changing the size and/or shape of an e.u. to minimize or eliminate the effect of competition, or by constructing EDs and statistical analyses to estimate treatment effects free of competition effects. For the last, Kempton (1982), Besag and Kempton (1986), and Federer and Basford,(1991) have constructed various EDs and statistical analyses for this purpose. Kempton (1982), e.g., has constructed experiment designs balanced for competition effects from the two opposite sides of an e.u. Sequences of treatments (letters) in the various blocks may be of the form ABA (or BAB) for two treatments, ABACBCA for three treatments, ABCDADBACBDCA for four treatments, etc. Note that each letter precedes and follows each of the other letters but not itself and that the length of the sequence increases rapidly with the number of letters  $v$ . The sequences may be shortened if it is only required that a letter be adjacent to each of the other letters. E.g., for three letters ABCA and for four letters ABCDACBD are sufficient for each letter to be adjacent to each of the other letters. For large  $v$ , the length of a sequence becomes highly impractical.

For two-dimensional arrangements of e.u.s, Federer and Basford (1991) constructed EDs balanced for competition effects in both rows and columns. Three families of designs were given. For the first family, row-column latin square type designs are selected such that treatments precede and follow each other an equal number of times in both rows and columns. These have been denoted as complete latin squares but could have equally well have been denoted as double change-over designs in both rows and columns. For  $v = 4$  and  $6$ , the EDs are:

$v = 4$					$v = 6$						
column					column						
period	1	2	3	4	period	1	2	3	4	5	6
1	A	D	B	C	1	A	F	B	E	C	D
2	D	C	A	B	2	F	E	A	D	B	C
3	B	A	C	D	3	B	A	C	F	D	E
4	C	B	D	A	4	E	D	F	C	A	B
					5	C	B	D	A	E	F
					6	D	C	E	B	F	A

Designs of the above type for  $v = 4$  and  $6$  do not allow solutions for competition effects for each of the treatments under the following response model:

$$Y_{fghijkm} = \mu + \rho_f + \gamma_g + \tau_h + \alpha_{1i} + \alpha_{2j} + \alpha_{3k} + \alpha_{4m} + \epsilon_{fghijkm}, \quad (5.1)$$

where the first four terms and the last term are defined as for equation (2.5) and the  $\alpha_{1i}$ ,  $\alpha_{2j}$ ,  $\alpha_{3k}$ , and  $\alpha_{4m}$  denote the competition effects from the four adjacent e.u.s of the fghth e.u. When the e.u. is rectangular rather than square, the authors show how to change (5.1) to account for the unequal areas of adjacency. Also, they give a suggestion of how to handle edge effects of an experiment. For this first family of EDs, solutions for all competition effects are only available for all even  $v \geq 8$ . For  $v = 4$ , only one linear combination (i.e., only one non-zero eigenvalue for the competition effects matrix) of the  $\alpha$ 's has a solution; for  $v = 6$ , four linear combinations have solutions (four non-zero eigenvalues) (See Federer and Basford, 1991.). If the balance in rows (or columns) is changed, then a solution for all  $\alpha$ 's results. A second family of EDs may be constructed by repeating the last row of the designs from the first family of designs. Then, solutions for  $v = 4$  and  $6$  are now possible. The third family of designs may be constructed from a particular type of F-square constructed as follows for  $v = 3$  treatments:

$$F(6; 2, 2, 2) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \begin{bmatrix} A & B & C \\ B & C & A \\ C & A & B \end{bmatrix} = \begin{bmatrix} A & B & C & A & B & C \\ B & C & A & B & C & A \\ C & A & B & C & A & B \\ A & B & C & A & B & C \\ B & C & A & B & C & A \\ C & A & B & C & A & B \end{bmatrix}$$

where  $\otimes$  denotes a Kronecker product. Instead of the first matrix being  $2 \times 2$ , it may be any size not necessarily square (Federer, 1992, chapter 9). Then the rows and columns of the resulting F-square are permuted in the same manner as for the first family of EDs to obtain row and column balance for competition effects. For  $v = 3$ , the design is:

	column					
row	1	2	3	4	5	6
1	A	C	B	B	C	A
2	C	B	A	A	B	C
3	B	A	C	C	A	B
4	B	A	C	C	A	B
5	C	B	A	A	B	C
6	A	C	B	B	C	A

The competition effect matrix for  $v = 3$  treatments in the above design has only one non-zero eigenvalue but has  $v - 1$  non-zero eigenvalues for  $v \geq 4$ , i.e., solutions are possible for all competition effects. The above EDs are balanced for competition effects in both rows and columns. Note that all treatments precede and follow each other including themselves in both rows and columns an equal number of times.

For competition experiments, one could compute a single degree of freedom sum of squares corresponding to that for the largest eigenvalue from the competition effects matrix. This will be denoted as Kempton's one degree of freedom sum of squares for competition as the idea came from his papers. This one degree of freedom sum of squares could be computed for every experiment where competition might occur, as a diagnostic statistic for competition and could be used in much the same manner as Tukey's one degree of freedom for non-additivity. These two diagnostic statistics could be included in computer software packages and used on a routine basis as aids in detecting deviations from the usual linear model when using an ANOVA.

In animal experiments wherein animals are in the same pen or are born to the same litter, competition among animals for food and space is a fact of life and must be dealt with in the design and analysis of experiments. One procedure for doing this is to estimate a component of variance due to competition. Competition may change with litter or pen size; if it can be modeled as a function of litter or pen size, then a component of variance due to competition may be estimated (see Federer and Ladipo, 1978).

Another type of non-independence occurs when there are gradients in the responses within strata or blocks (type (iv) above). A serial correlation between adjacent e.u.s is introduced. This type of situation has received much attention in the literature under the name nearest neighbor (NN) design and analysis. NN analyses are designed to remove the effect of this serial correlation from the estimates of treatment effects. The idea has been around for many years and has recently been rejuvenated (See, e.g., Wilkinson et al., 1983, Stroup and Mulitze, 1991, and the list of references in these papers.). For a NN analysis, the serial order of the e.u.s in the investigation is taken into account. A response model of this form for a RCBED is:

$$Y_{ijk} = \mu + \tau_i + \rho_j + f_k + \epsilon_{ijk}, \quad (5.2)$$

where  $f_k = f_{k-1} + \sigma_f \phi_k$  for  $k = 2, \dots, n$ ,  $f_k = \sigma_f \phi_k$  for  $k = 1$  and  $k$  is the order within a block, and the other parameters are defined as in equation (1.7);  $\sigma_f$  is a constant and  $\phi_k$  are random deviations associated with deviations from serial order within a block. For a standard RCBED,  $n = v$ . First order differences between adjacent e.u.s within a block are computed as:

$$Y_{ij2} - Y_{ij1}, Y_{ij3} - Y_{ij2}, Y_{ij4} - Y_{ij3}, \dots, Y_{ijn} - Y_{ij(n-1)} \quad (5.3)$$

and second order differences are computed as:

$$Y_{ij2} - (Y_{ij1} + Y_{ij3}) / 2, Y_{ij3} - (Y_{ij2} + Y_{ij4}) / 2, \dots, Y_{ij(n-1)} - (Y_{ij(n-2)} + Y_{ijn}) / 2, \quad (5.4)$$

where k runs serially in each block. Using first order or both first and second order differences, a NN analysis is performed. Since blocking does not control gradients within a block, nearest neighbor analyses have been found useful in analyzing results from several types of experiments.

## 6. Study of Residuals

For any response model, the residuals may be computed and investigated for patterns, trends, outliers, or other types of behavior which would affect the assumptions involved in the use of the results in an ANOVA. For example, a serial correlation of the serially ordered residuals within each of r blocks could be computed with the associated sum of squares to obtain r single degree of freedom sums of squares for detecting gradients within blocks. If the gradients or trends were the same in every block, a single degree of freedom sum of squares could be computed as a single degree of freedom diagnostic test for trend. Such a test could also be included in computer packages for routine use in testing for trend.

Major advances in studying residuals have been made by J. W. Tukey and associates over the last 40 years. The terms modern data analysis, exploratory data analysis, and study of residuals are used to include many aspects of studying residuals from a response model. Two of several books by Tukey and associates are by Hoaglin, Mosteller, and Tukey (1985, 1985). The numerical example in Table 6.1 is presented to illustrate how a study of residuals pointed to a single outlier which greatly affected the interpretation of the experimental results from an ANOVA. The largest residual in the table is 2.461 / 16 which is associated with the observation 1.035. Note that this is an impossible result in that the dry weight measurement divided by the wet weight measurement on the same material must be less than or equal to one! As is often the situation, an investigator submits the data for statistical analysis using computer software and fails to note this discrepant result. A study of the original data for possible gross errors is usually not made, especially for large data sets. After computing the residuals, it may be noted that a pattern exists in that every residual in the same row or column as the observation 1.035 is negative while the one associated with the observation is a large positive residual. This is only one type of pattern that residuals may take. For this particular example, the experimenter would have used a residual mean square that was eight times too large, thus affecting significance tests and confidence intervals. Biological, medical, nutritional, and engineering researchers often use ratios as used here. It should be noted that such ratios often tend to eliminate differences among numerator or among denominator responses in that ratios tend to be a constant proportion. For this example, the

coefficient of variation after removing the effect of the outlier, was about three percent which is much smaller than for wet or dry weights analyzed separately.

H.C. Kirton (1984) suggested that an ANOVA be computed using the absolute values of the residuals and adjusting the “degrees of freedom” column so that they add up to  $(r - 1)(v - 1)$ , or  $[(r - 1)(v - 1) - k]$  if there are  $k$  missing plot values. The residuals using the computed missing plot value of 0.762 are given in Table 6.2. ANOVAs on the two sets of residuals are also presented. For the original values in Table 6.1, the “mean squares” for blocks and for treatments is about five times that for the remainder. When residuals are computed using the missing plot value 0.762 and computing an ANOVA on the absolute values of the residuals, the “mean squares” for blocks and for treatments is approximately equal to that for remainder. Of course, these ratios of “mean squares” do not follow an F distribution but since F is robust to non-normality, it can be considered to be a fair approximation. A study of the block and treatment of absolute values of the residuals in Table 6.1, indicates that the treatment early and block 3 means were much higher than the other means. This again would point to the discrepant observation 1.035.

As stated, there are many methods for investigating whether patterns, trends, and outliers occur in the residuals used to compute an error mean square for statistical analyses. Instead of studying this type of residual, the investigator may focus attention on the interaction terms from a two factor factorial and use some of the same methods. It is often desirable to model factorial responses with as few parameters as possible (parsimony). The interaction terms are treated as residuals. Bradu and Gabriel (1978) present a method known as bi-plot as a diagnostic tool in searching for an appropriate model. Their general results contain several of the procedures described previously such as Tukey’s one degree of freedom for non-additivity and Mandel’s procedure. They discuss bi-plotting using the original observations, deviations from the overall mean, and residuals or interaction terms. Gauch (1988) made use of their ideas to develop an additive main effects and multiplicative interaction (AMMI) model for two factor studies such as genotype and environment. The method has been successfully applied to a variety of experiments in agriculture. A principal components analysis is applied to the residuals (interactions). Often only the first and perhaps the second principal components are sufficient to model the response. Using a bi-plot aids in the interpretation of the data.

An AMMI response model for a RCBED with  $ab$  treatments in a two factor factorial with  $a$  levels of factor A and  $b$  levels of factor B is:

$$Y_{ijk} = \mu + \rho_i + \alpha_j + \beta_k + \sum_{h=1}^n \lambda_h \gamma_{hj} \delta_{hk} + \pi_{jk} + \epsilon_{ijk}, \quad (6.1)$$

where  $\mu$ ,  $\rho_i$ , and  $\epsilon_{ijk}$  are as defined for (1.7),  $\alpha_j$  is the additive effect of the  $j$ th level of factor A,  $\beta_k$  is the additive effect of the  $k$ th level factor B,  $\lambda_h$  is the singular value for interaction principal component  $h$ ,  $\gamma_{hj}$  and  $\delta_{hk}$  are the two factor eigenvectors for principal component  $h$ , and  $\pi_{jh}$  is the residual left

for interaction after fitting  $n$  principal components to the interaction terms  $\bar{y}_{.jk} - \bar{y}_{.j} - \bar{y}_{.k} + \bar{y}_{...} = \alpha\hat{\beta}_{jk}$ . Note that  $\hat{\alpha}_j = \bar{y}_{.j} - \bar{y}_{...}$  and  $\hat{\beta}_k = \bar{y}_{.k} - \bar{y}_{...}$  and the usual principal components analysis constraints are used, i.e.,

$$\sum_{j=1}^a \alpha_j = \sum_{k=1}^b \beta_k = 0, \quad \sum \gamma_j^2 = \sum \delta_k^2 = 1,$$

and every eigenvector is constrained to be orthogonal to all previous eigenvectors, so that for  $h \neq h'$

$$\sum_{j=1}^a \gamma_{hj} \gamma_{h'j} = \sum_{k=1}^b \delta_{hk} \delta_{h'k} = 0.$$

The maximum number of principal components to be fitted is the minimum of  $(a - 1)$  and  $(b - 1)$ . For any data set, zero to  $\min \{a - 1, b - 1\}$ , principal components will be fitted in the AMMI model family, i.e., AMMI0, AMMI1, AMMI2, ..., AMMIF,  $F = \min \{a - 1, b - 1\}$ . AMMI0 corresponds to the no interaction case and AMMIF corresponds to a consideration of the means  $\bar{y}_{.jk}$  and comparisons among these  $ab$  means, e.g. multiple comparisons. The eigenvalue for component  $h$  is  $r \lambda_h^2$ . There is a controversy in the literature about the number of degrees of freedom to assign to the sum of squares associated with each principal component. Gauch (1992) appears to have resolved this dilemma.

The steps in the AMMI model analysis for two-factor factorials are

- (i) compute  $\alpha\hat{\beta}_{jk}$ ,
- (ii) select  $\min \{a - 1, b - 1\}$  to determine which levels of factors A and B are to be used as variates,
- (iii) fit a principal components analysis,
- (iv) let  $n$ , usually only  $h = 1$  or  $h = 2$ , be the number of principal components to be used as determined by having the residual interaction mean square approximately equal to the error mean square,
- (v) obtain the estimated  $jk^{\text{th}}$  cell means as  $\hat{y}_{.jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + \sum_{h=1}^n \lambda_h \gamma_{hj} \delta_{hk}$ , and
- (vi) prepare a bi-plot of the two-way array of the values from (v) using the adjusted values  $\hat{y}_{.j}$  and  $\hat{y}_{.k}$  as the abscissas for the plot and the  $\gamma_{hj}$  and  $\delta_{hk}$  values as the ordinates for one bi-plot

(The name bi-plot is used to denote that two sets of bivariate,  $\hat{y}_{.j}$  and  $\gamma_{hj}$  and  $\hat{y}_{.k}$  and  $\delta_{hk}$ , are plotted on the same graph.). A second bi-plot of  $\gamma_{1j}\sqrt{\lambda_1}$  values against  $\gamma_{2j}\sqrt{\lambda_2}$  values is also used to aid in interpreting results from factorial experiments; the  $\delta_{1k}\sqrt{\lambda_1}$  values are plotted against the  $\delta_{2k}\sqrt{\lambda_2}$  values on the same bi-plot. Patterns in this latter bi-plot are indicative of particular types of models for the interaction terms. Plots of this nature can be most helpful in the interpretation of results from an experiment.

Table 6.1. Wheat yields of dry/ wet grain weight from a RCBED with four nitrogen treatments and four blocks.

Block	nitrogen applied				Total Y <sub>.j</sub>
	none	early	middle	late	
1	.718	.732	.734	.792	2.976
2	.725	.781	.725	.716	2.947
3	.704	1.035	.763	.758	3.260
4	.726	.765	.738	.781	3.010
Total	2.873	3.313	2.960	3.047	12.193

16 Residuals ( $Y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$ )

Block	none	early	middle	late	Total	Sum of
						absolute values
1	.285	-1.251	.193	.773	0	0.156375
2	.513	-.351	.165	-.327	0	0.084750
3	-1.075	2.461	-.479	-.907	0	0.307625
4	.277	-.859	.121	.461	0	0.107375
Total	0	0	0	0	0	-
Sum of absolute values	0.134375	0.307625	0.059875	0.154250	-	0.656125

ANOVA

Source of variation	df	sum of squares	mean suares
Total	16	9.381715	
Correction for mean	1	9.291828	
Block	3	.015443	.005147
Nitrogen	3	.027149	.009050
Block × Nitrogen	9	.047295	.005255
Outlier suspect	1	.042059	.042059
Remainder	8	.005236	.000655

$$CV = \sqrt{.000655} / (11.920) / 16 = 3.4\%$$

Table 6.2. Missing plot value 0.762 inserted for discrepant value 1.035 of Table 6.1.

ANOVA on  $|\hat{\epsilon}_{ij}|$  values of Table 6.1 as per H.C. Kirton

Source of variation	“degrees of freedom”*	sum of squares	“mean square”
Total	$(r-1)(v-1) = 9$	0.047295	0.005255
Mean	$c = 9/16$	0.026906	-
Blocks	$c(r-1) = 27/16$	0.007543	0.004470
Treatments	$c(v-1) = 27/16$	0.008111	0.004807
Remainder	$c(r-1)(v-1) = 81/16$	0.004735	0.000935

\* $c = (r-1)(v-1)/rv = 9/16$ .

Residuals times 16 using 0.762 in place of 1.035.

Block	Treatment				Sum of absolute residuals
	none	early	middle	late	
1	.012	-.432	-.080	.500	.0640
2	.240	.468	-.108	-.600	.0885
3	-.256	.004*	.340	-.088	.0430
4	.004	-.040	-.152	.188	.0240
Sum of absolute residuals	.0320	.0590	.0425	.0860	.2195

\*zero within rounding error on 0.762

ANOVA on absolute values of residuals in above Table as per H.C. Kirton

Source of variation	“degrees of freedom”*	sum of squares	“mean square”
Total	$(r-1)(v-1)-1 = 8$	0.005236	0.000655
Mean	$c' = 8/15$	0.003011	-
Blocks	$c'(r-1) = 24/15$	0.000577	0.000361
Treatments	$c'(v-1) = 24/15$	0.000416	0.000260
Remainder	$c'(rv-r-v) = 64/15$	0.001232	0.000289

\* $c' = [(r-1)(v-1)-1]/(rv-1) = 8/15$ .

## 7. Post Mortems on an Investigation

In light of the various procedures employed in the analysis of the results of an investigation, the investigator makes a decision on the model and the particular statistical analysis to use for the data. In addition, the nature of the observations, the manner in which they were obtained, and the theoretical background for such observations should be studied carefully. From a study of numerous data sets, it has been found that nearly always there is something peculiar about the results, about some or all of the observations, or about the way in which the observations were obtained. There usually is "something wrong" somewhere with a data set. Thus, a healthy attitude for an approach to analyzing any data set is to critically examine all aspects of the investigation prior to performing any statistical computations. Then, after all analyses have been made and conclusions drawn, a post mortem diagnosis of the entire procedure is in order. Unless the results and conclusions are repeatable by other researchers in the field, the results of an investigation are not of much use.

Another situation in experimentation is the occurrence of unequal sample sizes. Any of several books on linear models provide statistical procedures for unbalanced data. There appears, however, to be something missing in their descriptions in that all analyses presented are conditional upon the particular sample size configuration that resulted for this experiment. If, however, it can be demonstrated that treatment response is not a function of sample patterns in an experiment, then the analyses are also unconditional. One place where these analyses have been used extensively is in animal breeding studies. Here sample size is often related to treatment response, e.g., the best bulls always get the most cows!

Examples IV-1 and IV-2 in Federer (1955) are illustrations of what can go wrong in performing a statistical analysis on a set of data. The first example was purported to be a CRED whereas it most likely was a single replicate of a RCBED with multiple sampling units in each experimental unit. This type of mistaken ED is of frequent occurrence in published literature (See Federer, 1975). The second example had random sample sizes for the three types of plant. Here sample size is most likely related to treatment in that the germination and survival rate for off-types is lower than for the other two types.

## 8. Discussion

Some general considerations for checking on the assumptions behind a statistical procedure are discussed. Several more could have been included. Several of the procedures described herein have as their goal the partitioning of the residual (block  $\times$  treatment, for example) sum of squares into a part attributable to an effect such as non-additivity and a part due to error variation. The Tukey one degree of freedom sum of squares for non-additivity is an example of this as this sum of squares is removed from the block  $\times$  treatment sum of squares to obtain the estimated error mean square. The

estimated treatment effects are unaffected whether or not the blocks  $\times$  treatment sum of squares is partitioned. Likewise, the interaction sum of squares for a two factor factorial may be partitioned into two parts, one part to explain the interaction and a second part representing the estimated noise or error mean square. The AMMI procedure is an example of this kind of partitioning. The factor effects are considered to be additive and unaffected by the manner in which the interaction sum of squares is partitioned.

Other procedures discussed above have as their goal the adjustment of treatment effects for other effects. When competition effects between adjacent e.u.s are present, it is desirable to adjust direct treatment effects for effects of competition of adjacent units. Direct effects of treatments are adjusted for residual effects of previous treatments in repeated measures experiments. Nearest neighbor analyses have as their goal the adjustment of treatment effects for local gradients within blocks. The two different goals described above are not incompatible, and both may be desired when analyzing the results from an experiment. The extraction of all available information from an experiment should be the ultimate target for any data analyst.

## 9. References Cited

- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society (London), Series A*, 160:268-282.
- Besag, J. and R. A. Kempton (1986). Statistical analysis of field experiments using neighboring plots. *Biometrics* 42:231-251.
- Box, G. E. P. (1953). Non-normality and test on variances. *Biometrika* 40:318-335.
- Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A*, 143:383-430.
- Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211-252.
- Bradu, D. and K. R. Gabriel (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20:47-68.
- Cochran, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. *Annals of Eugenics* 11:47-52.

- Cox, C. P. (1958). The analysis of Latin square designs with individual curvatures in one direction. *Journal of the Royal Statistical Society, Series B*, 20:193-204.
- Federer, W.T. (1955) *Experimental Design - Theory and Application*. Macmillan Co., New York (Republished by Oxford & IBH Publishing Co., New Delhi in 1967 and in 1974).
- Federer, W. T. (1970). An application of three tests for nonadditivity. BU-343-M in the Technical Report Series of the Biometrics Unit, Cornell University, October.
- Federer, W.T. (1975). The misunderstood split plot. In *Applied Statistics* (R.P. Gupta, editor), North Holland Publishing Co., Amsterdam, Oxford.
- 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 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. *Biometrical Journal* 19(3):181-200.
- Federer, W. T. (1979). Statistical designs and response models for mixtures of cultivars. *Agronomy Journal* 71:701-706.
- Federer, W. T. (1979). Response equations and data analysis for some simple experiment designs. BU-676-M in the Technical Report Series of the Biometrics Unit, Cornell University, June.
- Federer, W. T. (1991). *Statistics and Society. Data Collection and Interpretation*, second edition. Marcel Dekker, Inc. New York.
- Federer, W. T. (1992). *Statistical Design and Analysis for Intercropping Experiments*. Springer Verlag, New York
- Federer, W. T. and K. E. Basford (1991). Competing effects designs and response models for two-dimensional field arrangements. *Biometrics* 47:1461-1472,

- Federer, W. T. and O. O. Ladipo (1978). A component of variance due to competition. *Australian Journal of Statistics* 20(2): 99-110.
- Federer, W. T. and D. Raghavarao (1987). Response models and minimal designs for mixtures of  $n$  of  $m$  items useful for intercropping and other investigations. *Biometrika* 74: 571-577.
- Federer, W. T. and C. S. Schlottfeldt (1954). The use of covariance to control gradients in experiments. *Biometrics* 10:282-290.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* lii:399-433.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*, first edition. Oliver and Boyd, Edinburgh.
- Fisher, R. A. (1935). *The Design of Experiments*, first edition. Oliver & Boyd, Edinburgh
- Fisher, R. A. and W. A. Mackenzie (1923). Studies in crop variation II. The manurial response of different potato varieties. *Journal of Agricultural Science* 13:311-320.
- Gauch, H. G., Jr. (1988). Model selection and validation for yield trials with interaction. *Biometrics* 44:705-715.
- Gauch, H. G., Jr. (1992). Degrees of freedom in the Mandel model for two-way tables. (In the process of publication).
- Griffing, B. (1956). Concept of general and specific ability in relation to diallel crossing systems. *Australian Journal of Biological Science* 9:463-493.
- Grimes, B. A. and W. T. Federer (1984). Comparison of means from populations with unequal variances. In *W. G. Cochran's Impact on Statistics* (Editors: P. S. R. S. Rao and J. Sedransk), John Wiley and Sons, Inc., New York, pp. 353-374.
- Harris, J. A. (1913). On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika* 9:446-472.

- Hartley, H. O. (1950). The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37:308-312.
- Hoaglin, D. C., F. Mosteller, and J. W. Tukey (1985). *Fundamentals of Exploratory Analyses*. John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore.
- Hoaglin, D. C., F. Mosteller, and J. W. Tukey (1985). *Exploring Data Tables, Trends, and Shapes*. John Wiley & Sons, Inc., New York Chichester, Brisbane, Toronto, Singapore.
- Kempthorne, O. (1952). *Design and Analysis of Experiments*. John Wiley & Sons, Inc., New York.
- Kempton, R. A. (1982). Adjustment for competition between varieties in plant breeding trials. *Journal of Agricultural Science* 98:599-611.
- Kempton, R.A. (1984). The use of biplots in interpreting variety by environment interactions. *Journal of Agricultural Science, Cambridge*, 3, 123-135.
- Kershner, R. P. and W. T. Federer (1981). Two-treatment cross-over designs for estimating a variety of effects. *Journal of the American Statistical Association* 76:612-619.
- Kirton, H. C. (1984). Personal communication.
- Mandel, J. (1961). Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association* 56:878-888.
- Mandel J. (1971). A new analysis of variance model for non-additive data. *Technometrics* 13: 1-18.
- Martin, K. J. (1980). A partition of a two-factor interaction, with an agricultural example. *Applied Statistics* 29:149-155.
- Outhwaite, A. D. and A. Rutherford (1955). Covariance analysis as an alternative to stratification in the control of gradients. *Biometrics* 11:431-440.
- Pearson, E. S. and H. O. Hartley (1962). *Biometrika Tables for Statisticians*, 2nd edition, volume 1. Cambridge University Press, London.

- Robson, D. S. (1970). Tests for non-additivity viewed as tests of the hypothesis of no interaction. A preliminary report. BU-334-M in the Technical Report Series of the Biometrics Unit, Cornell University, October.
- Scheffé, H. (1959). The Analysis of Variance. John Wiley & Sons, Inc. New York.
- Snedecor, G. W. and W. G. Cochran (1980). Statistical Methods, seventh edition. Iowa State University Press, Ames, Iowa.
- Sprague, G. F. and L. A. Tatum (1941). General vs. specific combining ability in single crosses of corn. Journal of the American Society of Agronomy 34:923-932.
- Stroup, W. W. and D. K. Mulitze (1991). Nearest neighbor adjusted best linear unbiased prediction. The American Statistician 45(3): 194-200.
- Tukey, J. W. (1949). One degree of freedom for non-additivity. Biometrics 5:232-242.
- Votaw, D. F., Jr. (1948). Testing compound symmetry in a normal multivariate distribution. Annals of Mathematical Statistics 19:447-473.
- Wilkinson, G. N., S. R. Eckert, T. W. Hancock, and O. Mayo (1983). Nearest neighbour (NN) analysis of field experiments (with discussion). Journal of the Royal Statistical Society, Series B 44:151-211.