OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

WHAT THE TEXTBOOKS SAY ABOUT THE DESIGN OF EXPERIMENTS

by

Bruce W. Lamar

OR 073-78 March 1978

Prepared under Grant Number 78NI-AX-0007 from the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Department of Justice.

TABLE OF CONTENTS

ABSTRACT

This report reviews classical experimental designs including single and multiple factor analysis of variance, analysis of covariance, and Latin squares designs. Assumptions used in the models are presented, and tests for violations of the assumptions are described. Examples illustrating primary designs and remarks discussing further model extensions and considerations are also included.

1 INTRODUCTION

Evaluations of criminal justice systems frequently involve the testing of alternative programs or treatments. Foremost among the questions that may be asked in such evaluations is: "Did the treatments make any difference?" To answer this question effectively, an organized statistical plan, that is an experimental design, must be developed and implemented.

This report is a review of the textbook material relating to experimental designs. The Selected Bibliography contained at the end of this report lists a few of the plethora of mathematical statistics and specialized books available in the M.I.T. libraries on the subject. Such books range from the quite descriptive (Chapin) to the quite mathematical (Winer).

There are two main sections into which this report is organized. The first section reviews the fundamental experimental designs in which all the assumptions in implementing the model are satisfied. The other section reviews the procedures undertaken when the basic assumptions are violated. This latter section includes procedures for testing for violations as well as alternative designs that may be implemented when the assumptions are not satisfied or when policy decisions cause a change in the experimental environment during the course of the experiment.

2 FUNDAMENTAL EXPERIMENTAL DESIGNS

The most extensively employed technique used in experimental designs is the analysis of variance (ANOVA) which tests whether or not there is variation in the treatments under consideration by assigning the variations observed in experimental data to known sources (Ferguson, p. 223). In the experiment, observations or measurements are made on experimental units which are subjected to the various treatments. The experimental units may be individuals, police squads, townships, or the like (Neter, p. 674).

This section introduces the ANOVA techniques by first summarizing the basic assumptions involved and then by applying ANOVA to several fundamental designs.

2.1 Assumptions

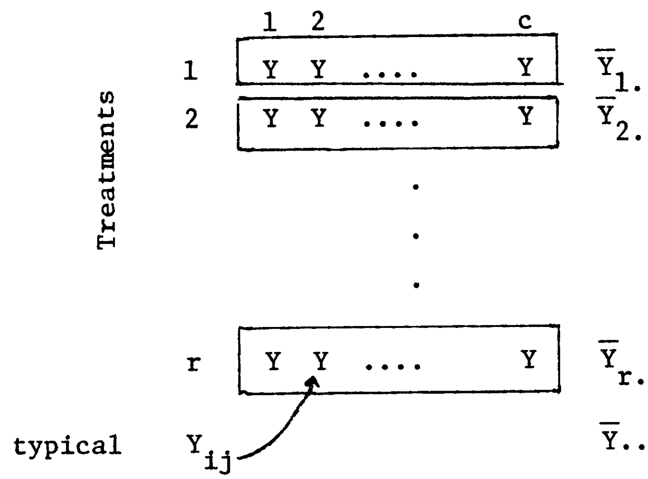
The assumptions underlying the fundamental ANOVA models described in this section are as follows (Kirk pp. 102-103; Neter, p. 426):

- The experimental errors within each treatment population are normally distributed.
- The experimental errors within each treatment population have the same variance.
- Each observation may be represented as a linear combination of terms.
- The treatments are randomly assigned to experimental units to ensure independence between observations.

2.2 Single Factor Design

Most fundamental of the fundamental designs is the single factor design which tests only for differences among treatments. The experimental layout is shown in Figure 2-1.

Figure 2-1



where: Y_{ij} is experimental observations

$\bar{Y}_{i.}$ is treatment mean

$\bar{Y}_{..}$ is overall mean

The model under consideration is

Eqn 2-1 $$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \text{with} \quad \sum_i \alpha = 0$$

where: Y_{ij} is the observation of experimental unit j under treatment i;

μ is the overall mean;

α_i is the deviation from the mean due to the treatment i;

ε_{ij} is the random error distributed $N(0,\sigma^2)$.

and is used to test the hypothesis

Eqn 2-2 $$H_o: \quad \alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$$

H_1: otherwise.

To test this hypothesis two equivalent approaches may be employed.

2.2.1 First Approach

The first approach (Dixon, pp. 147-148; Hoel pp. 289-290) starts by noting that Y_{ij} is distributed $N(\mu + \alpha_i, \sigma^2)$ since ε_{ij} is distributed $N(0, \sigma^2)$ and μ and α_i are parameters (constants).' Since Y_{ij} is distributed

Normally, the within treatment sample variance $s_i^2 = 1/c\Sigma(Y_{ij} - \overline{Y}_{i.})^2$ provides the ratio cs_i^2/σ^2 which is chi-square distributed with c-1 degrees of freedom. The sum of these values,

$$k_1 = \frac{c(s_1^2 + \ldots + s_r^2)}{\sigma^2}$$

is also chi-squared distributed with $d_1 = r(c - 1)$ degrees of freedom.

In addition, since Y_{ij} is Normal,this implies that $\overline{Y}_{i.}$ is distributed $N(\mu + \alpha_i, \sigma^2/r)$. So in a similar fashion

$$k_2 = \frac{rs^2_{\overline{Y}_{i.}}}{\sigma^2/c}$$

is also chi-square distributed with $d_2 = (r - 1)$ degrees of freedom. Taking the ratio of k_2 to k_1 divided by their respective degrees of freedom, we achieve the formula

$$F = \frac{k_2/d_2}{k_1/d_1}$$

which is F-distributed with d_2, d_1 degrees of freedom. By noting that k_2 should be small if H_o is true, we have our means of testing H_o, namely to reject H_o if F is too large.

2.2.2 Second Approach

The second approach (Ferguson pp. 226-228; Neter pp. 436-441; Winer pp. 152-155) is developed by first observing the deviation of sample values about the estimate of the mean via the following identity:

$$Y_{ij} - \overline{Y}_{..} = Y_{ij} - \overline{Y}_{i.} + \overline{Y}_{i.} - \overline{Y}_{..}$$

Then by squaring both sides and summing over i and j we obtain:

Eqn 2-3

$$\sum_{i,j}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i,j}(Y_{ij} - \overline{Y}_{i.} + \overline{Y}_{i.} - \overline{Y}_{..})^2$$

$$= \sum_{i,j}(\overline{Y}_{i.} - \overline{Y}_{..})^2 + 2\sum_{i,j}(\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.})$$

$$+ \sum_{i,j}(Y_{ij} - \overline{Y}_{i.})^2$$

$$= \sum_{i,j}(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i,j}(Y_{ij} - \overline{Y}_{i.})^2$$

where $2\sum_{i,j}(\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) = 0$ since it is a sum of deviations about the mean.

Equation 2-3 may be interpreted as the total variation (SS_y) equaling the variation between rows (i.e., treatments) (SSR_y) plus the unexplained (i.e., residual) variation within treatments (SSU_y). That is,

$$SS_y = SSR_y + SSU_y.$$

Dividing SSU_y by its appropriate degrees of freedom $[d_1 = r(c - 1)]$ we arrive at the mean square of residuals (MSU_y) which is an unbiased estimate of σ^2. Similarly, the mean square of treatments (MSR_y) is obtained by dividing SSR_y by its degrees of freedom ($d_2 = r - 1$). MSR_y is an unbiased estimate of σ^2 if H_o is true; else $E(MSR_y) > \sigma^2$. Hence, we again arrive at the ratio

$$F = \frac{MSR_y}{MSU_y}$$

which is F-distributed with d_2, d_1 degrees of freedom. Again, we reject H_o if F is too large.

2.2.3 Example

Consider an experiment in which three different dispatching methods (treatments) are randomly assigned to police officers (experimental units) and the response times (observations) are measured. Typical data for this experiment is shown in Figure 2-2.

Figure 2-2

Experimental Units

Treatments	1	2	3	4	$\bar{Y}_{i.}$
1	50	46	39	47	45.5
2	50	58	55	50	53.25
3	55	57	46	54	53.0
$\bar{Y}_{.j}$	51.67	53.67	46.67	50.33	$\bar{Y}_{..}$ = 50.5833

To test whether or not differences between treatments exist the ANOVA calculations are compiled in an ANOVA Table such as in the one below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Statistic
Row (treatment)	SSR_y = 155.167	r-1 = 2	MSR_y = 75.58	F = 3.84
Unexplained	SSU_y = 181.75	r(c-1)=9	MSU_y = 20.19	---
Total	SS_y = 336.91	rc-1 =11	---	---

Since at the 95% level the critical F value is 4.26 (Hoel, p. 395), then F = 3.84 < 4.26 implies that we accept H_o and infer that no difference between treatments exists.

2.2.4 Remarks

1) In order to aid in comparison and validity of experimental results, one of the treatments is frequently a control group (Campbell, p. 13).

2) Instead of absolute measurements, the difference between pretreatment and post-treatment measurement may be used. This helps to eliminate external effects and so increases the internal validity of the model but makes it less generalizable to situations without pretreatment measurements and so decreases the model's external validity (Campbell, p.25).

3) "Tea for two." If just two treatments are under consideration, then the assumptions described in Section 2.1 equivalently allow for a

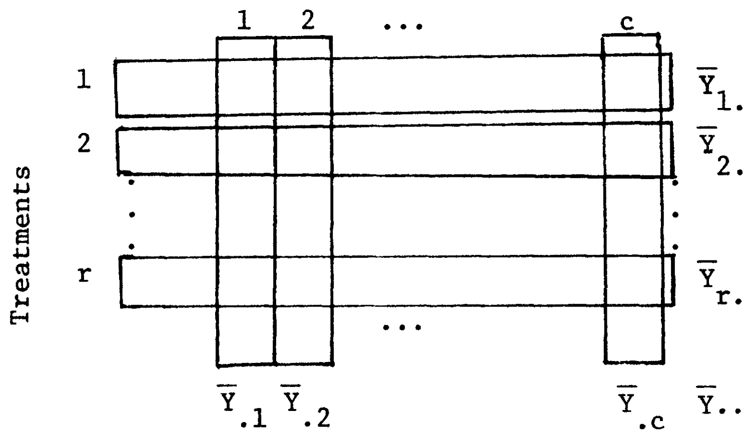
T-test between two means to be employed (Chapin, p. 197).

2.3 Two Factors Design

There are many extensions that may be made to the single factor design. One such extension is the two factor design which takes into account variations in both treatments and experimental units. The experimental layout is shown in Figure 2-3 which is identical to Figure 2-1 save for the inclusion of the experimental unit means.

Figure 2-3

Experimental Units



where: Y_{ij}, $\bar{Y}_{i.}$, $\bar{Y}_{..}$ are as described in Figure 2-1

$\bar{Y}_{.j}$ is the experimental unit mean.

The model of Eqn 2-1 is extended in the two factor design to

Eqn 2-4 $$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \text{with } \sum_i \alpha_i = 0, \sum_j \beta_j = 0$$

where: Y_{ij}, μ, ε_{ij} are as described in Eqn 2-1

β_j is the deviation from the mean due to experimental units.

Here, in addition to the hypothesis, "Is there a difference in treatments?." given in Eqn 2-2, the model also simultaneously tests the hypothesis, "Is there a difference in experimental units?." in the following form:

H_o: $\beta_1 = \beta_2 = \dots = \beta_r = 0$

H_1: Otherwise

To test each of these hypotheses the procedure for partitioning the sum of squares is utilized. The equation

$$\sum_{i,j}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i,j}(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i,j}(\overline{Y}_{.j} - \overline{Y}_{..})^2 + \sum_{i,j}(Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..})^2$$

may be rewritten using the acronyms

Eqn 2-5 $$SS_y = SSR_y + SSC_y + SSU_y$$

where: SS_y is the total variation

SSR_y is the variation between rows (treatments)

SSC_y is the variation between columns (experimental units)

SSU_y is the unexplained variation.

As before, the sum of squares divided by their respective degrees of freedom provide the mean squares (MSR_y, MSC_y, MSU_y) as estimates of σ^2. The statistic

$$F_1 = \frac{MSR_y}{MSU_y}$$

tests for treatment effects, while

$$F_2 = \frac{MSC_y}{MSU_y}$$

tests for experimental unit effects.

2.3.1 Example

Consider again the data in Figure 2-2. The ANOVA Table incorporating experimental unit effects is given in the table below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Statistics
Row(treatment)	SSR_y = 155.167	r-1 = 2	MSR_y = 77.58	F_1 = 4.49
Column(exp.unit)	SSC_y = 775.99	c-1 = 3	MSC_y = 258.66	F_2 = 14.99
Unexplained	SSU_y = 103.47	(r-1)(c-1) = 6	MSU_y = 17.25	---
Total	SS_y = 1034.64	rc - 1 =11	---	---

Comparing F_1 with its critical value (5.14), we decide to accept H_o (i.e., no row effect). Comparing F_2 with its critical value (4.76), we decide to reject H_o (i.e., column effect exists).

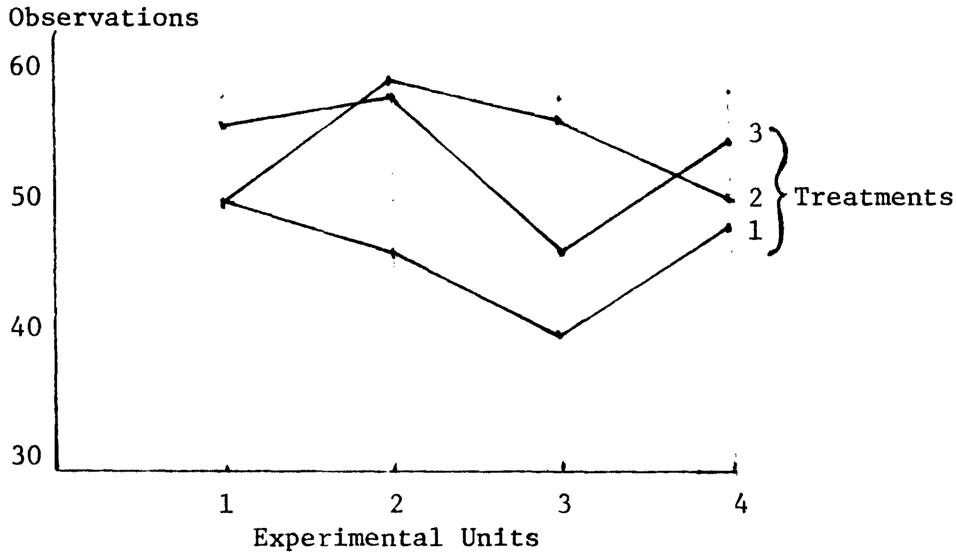
2.3.2 Remark

If each of the treatments is assigned to an experimental unit, then observations may not be independent as required by the randomness assumption in Section 2.1. In order to control for this dependence, a block design as described in Section 3.6 may be required (Neter, p. 429).

2.4 Two Factors with Interaction Design

Another possible design extension is the inclusion of interaction terms in the model. To illustrate the effect of interaction, consider the data of Figure 2-2 which is graphed in Figure 2-4.

Figure 2-4



Note that the differences between treatments varies with the experimental units. This variance implies interaction between treatments and experimental units. Had the lines in Figure 2-4 been mutually parallel then no interaction between factors would have been present (Campbell, p. 27-29).

The presence of interaction may be tested by having multiple observations per treatment/experimental unit cell and using the model [Neter, p. 568]:

Eqn 2-6 $$Y_{ij\ell} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij\ell}$$

with $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, $\sum_i (\alpha\beta)_{ij} = 0$, $\sum_j (\alpha\beta)_{ij} = 0$

where: μ, α_i, β_j, are as described in Eqn 2-4

$Y_{ij\ell}$ is the ℓ^{th} replication of observations of experimental unit j under treatment i;

$(\alpha\beta)_{ij}$ is the deviation from the mean due to the interaction between experimental unit j and treatment i;

$\varepsilon_{ij}\ell$ is the random error term.

Hypotheses about differences in the treatment, experimental unit, and interaction means may be tested in the usual manner by partitioning the sum of squares in the form

$$SS_y = SSR_y + SSC_y + SSRC_y + SSU_y$$

where SS_y, SSR_y, SSC_y, SSU_y are as described in Eqn 2-5

$SSRC_y$ is the interaction variation.

The F-statistics to test for treatment, experimental unit, and interaction effects are, respectively,

Eqn 2-7 $$F_1 = \frac{MSR_y}{MSU_y}, \quad F_2 = \frac{MSC_y}{MSU_y}, \quad F_3 = \frac{MSRC_y}{MSU_y}.$$

2.4.1. Example

If the data in Figure 2-2 is supplemented with a second observation in each cell, as indicated in Figure 2-5, then ANOVA may also include

Figure 2-5

First set of observations

Experimental Units

Treatments	1	2	3	4
1	50	46	39	47
2	50	58	55	50
3	55	57	46	54

Second set of observations

Experimental Units

Treatments	1	2	3	4
1	52	43	36	50
2	57	57	44	62
3	45	44	42	56

interactions as is the case in the table below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Statistics
Row (treatment)	$SSR_y = 306.33$	$r-1 = 2$	$MSR_y = 153.17$	$F_1 = 5.78$
Column (exp.unit)	$SSC_y = 317.46$	$c-1 = 3$	$MSC_y = 105.82$	$F_2 = 4.0$
Layer (interaction)	$SSRC_y = 86.67$	$(r-1)(c-1) = 6$	$MSRC_y = 14.44$	$F_3 = 0.54$
Unexplained	$SSU_y = 316.5$	$rc(h-1) = 12$	$MSU_y = 26.46$	---
Total	$SS_y = 1027.95$	$rch-1 = 23$	---	---

The conclusions to be drawn from the table above are summarized as follows:

F-Statistic	Critical Value	Conclusion
$F_1 = 5.78$	3.88	row effect
$F_2 = 4.0$	3.49	column effect
$F_3 = 0.54$	2.85	no interaction effect

2.5 Additional Design Considerations

Other possible fundamental experimental design considerations are briefly reviewed in this section.

2.5.1 General Model

Additional factor and interaction term effects may be added to the model. For example, the three factor with interaction model has the form

$$Y_{ijk\ell} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk\ell}$$

where terms are defined analogously to Eqn 2-6.

2.5.2 Unequal Sample Sizes

In the designs presented so far in Section 2, the number of replications of observations for each combination of factors has been assumed to be equal. However, in real life applications, this situation may not be the case.

The simplest case of unequal sample sizes is where the number of replications between any two factors is proportional. For example, in the two factor case in Section 2.4, the number of replications will be proportional if

$$h_{ij} = \frac{(\sum_i h_{ij})(\sum_j h_{ij})}{\sum_{ij} h_{ij}}$$

where: h_{ij} is the number of replications of treatment i and experimental unit j.

Here, ordinary ANOVA may be performed by simply weighing observations by their sample size (Neter, p.613; Winer, p.212)

In the case where unequal sample sizes are not proportional then ordinary ANOVA is not appropriate since the sum of square variations is not orthogonal and does not add to the total sum of squares (SS_y). Instead, an approximate ANOVA technique, the Method of Unweighted Means, may be used. Here, replications in each factor cell are averaged and this average value is used in the ANOVA calculations rather than the observations themselves (Neter, pp. 614-615; Winer, pp. 402-404).

2.5.3 Analysis of Factor Effects

The initial question asked in the design of experiments is: "Did the treatments make any difference?" If the answer is "yes", then the next question is "How much difference did the treatments make?" In other words, a ranking of the treatments is required.

There are numerous tests which make multiple comparisons among treatment or other factor means, most of which use the unbiased estimate of variance (MSU_y) to develop a T-distributed statistic (Winer, p. 185). One such test is the Tukey Honesty Significant Difference Test which makes comparisons between all pairs of factor means (Kirk, pp. 88-90; Lee, pp.300-301; Neter, pp.473-477). A second test is the Scheffé S Test which allows any number of factor means to be compared simultaneously (Kirk, pp.90-91; Lee, pp.301-302; Neter, pp.477-480). Finally, a third test, the Newman-Keuls Test, compares selected pairs of factor means in a stepwise manner (Kirk, pp.91-92; Lee, pp. 302-304).

2.5.4 Fixed, Random, and Mixed Models

In some cases, all of the factors under consideration are tested directly in the experiment. Such experiments, like the ones described in Section 2, are referred to as Fixed Model experiments. However, in other cases, only a random sample of factors, e.g., five police squads out of 60, are selected for testing in the experiment and then inferences to the rest of the factor population is made. This form of experiment is called a Random Model experiment (Campbell, p.31).

The Random Model has the same form as the Fixed Model, but a different interpretation is placed upon the terms. For example, in Eqn 2-6 α_i, β_j, and $(\alpha\beta)_{ij}$ are no longer fixed parameters but are random variables sampled from the factor population. This results in alternative calculations of the F-statistics. Specifically, the calculations in Eqn 2-7 are replaced (Neter, p.623) by

$$F_1 = \frac{MSR_y}{MSCR_y} \quad , \quad F_2 = \frac{MSU_y}{MSRC_y} \quad , \quad F_3 = \frac{MSRC_y}{MSU_y}$$

Lastly, an experiment which contains factors, some of which are random and some of which are fixed, is referred to as a Mixed Model experiment.

3 VIOLATIONS OF ASSUMPTIONS

In criminal justice evaluations, along with most investigations of social science behavior, it is not always possible to comply with all the conditions assumed present in employing a mathematical model. Therefore, it is important to be able to judge what effect a violated assumption will have on the overall validity of the model results.

The assumptions under which the designs in Section 2 were developed are summarized below:

- Experimental errors are Normally distributed
- Experimental errors have the same variance
- Observations are represented by a linear combination of terms
- Treatments are randomly assigned to experimental units.

This section considers the robustness of the ANOVA designs with respect to each of these assumptions. Tests for compliance with the assumptions as well as procedures to control for violations are also considered.

3.1 Normal Distribution Variations

Inherent in the formulation of the ANOVA model was the assumption that each observation was sampled from a Normal distribution. Fortunately, unless a departure from Normality is very extreme in either skewness or kurtosis, it will have little effect on the probability associated with the F-test of significance (Kirk, p.61; Neter, p. 513). Of the two, the F-test is less sensitive to skewness than to kurtosis (flatness or peakedness) of the distribution.

To test for Normality, standard tests such as the chi-square and Kulmorgorov-Smirnov tests may be employed. Alternatively, tests that do not require the estimation of distribution parameters (mean and variance) such as the Shapiro and Wilk W Test (Anderson, p.25) may be used.

If, indeed, the population distributions are far from Normal, then two options are available to circumvent this difficulty. The first option is to transform the data into a form that exhibits Normal behavior by the techniques described in Section 3.3. The second option is to abandon the F-statistic and its Normal dependency in favor of nonparametric statistics such as the median or the Kruskal-Wallis ank statistic (Neter, pp. 520, 522; Winer, pp. 848-849).

3.2 Unequal Variances

The equality of variances is another basic assumption in the designs of Section 2. However, like violations of Normality, the ANOVA model is quite robust to violations of the equal variance assumption (Kirk, p.61; Neter, p.514). Nevertheless, for some of the extended designs described in Section 2.5 (specifically, the designs encompassing unequal sample sizes and random effects), the effect of unequal variances becomes more pronounced and can result in misguided inferences from the F-test.

There are several methods available for testing the equality of variances among sample observations. One set of tests, such as the Bartlett test and the Bartlett-Kendall test, uses $\ell n\ S_i^2$ where S_i^2 is the sample variance of treatment i (Anderson, pp.20-21; Dixon, p.179) These tests capitalize on the fact that $\ell n\ S_i^2$ is approximately Normally distributed so ordinary ANOVA may be applied to the $\ell n\ S_i^2$ values themselves. A second set of tests, such as the Hartley test and Cochran test, use ratios of $\max(S_i^2)$ and $\min(S_i^2)$ to test for the equivalence of variances (Dixon pp.180,181; Kirk p.62; Neter p.512).

A third set of tests, such as the Burr-Foster Q-test, is derived from the sum of the sample variances squared (Anderson, p.22).

The main technique for equalizing variances is to transform the data via techniques described in Section 3.3.

3.3 Nonadditive Terms

The models presented in Section 2, such as Eqn 2-1, were a summation of component terms. In certain circumstances this form of a model may not accurately describe the real situation and a transformation of data may be required in order to express the model in additive terms. Frequently, if the terms are not additive, then the assumptions of Normality and equal variances may also be unsatisfied; so a judicious choice of transforms may serve to remedy all of these problems.

For example, in experiments including growth, such as the effect of diets on the weight of animals, the "true" model may be in the form:

$$Y_{ij} = e^{(\mu + \alpha_i + \varepsilon_{ij})}$$

So, the logarithmic transform

$$\ln Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

reduces the data to the standard additive form (Anderson, p.25). In addition, this transform is appropriate if treatment means are proportional to treatment standard deviations (Kirk, p. 65).

3.4 Nonrandomization

There are two major reasons for randomly assigning treatments to experimental units. First, randomization is used to ensure the neutralization of effects not under consideration in the experiment (Campbell, pp. 13,34) Second, randomization ensures the independence of observations within and between treatments.

Unfortunately, because of costs and the limited supply of experimental units, investigations of criminal justice systems are not always able to employ complete randomization. For example, in a survey, one judge (experimental unit) may be questioned about each of the punitive programs (treatments) under consideration. As a second example, instead of randomly assigning treatments to police units, all of one squad may receive the first treatment, all of another squad may receive the second treatment, and so on.

These problems, along with the lack of control created by policy changes made while the experiment is in progress require modifications to the designs described in Section 2. Two such modifications are the Analysis of Covariance design, discussed in Section 3.5, and the Block design, discussed in Sections 3.6 and 3.7.

3.5 Analysis of Covariance

In order to control for external effects, including those introduced by policy changes undertaken during the course of the experiment, the ANOVA model may be augmented by one or more independent regression variables. This augmented model, combining analysis of variance and regression, is referred to as an Analysis of Covariance (ANOCOVA) model and the independent regression variables are called covariates.

To illustrate, consider the single factor design of Section 2.2. Say that, to do policy changes, police officers with fewer years experience are made available for the experiment. This effect may be controlled by explicitly incorporating the years of experience in the model in the form of an independent regression term, i.e., a covariate.

The single factor ANOCOVA model is

$$Y_{ij} = \mu + \alpha_i + \psi(X_{ij} - \overline{X}..) + \varepsilon_{ij} \qquad \text{with} \quad \sum_i \alpha_i = 0$$

where: Y_{ij}, μ, α_i, ε_i are as described in Eqn 2-1

ψ is the covariate coeficient

X_{ij} is the covariate variable normalized about its mean $\overline{X}..$

The model is used to test the same hypothesis as Section 2.2 (all $\alpha_i = 0$) by calculating the total sum of square deviations about the regression line (instead of the mean) as follows (Ferguson, pp.350-351; Neter, pp.704-706):

$$\sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2 = \sum_{i,j} (Y_{ij} - \overline{Y}..)^2 - \frac{[\sum_{ij}(X_{ij} - \overline{X}..)(Y_{ij} - \overline{Y}..)]^2}{\sum_{ij}(X_{ij} - \overline{X}..)^2}$$

where: Y_{ij}, $\overline{Y}..$ are as defined in Figure 2-1

X_{ij}, $\overline{X}..$ are defined analogously for the covariate

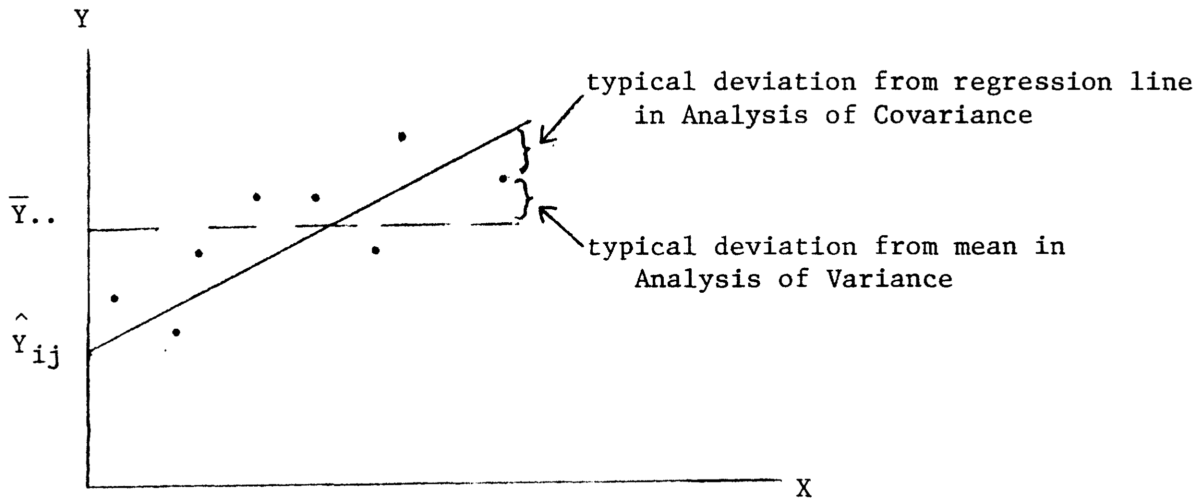
$\hat{Y}_{ij}$ is the overall regression lined predicted values.

This equation may be rewritten as

$$SS = SS_y - \frac{[SP_{xy}]^2}{SS_x}$$

The distinction between the ANOVA and the ANOCOVA models is illustrated in Figure 3-1.

Figure 3-1



In an analogous fashion, the unexplained variation (SSU) may be written as follows:

$$\sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2 = \sum_{i,j} (Y_{ij} - \bar{Y}_{i.})^2 - \frac{[\sum_{ij}(X_{ij} - \bar{X}_{i.})(Y_{ij} - Y_{i.})]^2}{\sum_{i,j} (X_{ij} - \bar{X}_{i.})^2}$$

where: Y_{ij}, $\bar{Y}_{i.}$ are as defined in Figure 2-1

X_{ij}, $\bar{X}_{i.}$ are defined analogously for the covariate

$\hat{Y}_{ij}$ is the within treatment regression line predicted values

In other words:

$$SSU = SSU_y - \frac{[SPU_{xy}]^2}{SSU_x}$$

Then, the treatment variation in each row (SSR) may be obtained by subtraction:

$$SSR = SS - SSU$$

Next, mean square estimates of the variance (MSU and MSR) are formulated by dividing SSU and SSR by their degrees of freedom ($d_1 = r(c-1)-1$ for MSU and $d_2 = r-1$ for MSR). Finally, the hypothesis is tested via the F statistic

$$F = \frac{MSR}{MSU}$$

which is F-distributed with d_2, d_1 degrees of freedom. H_o is rejected if F is too large.

3.5.1 Example

Suppose the data in Figure 2-2 did not control for the years of experience of the police officers. The response time data may be refined by the years of experience data, contained in Figure 3-2,

Figure 3-2

Experimental Units

Treatment	1	2	3	4
1	50	46	39	47
2	50	58	55	50
3	55	57	46	54

Observed Values

Experimental Units

Treatment	1	2	3	4
1	4	6	7	5
2	7	3	5	7
3	4	3	7	4

Corresponding Covariate Values

by use of the ANOCOVA model as summarized below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Statistic
Row (treatment)	SSR = 122.75	r - 1 = 2	MSR = 61.3	F = 32.0
Unexplained	SSU = 15.34	r(c-1)-1=8	MSU = 1.91	---
Total	SS = 181.75	rc-2 = 10	---	---

At the 95% level, F (32.0) is far greater than the critical value (4.46). So, we conclude there is a difference between treatments. Note that by controlling for the years of experience we can detect differences between treatments whereas, without this data, no differences can be detected (as in the example in Section 2.2.3).

3.5.2 Remarks

1) Although the ANOCOVA design does not require that the randomization of treatments assumption be met, the other assumptions (Normality, equal variances, independence of observations) are still assumed to be present.

2) The ANOCOVA model may be extended to incorporate more elaborate regression terms in a natural manner. This includes multiple covariates as well as non-linear covariate terms.

3) Another natural extension is to include multiple factors in the ANOCOVA model (Neter, p.713) such as those described in Section 2.

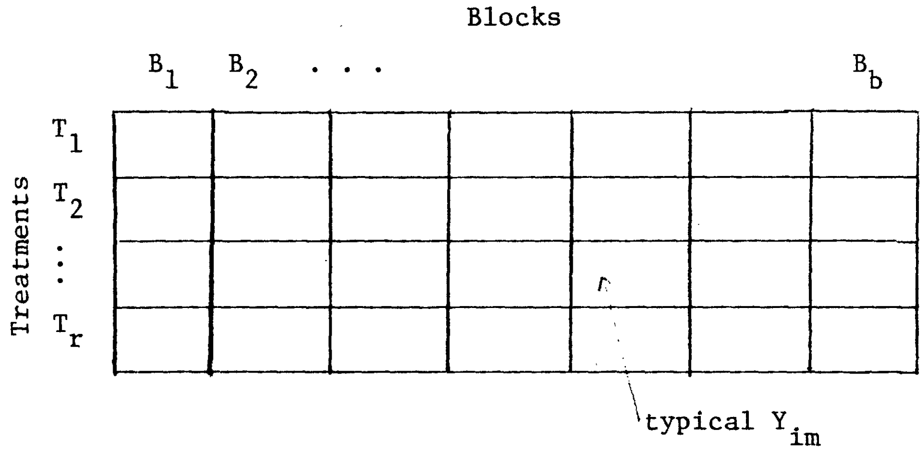
3.6 Complete Block Designs

An alternative method to the ANOCOVA design for controlling for external effects is to use a Block design. Here, experimental units that are not independent may be gathered into homogenous groups, i.e., blocks, and an additional term to account for the block effects may be explicitly added to the model. Complete Block designs, i.e., designs which assign each treatment to each block level (Anderson, p.124), are described in this section. The description of Incomplete Block designs is deferred to Section 3.7.

3.6.1 Single Block Designs

Consider, for example, an experiment in which treatments are being applied to several different squads of police officers. There may well be differences in performance between squads. To account for these differences the experimental units (the police officers) may be segmented into blocks (the squads) and treatments may then be randomly assigned to experimental units within each block. This design is illustrated in Figure 3-3

Figure 3-3



and is modelled by

Eqn 3-1 $$Y_{im} = \mu + \alpha_i + \rho_m + \varepsilon_{im} \qquad \text{with} \quad \sum_i \alpha_i = 0, \quad \sum_m \rho_m = 0$$

where: μ, α_i are as described in Eqn 2-1

Y_{im} is the i^{th} observation of block level m

ρ_m is effect due to block level m

ε_{im} is the random error distributed $N(0,\sigma^2)$

The similarity between this model and the two factor- no interaction model given in Eqn 2-4 implies that the analysis of the single block design is identical to the design in Section 2.3. This is in fact the case where the columns of experimental units are replaced by blocks (Neter, p.727).

3.6.2 <u>General Block Designs</u>

Complete Block designs may be generalized in two dimensions. One dimension refers to the number of factors included in the model. Here, additional terms may be added to Eqn 3-1 to represent additional factors as well as interactions between factors and blocks. For example, the two factor - single blocking variable model with interaction is

$$Y_{ijm} = \mu + \alpha_i + \beta_j + \rho_m + (\alpha\beta)_{ij} + (\alpha\rho)_{im} + (\beta\rho)_{jm} + \varepsilon_{ijm}$$

The other dimension into which Complete Block designs may be expanded refers to the number of blocking variables introduced into the model. A design with two blocking variables, each with three levels is shown in Figure 3-4

Figure 3-4

	A_1			A_2			A_3		
	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
Treatments T_1									
Treatments T_2									
Treatments T_3									

where one block is nested within the other (Ferguson, pp. 324-325). Letting a and b represent, respectively, the number of levels of the first and second block variables, the analysis of multiple block designs is identical to the single block design described in Section 3.6.1 For the multiple block designs the block index, m, ranges from 1 to a·b.

3.6.3 Example

Again, starting with Figure 2-2, the single block design may be exemplified by compiling the experimental units into two blocks as shown in Figure 3-5

Figure 3-5

Treatments	Block 1		Block 2	
1	50	46	39	47
2	50	58	55	50
3	55	57	46	54

The analysis is summarized below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Statistic
Row (treatment)	SSR_y = 77.58	r - 1 = 2	MSR_y = 38.79	F_1 = 0.59
Block	SSB_y = 26.13	b - 1 = 1	MSB_y = 26.125	F_2 = 0.40
Unexplained	SSU_y = 129.67	(r-1)(b-1)=2	MSU_y = 64.835	---
Total	SS_y = 223.37	rb - 1 = 5	---	---

Since the F-statistics (0.59 and 0.40) are far below the critical values (19.00 and 18.51) we accept the null hypotheses that there is no treatment effect and no block effect.

3.6.4 Remark

If the assumptions of both the ANOCOVA and Complete Block designs are met the external effects may be controlled by either method. The ANOCOVA design has the advantage that it may be implemented after the data have been collected while the Complete Block design must be constructed before the data are gathered, since experimental units are assigned within blocks (Kirk, p.488). The Complete Block design has the advantage that it is free of assumptions about the relationship between the observations and the external variables while the ANOCOVA model must incorporate a linear or nonlinear regression formulation (Neter, p. 757).

3.7 Incomplete Block Designs

Complete Block designs, that is, those in which all treatments are undertaken for each level of the blocking variables, may become rather cumbersome. As a case in point, a design using two 6-level blocking variables and six treatments would require 216 observations. One method to reduce the number of observations required would be to undertake only some of the treatments for each level of the blocking variables. That is, use an Incomplete Block design (Neter, p.764). Common Incomplete Block designs include the Latin Squares design, the Graeco-Latin Squares design, and the Youden Square design.

Although Incomplete Block designs have the advantage of reducing the number of observations, they have the disadvantage of being restricted to applications where blocks have only negligible interaction effects between treatments and other blocks (Neter, p.767). This non-interaction assumption is required to insure that variations associated with interaction will not be interpreted as variations due to treatments (Campbell, p.51; Ferguson, p.332).

3.7.1 Latin Squares Design

The Latin Square design* uses two blocking variables with one treatment per block level. The number of levels of each blocking variable must equal the number of treatments. In addition, each treatment must occur only once for each blocking level (Ferguson, p.330; Kirk, p.151; Neter, P.767). To illustrate, consider the complete Block design of Figure 3-4 with r = 3 treatments. There are $r!(r-1)! = 12$ possible Latin Square designs (Kirk, p.153) into which this design may be converted. Any of these designs may be chosen randomly. One such design is shown in Fig. 3-6.

Figure 3-6

Second Blocking Variable

First Blocking Variable	B_1	B_2	B_3
A_1	T_2	T_1	T_3
A_2	T_1	T_3	T_2
A_3	T_3	T_2	T_1

* The Latin Square design derives its name from an ancient puzzle that dealt with the number of ways Latin letters could be arranged in a square so that each letter appeared only once in each row and column (Kirk, p. 151).

The model used for the Latin Square design is

Eqn 3-2 $$Y_{imn} = \mu + \alpha_i + \rho_m + \tau_n + \varepsilon_{imn} \quad \text{with} \quad \sum_i \alpha_i = 0,\ \sum_m \rho_m = 0,\ \sum_n \tau_n = 0$$

where: μ and α_i are as described in Eqn 2-1

ρ_m is as described in Eqn 3-1

Y_{imn} is the i^{th} observation for block level m,n

τ_n is the effect due to the n^{th} level of the second blocking variable

ε_{imn} is the random error distributed $N(0,\sigma^2)$

The main hypothesis, testing differences between treatments, is

H_o: $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$

H_1: otherwise

In addition, hypotheses about blocking effects may be tested. Specifically,

H_o: $\rho_1 = \rho_2 = \ldots = \rho_a = 0$

H_1: otherwise

and

H_o: $\tau_1 = \tau_2 = \ldots = \tau_b = 0$

H_1: otherwise

These hypotheses are tested in the conventional manner by breaking the variations into sums of squares as follows:

$$\sum_{m,n} (Y_{imn} - \overline{Y}_{...})^2 = r\sum_i (\overline{Y}_{i..} - \overline{Y}_{...})^2 + \sum_{m,n} (\overline{Y}_{.m.} - \overline{Y}_{...})^2 + \sum_{m,n} (\overline{Y}_{..n} - \overline{Y}_{...})^2 +$$

$$\sum_{m,n} (Y_{imn} - \overline{Y}_{i..} - \overline{Y}_{.m.} - Y_{..n} + 2Y_{...})^2$$

Note that, except for the second term, the sums of squares are not indexed over i since m and n uniquely identify the treatment undertaken. This equation may then be rewritten using the acronyms

$$SS_y = SSR_y + SSA_y + SSB_y + SSU_y$$

The mean squares (MSR_y, MSA_y, MSB_y, MSU_y) are then obtained by dividing the sum of squares by their respective degrees of freedom. The F-statistics

$$F_1 = \frac{MSR_y}{MSU_y}, \quad F_2 = \frac{MSA_y}{MSU_y}, \quad F_3 = \frac{MSB_y}{MSU_y}$$

are then used to test for treatment effects, first blocking variable effects, and second blocking variable effects, respectively.

3.7.2 Other Designs

The Latin Squares design described in Section 3.7.1 used two blocking variables, each with the same number of levels. A design that uses three blocking variables, each with the same number of levels, is referred to as a Graeco-Latin Squares design. Designs using more than three blocking variables are referred to as Hyper-Graeco-Latin Squares designs (Kirk, pp. 166-168; Neter, p.794). The analysis of these designs is analogous to that of Section 3.7.1 where additional terms for each blocking variable are included in the model given in Eqn 3-2.

Another incomplete Block design is the Youden Squares Design. This design allows for differences between the number of levels in the two blocking variables. The analysis of these designs are more complex than the Latin Squares designs because not all treatments are undertaken for each level of the blocking variables (Kirk, pp. 441-448).

3.7.3 Example

Using the Latin Squares design in Figure 3-6, with three treatments and two 3-level blocks, data about response times of police officers such as that in Figure 3-7 may be gathered. This data, assuming no interaction, may be

Figure 3-7

	B_1	B_2	B_3
A_1	T_2 50	T_1 46	T_3 48
A_2	T_1 50	T_3 52	T_2 55
A_3	T_3 55	T_2 57	T_1 46

analyzed for differences in treatments and differences in blocks as shown in the table below:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Statistic
Treatment	$SSR_y = 68.67$	$r - 1 = 2$	$MSR_y = 34.33$	$F_1 = 5.42$
Row block	$SSA_y = 40.67$	$r - 1 = 2$	$MSA_y = 20.33$	$F_2 = 3.21$
Column block	$SSB_y = 8.0$	$r - 1 = 2$	$MSB_y = 4.0$	$F_3 = 0.63$
Unexplained	$SSU_y = 12.67$	$(r-1)(r-2) = 2$	$MSU_y = 6.33$	---
Total	$SS_y = 130.0$	$r^2 - 1 = 8$	---	---

Since all of the F values (5.42, 3.21, and 0.63) are below the critical values (19.0, 19.0, and 19.0) we accept the hypotheses that there is no difference between treatments and between blocks.

3.7.4 Remarks

1) As emphasized at the start of Section 3.7, the aptness of Incomplete Block designs, such as the Latin Squares design, are dependent upon negligible interaction effects. Therefore, it is important to be able to test for the significance of the interaction effects. One such test is the Tukey Test for Additivity which employes use of the model's unexplained variation SSU_y (Kirk, p. 160; Neter, pp. 780-781).

2) Note the large critical values (19.0) encountered in the example in Section 3.7.3. This occured because of the small number of treatments and block levels used in the experiment. Therefore, it is recommended that Latin Squares designs employ at least five treatments and block levels. This will ensure a sufficiently large number of degrees of freedom, which, in turn, will enable the critical values to be sufficiently small (Kirk, pp. 151-152).

SUMMARY OF NOTATION

Indices and Ranges

a	number of blocking levels in first blocking variable
b	" " second "
c	" experimental units
d_1, d_2	" degrees of freedom in F-statistic
h, h_{ij}	" repeated observations in factor cell
r	" treatments
i	index of first factor (treatments)
j	" second factor (experimental units)
k	" third factor
l	" repeated observations
m	" first blocking variable
n	" second blocking variable

Model Components

α_i	effect of first factor (treatments)
β_j	" second factor (experimental units)
γ_k	" third factor
ρ_m	" first blocking variable
τ_n	" second "
ψ	covariate (regression)coeficient
σ^2	variance of observations
μ	mean "
$(\alpha\beta)_{ij}$	effect of first factor/second factor interaction
$(\alpha\gamma)_{ik}$	" first factor/third factor "
$(\beta\gamma)_{jk}$	" second factor/third factor "
$(\alpha\rho)_{im}$	" first factor/first block "
$(\beta\rho)_{jm}$	" second factor/first block "
ε_{ij}, $\varepsilon_{ij\ell}$, $\varepsilon_{ijk\ell}$, ε_{im}, ε_{ijm}, ε_{imn}	random error distributed $N(0,\sigma^2)$

Observations and Statistics

A_m	label for first blocking variable
B_n	" second "
F, F_1, F_2, F_3	F-statistics
K_1, K_2	chi-square statistics
S_i^2	sample variance of first factor
$S_{\overline{Y}_i}^2$	" " first factor mean
T_i	label for first factor
Y_{ij}, $Y_{ij\ell}$, $Y_{ijk\ell}$, Y_{im}, Y_{ijm}, Y_{imn}	observed values
$\overline{Y}_{i.}$, $\overline{Y}_{i..}$	mean of first factor
$\overline{Y}_{.j}$	" second factor
$\overline{Y}_{.m.}$	" first blocking variable
$\overline{Y}_{..n}$	" second "
$\overline{Y}_{..}$, $\overline{Y}_{...}$	overall mean
$\hat{Y}_{ij}$	overall predicted regression values
$\hat{\hat{Y}}_{ij}$	within first factor "
X_{ij}	covariate values
$\overline{X}_{i.}$	covariate mean corresponding to first factor
$\overline{X}_{..}$	overall covariate mean

Mean Squares, Sum of Squares, and Sum of Products

MSR	adusted mean square for first factor (row)
MSU	" " unexplained variation
MSA_y	observed mean square for first blocking variable
MSB_y	" " second "
MSC_y	" " second factor (column)

MSR_y	observed mean square for first factor (row)
MSU_y	" " unexplained variation
$MSRC_y$	" " interaction
SS	adjusted sum of squares for total variation
SSR	" " first factor (row)
SSU	" " unexplained variation
SS_x	covariate sum of squares for total variation
SSU_x	" " unexplained variation
SP_{xy}	observed and covariate sum of products for total variation
SPU_{xy}	" " " unexplained variation
SS_y	observed sum of squares for total variation
SSA_y	" " first blocking variable
SSB_y	" " second blocking variable
SSC_y	observed sum of squares for second factor
SSR_y	" " first factor
SSU_y	" " unexplained variation
$SSRC_y$	observed sum of squares for first factor/second factor interaction

SELECTED BIBLIOGRAPHY

Mathematical Statistics

Dixon, W.J. and F.J. Massey Jr., Introduction to Statistical Analysis, York, Pa., McGraw-Hill, 1957.

Hoel, P.G., Introduction to Mathematical Statistics, New York, Wiley, 1971.

Neter, J. and W. Wasserman, Applied Linear Statistical Models, Homewood, Ill., Irwin 1974.

Specialized Books

Anderson, V.L. and R.A. McLean, Design of Experiments, A Realistic Approach, New York, Marcel Dekker, 1974.

Campbell, D. T. and J.C. Stanley, Experimental and Quasi-Experimental Designs for Research, Chicago, Rand McNally, 1963.

Chapin, F.S., Experimental Designs in Sociological Research, Westport, Conn., Greenwood Press, 1974.

Ferguson, G.A., Statistical Analysis in Psychology and Education, New York, McGraw-Hill, 1976.

Kirk, R.E., Experimental Design: Procedures for the Behavioral Sciences, Belmont, Calif., Brooks/Cole, 1968.

Lee, W., Experimental Design and Analysis, San Francisco, W.H. Freeman, 1975.

Mendenhall, W., Introduction to Linear Models and the Design and Analysis of Experiments, Belmont, Calif., Wadsworth, 1968.

Winer, B.J., Statistical Principles in Experimental Design, New York, McGraw-Hill, 1971.