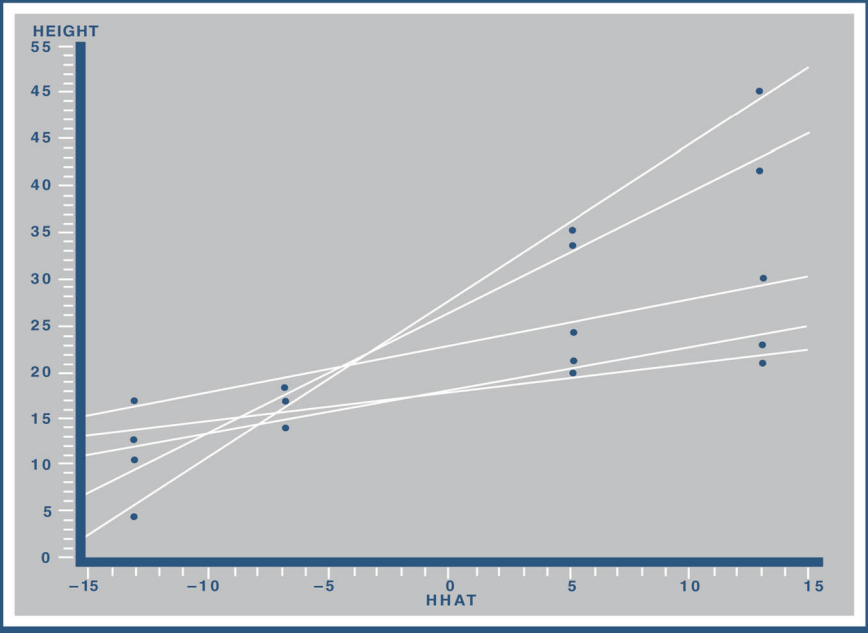
Analysis of Messy Data

VOLUME 2

NONREPLICATED EXPERIMENTS

George A. Milliken/Dallas E. Johnson



CHAPMAN & HALL/CRC

Analysis of Messy Data

VOLUME 2
NONREPLICATED EXPERIMENTS

George A. Milliken
Dallas E. Johnson

CHAPMAN & HALL/CRC

A CRC Press Company
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

(Revised for vol.2) Milliken, George A., 1943- Analysis of messy data.
Vol. 2 has imprint: New York : Van Nostrand Reinhold
Includes bibliographies and indexes
Content: v. 1. Designed experiments—v. 2. Nonreplicated experiments.
1. Analysis of variance. 2. Experimental design.
3. Sampling (Statistics) I. Johnson, Dallas, E.,
ISBN 0-412-06371-9
1938- . II. Title.
QA279.M48 1984
519.5'352 84-839
CIP

International Standard Book Number 0-412-06371-9
Library of Congress Card Number 84-839

Contents

Preface

This volume considers the analysis and design of nonreplicated experiments. It is very important that such experiments be designed efficiently and analyzed correctly, since these kinds of experiments are often much more expensive to conduct, and proper interpretation of the results is often very critical to decision making. It has been estimated that nearly 50% of all experiments are of the "nonreplicated" type and many of these are not being analyzed statistically because researchers are not aware of existing statistical methods which can be used. This book tries to provide researchers with statistical methods appropriate for nonreplicated experiments as well as some ways to make the required statistical computation feasible using existing statistical software.

Many experiments are very expensive to conduct, and requiring independent replications of all of the treatment combinations can be overly burdensome to experimenters. For example, suppose an automobile engineer wishes to determine optimal locations for seatbelt anchors. Locations must be determined so that safety and comfort are assured for drivers and occupants of many different sizes. Running an experiment which uses all possible combinations, the influential treatment factors may require 200 to 300 different runs for a single replication. Each run may cost several thousand dollars because simulated automobile crashes are expensive to conduct. Obviously even a single replication of such an experiment would be much too expensive to run.

The basic purpose of this book is to introduce several techniques and methods for analyzing experiments in which there are no independent replications of the treatment combinations being studied. Occasionally there have been clients who have several measurements on each experimental unit which they had planned to analyze as though these measurements represented independent replications of their treatment combinations. Some of the techniques presented in this book have helped to salvage information from these kinds of experiments. Many of the techniques discussed in this book are not currently available in any other books.

Users of this book will learn the following:

1. How to recognize whether replications are independent replications or dependent replications
2. How to test for interaction in nonreplicated experiments
3. How to obtain reasonable estimates of the experimental error variance in nonreplicated experiments

4. How to determine whether Tukey's model or Mandel's model can be used to model the data collected in an experiment
5. When to use a multiplicative interaction model and how to use it
6. How to use existing statistical software to fit these models
7. How to determine which treatment combinations are the primary causes of significant interactions
8. How to construct interaction plots and how to use them to interpret data
9. Simplified procedures for constructing half-normal plots
10. How to use half-normal plots to your advantage
11. How to use 2^n factorial experiments for exploratory purposes
12. How to use blocking to conduct 2^n factorial experiments more efficiently
13. How to use fractional replications of factorial experiments for exploratory purposes
14. How to use polynomial models to model certain kinds of nonreplicated experiments
15. How to use quadratic response surface models and contour plots to advantage when developing new products or improving old products

The approach used in this book is similar to that used in the first volume. That is, each topic is covered from a practical viewpoint, emphasizing the implementation of the methods much more than the theory behind the methods. Many real-world examples are used to illustrate the techniques introduced. Formulas are included for those readers who would like to program the techniques on their personal computers.

The book is intended for everyone who analyzes data. The reader should have a knowledge of analysis of variance techniques as well as basic statistical ideas. Although a knowledge of the contents of *Analysis of Messy Data Vol. 1—Designed Experiments* would be useful, such knowledge is not required.

The book contains several tables that are not available in many other books and examples that will help readers recognize the need for a particular method and show how the method should be correctly used for their own situations.

We would like to express our appreciation to Linda Kaufholz who did the initial typing of this manuscript, and we especially appreciate her willingness to learn word processing while working on this book. Her expertise played a major role in the final project. We would also like to thank Retha Parker for using her talents in the revisions of the original manuscript.

1

Analyzing Two-Way Treatment Structures with One Observation per Treatment Combination

CHAPTER OUTLINE

Many experiments are very expensive to conduct so that experimenters are often forced to limit the number of treatment combinations that can be studied in order to have adequate resources available to replicate the treatment combinations under study. The replication of treatment combinations is necessary in order to be able to have an independent estimate of the experimental error variance, which is denoted by σ^2 in this book. Having a good estimate of σ^2 is very important when the major objective of the experiment is confirmatory. However, if the major objective of the experiment is exploratory, it is often more desirable to study many different treatment combinations, each performed once, rather than a few treatment combinations each replicated many times. This book is devoted to methods that can help to extract the relevant information in experiments that are not replicated.

The first three chapters of this book are specifically devoted to analyzing two-way treatment structure experiments that have not been replicated. That is, there is only one independent observation for each treatment combination. If the cost of conducting the experiment is relatively insignificant, one cannot generally recommend designing experiments in this fashion. That is, if cost is not a factor, then one can almost always obtain better information from replicated experiments than from nonreplicated experiments.

1.1 INTRODUCTION

Suppose an experimenter wants to study the effect of temperature and humidity on the growth of a particular variety of sorghum when there are 12 growth chambers within which both the level of humidity and temperature can be controlled. The experimenter decides to study the effects of these two factors on sorghum growth by studying the 12 Temperature*Humidity combinations that are generated by considering all possible combinations of three temperature levels with four humidity levels. Typically, the experimenter places several plants of the specified variety within each growth chamber. Some researchers incorrectly analyze the observations on these plants as though they are independent replications of the 12 Temperature*Humidity treatment combinations. However, such observations are merely subsamples or repeated measures on the growth chambers rather than independent replications of the 12 Temperature*Humidity treatment combinations. The experimental units for this experiment are the 12 growth chambers; the fact that there are several plants within each growth chamber merely means that each growth chamber is being measured several times. With 12 treatment combinations and 12 experimental units, the experimenter has only one observation for each treatment combination and the usual methods for statistical analysis of the observed results do not apply because there are no independent replications from which to estimate σ^2. Thus, the experimenter is faced with several alternatives, none being very desirable. The alternatives are:

1. Decrease the number of Temperature*Humidity treatment combinations to be studied. For instance, if the number of combinations is reduced to 6, each combination could be assigned to two growth chambers resulting in two independent replications of each treatment combination.
2. Plan to repeat the experiment again at a later time using the same growth chambers, but rerandomizing the Temperature*Humidity combinations that are to be assigned to these growth chambers. This may be a viable alternative for fast-growing plants, but may not be realistic for slow-growing plants.
3. Conduct the experiment as planned, and use some of the analysis techniques described in this book.

This experiment illustrates a situation where it is possible to replicate the experiment even though the experiment may not actually get replicated. There are other situations where replicating the experiment is impossible. One such case is where an experimenter wants to compare the protein content of several varieties of wheat grown at many different locations. Such an experiment is impossible to replicate, since locations cannot be replicated. The methods described in this book will often enable experimenters to obtain usable information from experiments that are only replicated once.

Many experimenters have been observed to conduct experiments involving two-way treatment structures where the resulting data provides only one observation per treatment combination. These single-observation experiments often occur by accident. That is, the experimenter thought he or she was replicating the experiment while, in reality, the so-called replicates were really subsamples. Many experimenters have difficulty seeing the difference between true independent replications and subsampling. Those readers who have difficulty seeing the difference are advised to read Chapters 4 and 5 of Milliken and Johnson (1984). These chapters discuss split-plot and/or repeated-measures experiments. Subsampling is similar to a split-plot experiment except that no new treatments are applied to the subplot experimental units. An example involving only one independent replication per treatment combination is described in the next section.

1.2 AN EXAMPLE

Next, we give a numerical example. The example, complete with data, is used in subsequent sections to demonstrate some of the techniques for analyzing nonreplicated two-way experiments.

EXAMPLE 1.1: Growth Rate of Sorghum Plants

An experimenter has 20 growth chambers and conducts an experiment to study the effects of five temperature levels combined with each of four

Table 1.1 Mean Height of 10 Sorghum Plants

	Humidity, %			
Temperature,°F	*20*	*40*	*60*	*80*
50	12.3	19.6	25.7	30.4
60	13.7	16.9	27.0	31.5
70	17.8	20.0	26.3	35.9
80	12.1	17.4	36.9	43.4
90	6.9	18.8	35.0	53.0

humidity levels on the growth rate of sorghum plants. The experimenter places 10 sorghum plants of the same species in each of the 20 growth chambers and assigns Temperature*Humidity combinations randomly to the 20 chambers. The data given in Table 1.1 represent average heights in centimeters of the 10 plants from each growth chamber. These heights were measured after growing the plants for 4 weeks in the growth chambers. The average height of the 10 plants is used as the response because the experimental units for the Temperature*Humidity treatment combinations are the growth chambers. There is only one independent replication of each growth chamber. The variability existing between the 10 plants within a growth chamber does not measure the variability between growth chambers. That is, variability between plants growing within the same growth chamber may be much different than variability between plants growing in different growth chambers even if different growth chambers had been assigned the same Temperature*Humidity combinations.

As noted for split-plot and repeated-measures experiments, the within-growth-chamber variability is not an appropriate measure of variability with which to compare the effects of the treatments observed on the growth chambers. Several methods for analyzing this type of data are presented in the following sections.

1.3 MODEL DEFINITION AND PARAMETER ESTIMATION

A means model for experiments such as the one described in Example 1.1 is given by

$$y_{ij} = \mu_{ij} + \epsilon_{ij} \qquad i = 1,2, \ldots,t, \quad j = 1,2, \ldots,b \qquad (1.3.1)$$

where

$$\epsilon_{ij} \sim \text{i.i.d. } N(0, \sigma^2) \qquad i = 1,2, \ldots,t, \quad j = 1,2, \ldots,b$$

where i.i.d. means independently and identically distributed. In this model there are bt parameters and bt observations. The best esti-

mate of μ_{ij} is $\hat{\mu}_{ij} = y_{ij}$, $i = 1,2, \ldots t$, $j = 1,2, \ldots b$. However, no estimate of σ^2 is available, unless one is able to make some assumptions about the μ_{ij}'s. One assumption made by many statistical analysts is that the two sets of treatments do not interact. This is equivalent to assuming $\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0$ for all possible values of i, i', j, and j'. As stated in Chapter 7 of Milliken and Johnson (1984), it is also true that there is no interaction between the levels of the two sets of treatment combinations in an experiment if and only if there exist parameters μ, τ_1, τ_2, $\ldots, \tau_t$, β_1, β_2, $\ldots, \beta_b$ such that

$$\mu_{ij} = \mu + \tau_i + \beta_j \qquad i = 1,2, \ldots, t, \quad j = 1,2, \ldots, b$$

If it is true that there is no interaction between the levels of the treatments, the best estimate of σ^2 is

$$\hat{\sigma}^2 = \sum_{i=1}^{t} \sum_{j=1}^{b} (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2 / (b-1)(t-1)$$

The assumption of no interaction should not be made without some justification for it being true. However, we have found many experimenters more than willing to assume their treatments do not interact, especially when such an assumption enables them to calculate some test statistics. In fact, some experimenters do not put interaction terms in their models because they think they are not interested in the interaction. In reality, they have no choice but to be interested in interaction, if it exists; hence it is important to determine whether or not interaction exists.

1.4 WHAT CAN BE DONE?

What can an experimenter do when interaction is suspected to be present in the data or when the experimenter wants to test for it? Several methods are available to help answer this question. Unfortunately, no method is best for all situations, and most of the available methods assume that the μ_{ij}'s can be described by some sort of model other than a simple additive model. Because of this assumption, most available tests for interaction are good for certain types of interaction but not for all types.

Each of the remaining sections of this chapter presents a test for interaction for the two-way treatment structure with one observation per treatment combination.

1.5 A HEURISTIC TEST FOR INTERACTION

The test being considered in this section was presented by Milliken and Rasmuson (1977). One advantage this test has over the remaining tests presented in this chapter is that it does not require any assumptions about

the form of the interaction; however, a disadvantage is that there are some forms of interaction that the method cannot detect.

The procedure can be described as follows:

1. Partition the observations by the levels of factor T into the t sets,

$$\{y_{11}, y_{12}, \ldots, y_{1b}\}, \{y_{21}, y_{22}, \ldots, y_{2b}\}, \ldots, \{y_{t1}, y_{t2}, \ldots, y_{tb}\}$$

2. Determine the variance of the observations in each set. That is, let $v_i^2 = \sum_{j=1}^{b}(y_{ij} - \bar{y}_{i\cdot})^2/(b-1)$, $i = 1, 2, \ldots, t$. Next note that if there is no interaction, then $\mu_{ij} = \mu + \tau_i + \beta_j$ and each v_i^2 is an unbiased estimate of

$$\sigma^2 + \frac{1}{b-1}\sum_{j=1}^{b}(\beta_j - \bar{\beta}_\cdot)^2$$

However, when there is interaction, then $\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$, and each v_i^2 is an unbiased estimate of

$$\sigma^2 + \frac{1}{b-1}\sum_{j=1}^{b}(\beta_j - \beta_\cdot + \gamma_{ij} - \gamma_{i\cdot})^2 = \delta_i^2 \quad \text{(say)}$$

Hence, if one tests H_0: $\delta_1^2 = \delta_2^2 = \cdots = \delta_t^2$ and rejects, then one can conclude that there is interaction in the data.

3. To test H_0: $\delta_1^2 = \delta_2^2 = \cdots = \delta_t^2$, Milliken and Rasmuson recommend using any of the tests for homogeneity of variance given in Chapter 2 of Milliken and Johnson (1984).

For the interested reader, the v_i^2's are multiples of noncentral chi-square random variables, rather than central chi-square random variables as required for the tests in Milliken and Johnson. Thus, the homogeneity tests are only approximate for the situation described here.

One unfortunate aspect of the above test is that even when one accepts H_0, one is still not able to conclude that the data are additive. This is true since it is possible for there to be interaction in the data and still have all the δ_i^2 equal. Thus, if H_0 is rejected, there is interaction in the data, but if H_0 is accepted, one cannot guarantee no interaction. When H_0 is accepted, one could partition the data according to the levels of B, and test for equality of variances in the b sets of t observations each, using the procedure described above.

Unfortunately, even if this hypothesis is also accepted, one still cannot conclude that the data are additive. This can be illustrated by examining the following set of true cell means.

	μ_{ij}	B 1	B 2	B 3	Row Variance
T	1	8	6	4	4
	2	4	8	6	4
	3	6	4	8	4
Column Variance		4	4	4	

For the above data, all row variances are equal and all column variances are equal, and thus the heuristic test for interaction will not detect the interaction which actually exists between the two sets of treatment factors.

As an example, consider the data in Table 1.1. The column variances are 15.268, 1.828, 28.407, and 88.763, respectively. Hartley's F-max statistic is

$$F_{\max} = \frac{88.763}{1.828} = 48.56$$

which is significant at, approximately, the 1% level $[F_{\max,4,4}(.01) = 49.0]$. Critical points for this statistic are tabled in Milliken and Johnson (1984). Bartlett's test statistic is equal to

$$\chi_c^2 = 10.48$$

which is also almost significant at the 1% level ($\chi^2_{.01,3} = 11.345$).

Thus, one would conclude that interaction exists between the levels of the two sets of treatments in these data.

1.6 TUKEY'S SINGLE-DEGREE-OF-FREEDOM TEST FOR NONADDITIVITY

Tukey (1949) was the first writer to propose a test for interaction in the two-way treatment structure experiment with one observation per treatment combination. Although Tukey did not consider any particular model when he proposed the test, other authors, Ward and Dick (1952), Scheffé (1959), and Graybill (1961, 1977), showed that the test is most easily motivated by assuming that the cell means can be expressed as

$$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\tau_i\beta_j$$
$$i = 1,2,\ldots,t, \quad j = 1,2,\ldots,b \qquad (1.6.1)$$

That is, it is assumed that the interaction term γ_{ij} in the usual effects model

$$\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$$
$$i = 1,2,\ldots,t, \quad j = 1,2,\ldots,b \qquad (1.6.2)$$

is a scalar multiple of the product of the row and column main effects; i.e., $\gamma_{ij} = \lambda\tau_i\beta_j$, $i = 1,2, \ldots, t$, $j = 1,2, \ldots, b$. A test for interaction is made by testing H_0: $\lambda = 0$ versus H_0: $\lambda \neq 0$.

Tukey proposed using the sum of squares

$$\text{SSN} = \frac{\left[\sum_{ij}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})\right]^2}{\sum_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2} \tag{1.6.3}$$

as a measure of nonadditivity. When H_0: $\lambda = 0$ is true, SSN/σ^2 has a sampling distribution that is chi-square with 1 degree of freedom. The residual sum of squares after fitting the interaction term is

$$\text{SSR} = \sum_{ij}(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2 - \text{SSN} \tag{1.6.4}$$

When H_0: $\lambda = 0$ is true, SSR/σ^2 has a chi-square distribution with $(b-1)(t-1)-1$ degrees of freedom, and SSR is distributed independently of SSN. Tukey's single-degree-of-freedom test for nonadditivity is: Reject H_0: $\lambda = 0$ if

$$F_c = \text{SSN}/[\text{SSR}/(bt - b - t)] > F_{\alpha,1,bt-b-t}$$

Once again, if we fail to reject H_0, we can conclude there is no interaction of the form $\lambda\tau_i\beta_j$, but still cannot really conclude the data are additive.

1.6.1 Computing Tukey's Test Using Statistical Software

Tukey's test statistic for nonadditivity can be obtained using many existing statistical packages. The procedure requires two steps. The first step consists of fitting the additive model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{1.6.5}$$

and selecting the solution of the normal equations [see Milliken and Johnson (1984), Chapter 6] that satisfies $\sum \hat{\tau}_i = 0$ and $\sum \hat{\beta}_j = 0$. In this case, $\hat{\tau}_i$ and $\hat{\beta}_j$ are given by $\hat{\tau}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$, $i = 1,2, \ldots, t$ and $\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$, $j = 1,2, \ldots, b$. Then equation (1.6.3) simplifies to

$$\text{SSN} = \left[\sum_{ij}\hat{\tau}_i\hat{\beta}_j z_{ij}\right]^2 \Bigg/ \left[\sum_{ij}(\hat{\tau}_i\hat{\beta}_j)^2\right]$$

where

$$z_{ij} = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}$$

Secondly, one may obtain SSN using a statistical computing package by fitting the model

$$z_{ij} = \lambda \hat{\tau}_i \hat{\beta}_j + e_{ij}$$

Fitting this model is similar to fitting a simple linear regression model without an intercept. Hence,

$$\hat{\lambda} = \left[\sum_{ij} \hat{\tau}_i \hat{\beta}_j z_{ij} \right] \Bigg/ \sqrt{\sum_{ij} \hat{\tau}_i^2 \hat{\beta}_j^2}$$

and the sum of squares due to regression is

$$\left[\sum_{ij} \hat{\tau}_i \hat{\beta}_j z_{ij} \right]^2 \Bigg/ \left[\sum_{ij} \hat{\tau}_i^2 \hat{\beta}_j^2 \right] = \text{SSN}$$

The quantity SSR needed for the F-test is obtained by subtracting SSN from the residual sum of squares obtained from fitting the additive model (1.6.5). It is interesting to note that if one replaces z_{ij} in the above two equations by y_{ij}, the results are the same.

1.6.2 Computing Tukey's Test Using SAS®

Tukey's test can be obtained with SAS® by following the steps below:

Step 1: Fit model (1.6.5) by using

```
PROC GLM; CLASSES T B;
MODEL Y = T B/SOLUTION;
```

The solution vector given by SAS® satisfies $\hat{\tau}_t = 0$ and $\hat{\beta}_b = 0$ rather than $\sum \hat{\tau}_i = 0$ and $\sum \hat{\beta}_j = 0$. A new solution vector $[\hat{\mu}^*, \hat{\tau}_1^*, \ldots, \hat{\tau}_t^*, \hat{\beta}_1^*, \ldots, \hat{\beta}_b^*]'$, which satisfies $\sum \hat{\tau}_i^* = 0$ and $\sum \hat{\beta}_j^* = 0$ can be obtained from the SAS® solution by letting:

$$\begin{aligned} \hat{\mu}^* &= \hat{\mu} + \bar{\hat{\tau}}_\cdot + \bar{\hat{\beta}}_\cdot \\ \hat{\tau}_i^* &= \hat{\tau}_i - \bar{\hat{\tau}}_\cdot \qquad i = 1, 2, \ldots, t \\ \hat{\beta}_j^* &= \hat{\beta}_j - \bar{\hat{\beta}}_\cdot \qquad j = 1, 2, \ldots, b \end{aligned} \tag{1.6.6}$$

Step 2: Construct a new set of data by adding both $\hat{\tau}_i^*$ and $\hat{\beta}_j^*$ to the data card, which contains the treatment combination T_i and B_j, and then fit the model,

$$y_{ij} = \mu + \tau_i + \beta_j + \lambda \hat{\tau}_i^* \hat{\beta}_j^* + \epsilon_{ij}$$

by using

```
PROC GLM; CLASSES T B;
MODEL Y = T B THAT*BHAT/SOLUTION;
```

The F-test given by either the Type I or Type III analysis that corresponds to the row labeled by THAT*BHAT is Tukey's single-degree-of-freedom test for nonadditivity.

1.6.3 An Example Illustrating Tukey's Test

Consider the data in Table 1.1. The data were first analyzed with SAS® GLM using the statements:

```
PROC GLM;
CLASSES TEMP HUMIDITY;
MODEL HEIGHT = TEMP HUMIDITY / SOLUTION;
```

The results obtained from the SOLUTION option are shown in Table 1.2.

The estimates given in Table 1.2 are substituted into equation (1.6.6) to find the sum-to-zero solutions. First note that

$$\bar{\hat{\tau}}_{\cdot} = -3.395 \text{ and } \bar{\hat{\beta}}_{\cdot} = -13.810$$

Then

$$\hat{\mu}^* = 42.235 + (-3.395) + (-13.810) = 25.03$$

$$\hat{\tau}_1^* = -6.425 - (-3.395) = -3.030$$

$$\hat{\tau}_2^* = -6.150 - (-3.395) = -2.755$$

$$\hat{\tau}_3^* = -3.425 - (-3.395) = -.030$$

$$\hat{\tau}_4^* = -.975 - (-3.395) = 2.420$$

$$\hat{\tau}_5^* = .000 - (-3.395) = 3.395$$

$$\hat{\beta}_1^* = -26.280 - (-13.81) = -12.47$$

$$\hat{\beta}_2^* = -20.30 - (-13.81) = -6.49$$

$$\hat{\beta}_3^* = -8.66 - (-13.81) = 5.15$$

$$\hat{\beta}_4^* = .00 - (-13.81) = 13.81$$

The data set to be analyzed next is given in Table 1.3 and is equivalent to the original data set augmented by the τ_i^* and β_j^*.

The data in Table 1.3 are analyzed with SAS® GLM by using the commands:

```
PROC GLM;
CLASSES TEMP HUMIDITY;
MODEL HEIGHT = TEMP HUMIDITY THAT*HHAT/SOLUTION;
```

The Type I and Type III analysis of variance tables generated by SAS® GLM are given in Table 1.4 and the results from the SOLUTION option are given in Table 1.5.

Table 1.2 Results from SOLUTION Option of SAS® GLM

Parameter		*Estimate*	*T for H_0: Parameter = 0*	*PR > |T|*	*STD Error of Estimate*
Intercept		42.23500000 B	11.56	0.0001	3.65352341
Temp	50	−6.42500000 B	−1.57	0.1417	4.08476336
	60	−6.15000000 B	−1.51	0.1580	4.08476336
	70	−3.42500000 B	−0.84	0.4181	4.08476336
	80	−0.97500000 B	−0.24	0.8154	4.08476336
	90	0.00000000 B	.	.	.
Humidity	20	−26.28000000 B	−7.19	0.0001	3.65352341
	40	−20.30000000 B	−5.56	0.0001	3.65352341
	60	−8.66000000 B	−2.37	0.0354	3.65352341
	80	0.00000000 B	.	.	.

Table 1.3 Data for Tukey's Model

Temp	*Humidity*	*Y*	*THAT*	*HHAT*
50	20	12.3	−3.030	−12.47
50	40	19.6	−3.030	−6.49
50	60	25.7	−3.030	5.15
50	80	30.4	−3.030	13.81
60	20	13.7	−2.755	−12.47
60	40	16.9	−2.755	−6.49
60	60	27.0	−2.755	5.15
60	80	31.5	−2.755	13.81
70	20	17.8	−0.030	−12.47
70	40	20.0	−0.030	−6.49
70	60	26.3	−0.030	5.15
70	80	35.9	−0.030	13.81
80	20	12.1	2.420	−12.47
80	40	17.4	2.420	−6.49
80	60	36.9	2.420	5.15
80	80	43.4	2.420	13.81
90	20	6.9	3.395	−12.47
90	40	18.8	3.395	−6.49
90	60	35.0	3.395	5.15
90	80	53.0	3.395	13.81

Table 1.4 Type I and Type III Analysis of Variance Tables Generated by SAS® GLM

Source	DF	Sum of squares		Mean square
Model	8	2499.56700865		312.44587608
Error	11	111.79499135		10.16318103
Corrected Total	19	2611.36200000		
Model F =	30.74			PR > F = 0.0001
R-square	C.V.	Root MSE		Y mean
0.957189	12.7366	3.18797444		25.03000000
Source	DF	Type I SS	F value	PR > F
Temp	4	136.61700000	3.36	0.0498
Humidity	3	2074.29800000	68.03	0.0001
THAT*HHAT	1	288.65200865	28.40	0.0002
Source	DF	Type III SS	F value	PR > F
Temp	4	136.61700000	3.36	0.0498
Humidity	3	2074.29800000	68.03	0.0001
THAT*HHAT	1	288.65200865	28.40	0.0002

From Table 1.4 we see that Tukey's single-degree-of-freedom test for nonadditivity is $F = 28.40$ and H_0: $\lambda = 0$ is rejected at the $\alpha = .0002$ significance level. Also the value of SSN in (1.6.3) is 288.652 and the value of SSR in (1.6.4) is 111.795 and has 11 degrees of freedom. From the last row of Table 1.5, the estimate of λ in the model (1.6.1) is $\hat{\lambda} = .14273$.

Table 1.5 Results from SOLUTION Option of SAS® GLM with Tukey's Model

Parameter		*Estimate*	*T for* H_0: *Parameter* = 0	*PR* > |*T*|	*STD Error of Estimate*
Intercept		42.23500000 B	20.95	0.0001	2.01625207
Temp	50	−6.42500000 B	−2.85	0.0158	2.25423834
	60	−6.15000000 B	−2.73	0.0196	2.25423834
	70	−3.42500000 B	−1.52	0.1569	2.25423834
	80	−0.97500000 B	−0.43	0.6737	2.25423834
	90	0.00000000 B	.	.	.
Humidity	20	−26.28000000 B	−13.03	0.0001	2.01625207
	40	−20.30000000 B	−10.07	0.0001	2.01625207
	60	−8.66000000 B	−4.30	0.0013	2.01625207
	80	0.00000000 B	.	.	.
THAT*HHAT		0.14272970	5.33	0.0002	0.02678193