DETERMINING THE EFFECTS OF FERTILIZATION ON BLACK SPRUCE IN THE PRESENCE OF KALMIA ANGUSTIFOLIA

JOHN ROBERT HUMBER, B.Sc. (Hon)

DETERMINING THE EFFECTS OF FERTILIZATION ON BLACK SPRUCE IN THE PRESENCE OF *KALMIA ANGUSTIFOLIA*

BY

©JOHN ROBERT HUMBER, B.Sc. (Hon)

Submitted to the School of Graduate Studies in

partial fulfillment of the requirements

for the degree of Master of

Applied Statistics

Department of Mathematics and Statistics

Memorial University of Newfoundland

October, 1990

St. John's Newfoundland Canada

ISBN 0-315-65343-4

ABSTRACT

This study was initiated by Dr. Brain Titus who is employed with Forestry Canada. Dr. Titus was interested in the effects of fertilizers on black spruce seedlings in the presence of a shurb, *Kalmia angustifolia*, which is thought to inhibit spruce growth. Statistical analysis is presented to evaluate the differences between fertilizers for their contribution in promoting tree growth.

ACKNOWLEDGEMENTS

I would like to thank Dr. Brian Titus for giving me the opportunity to be a part of this study. Also thanks goes to Dr. Catherine Dalzell who has provided me with her knowledge and support over the many months of this project. I am grateful for this opportunity to shape my training in the areas of statistical analysis and consulting.

I am very grateful to the financial aid I have received during my years of being a graduate student. I must thank the School of Graduate Studies of Memorial University of Newfoundland for providing me with Bursary support. Finally, many thanks to Dr. Sutradhar and the Department of Mathematics and Statistics for giving me the opportunity to earn extra funding through teaching.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

With the Newfoundland economy depending greatly on the forest industry it is important that successful reforestation be carried out on sites that have been harvested. However, within the forests of Newfoundland and other parts of Eastern Canada a shrub (*Kalmia angustifolia L.*, hereafter referred to as *Kalmia*) is alleged to restrain the growth of black spruce (*Picea mariana* (Mill.) B.S.P.) and other varieties of trees (Mallik 1987). The *Kalmia* plant is a low, erect, woody shrub up to one meter in height, and is green in colour but turns a reddish-brown in late fall (Hall, Jackson and Everett 1973). It is hypothesized that *Kalmia* inhibits black growth through one or both of the following types of competition, *Resource* (or *Exploitative*) and *Allelopathic* competition. Tilman (1988) defines these types of competition as follows:

1. *Resource* competition occurs when one plant inhibits another plant through consumption of limiting resources.

2. *Allelopathic* competition occurs when one individual releases a compound that in some way inhibits growth or increases mortality of other plants.

For more discussion on these two types of competition see Walstad and Kuch (1987).

If one or both of the above types of competition is the cause of *Kalima* - induced growth inhibition in black spruce, fertilization may be a solution. To test this hypothesis and to decide which, if any, combinations of fertilizers prove effective in promoting black spruce tree growth, a greenhouse experiment was designed by Dr. B. D. Titus and Dr. A. U. Mallik.

Before the experiment was carried out, *Kalmia* plants were collected from the Botwood area on the fourteenth and fifteenth of September 1987. They were placed in pots (diameter 28.9 cm and depth 21.5 cm) and then stored at the Forestry Canada Badger Field Station waiting transportation. While stored each pot received water until they were delivered to Forestry Canada greenhouse located in St. John's on the eighteenth of September.

In the fall of the same year 240 black spruce seedlings were harvested and stored in a cold room awaiting planting. Before the seedlings were planted, seven pre-experimental variables (covariates) were measured. The covariates and a brief description of each follows:

1. Root volume : This measurement was done by displacement i.e. the difference in the weight of a large beaker of water before and after the root system was immersed (cm^3).
2. Total seedling fresh weight : Weight of each seedling at time of planting (g).
3. Root length : length of largest root (cm).
4. Stem length : measured from the base of the tree to the tip (cm).
5. First root collar diameter measurement : first measurement of the seedling's stem diameter at the base of the stem (cm).
6. Second root collar diameter measurement : second measurement of the seedling's stem diameter at the base of the stem (cm).
7. Height of seedling : above ground height of each individual seedling at time of planting (cm).

The procedure for arranging the experimental units, i.e. the pots containing the *Kalmia* plants, within the greenhouse was as follows : 48 pots were selected from the previously collected *Kalmia* plants. Each of the pots was numbered from 1 to 48 and then each was assigned randomly to the six rows and eight treatments. Next, the treatment locations were randomly assigned within a row. Finally, the 240 black spruce seedlings were planted in groups of five in each pot.

Seven fertilizers and a control were used in the experiment. The fertilizers consisted of all possible combinations of three major nutrients, N (ammonium nitrate), P (super triple phosphate) and K (potash). These combinations are denoted by:

1. OOO - Control
2. NOO - ammonium nitrate
3. OPO - phosphate
4. OOK - potash
5. NPO - ammonium nitrate + phosphate
6. NOK - ammonium nitrate + potash
7. OPK - phosphate + potash
8. NPK - ammonium nitrate + phosphate + potash

The above fertilizers were used in liquid form in order to minimize the disturbance and potential damage to the seedlings, *Kalmia* and soil microbes.

The fertilizer dosage (equivalent to 150, 160 and 100 kg ha^{-1} of elemental N, P and K, respectively) was calculated as follows:

1. Bucket diameter = 28.50 cm

   Bucket radius = 14.25 cm

   Surface area of bucket = $\pi \times r^2 = \pi \times (14.25)^2$

2. Equivalent to:

150 kg ha^{-1} elemental N = $0.9569 \text{ g bucket}^{-1}$

160 kg ha^{-1} elemental P = $0.3828 \text{ g bucket}^{-1}$

100 kg ha^{-1} elemental K = $0.6379 \text{ g bucket}^{-1}$

3. Percent nutrient content of fertilizers:

N (ammonium nitrate)	= 34.50 % N
P (super triple phosphate)	= 46 % P_2O_5 = 20.07 % P
K (potash)	= 60 % K_2O = 49.81 % K

4. Weight of fertilizers:

N : $\frac{5.7414}{x} = \frac{34.50}{100}$ $\quad x = 16.64 \text{ g replicate}^{-1}$

P : $\frac{2.2968}{x} = \frac{20.07}{100}$ $\quad x = 11.44 \text{ g replicate}^{-1}$

K : $\frac{3.8274}{x} = \frac{49.81}{100}$ $\quad x = 7.68 \text{ g replicate}^{-1}$

The environment of the greenhouse consisted of eighteen hours of light per day at a temperature of 25 degrees celsius. In the night the temperature was lowered to 20 degrees celsius. The relative humidity of the greenhouse was kept constant at 60%. Automated watering of the seedlings was carried out twice a week in the morning for two minutes per event.

The seedlings' heights were measured every three weeks up to and including week 32, which was the termination date for the experiment.

1.2 GRAPHICAL DISPLAY OF TREATMENT MEANS

Growth curves for each of the seven fertilizer treatment means over the six rows are displayed individually with the control group over time (Figures 1 to 7). Figure 8 displays all of the treatment group means over time. From this figure one should notice that the heights attained for treatment means containing N (ammonium nitrate) tend to be greater than those that do not contain N. From this it was decided to break the treatments into two groups, the first comprised of treatments containing N and the second without N (Figure 9). By viewing Figure 9 a difference in the growth curves of these two groups is indeed noticeable, especially after the period of 24 weeks.

For the purpose of this study we will only be concentrating on the final seedlings' heights. We will be interested in the effects of the different fertilizers on the height of the seedlings at week 32.

Figure 1.1

TREATMENT MEANS VS TIME

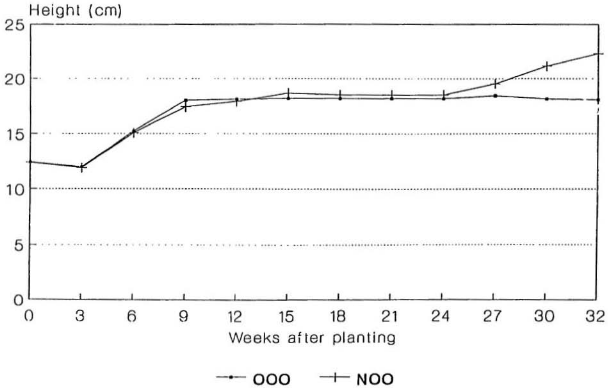


Figure 1.2
TREATMENT MEANS VS TIME

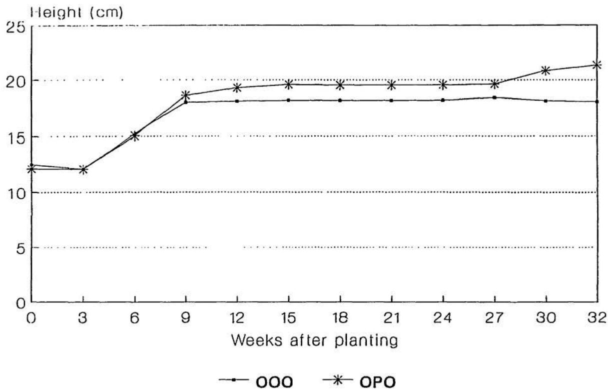


Figure 1.3
TREATMENT MEANS VS TIME

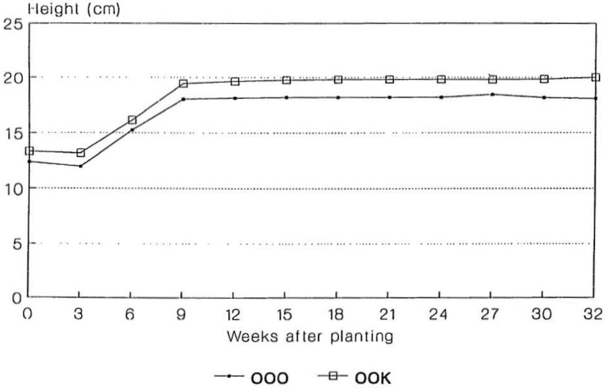


Figure 1.4
TREATMENT MEANS VS TIME

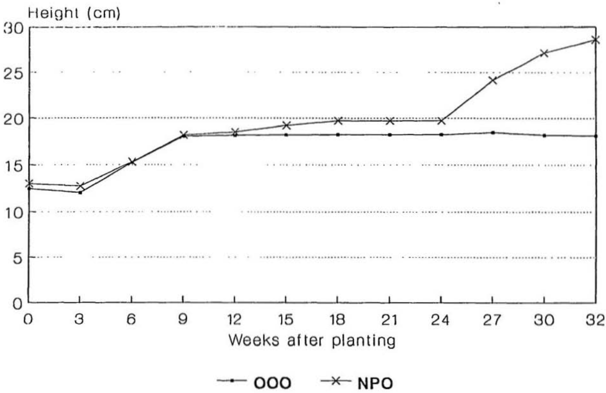


Figure 1.5
TREATMENT MEANS VS TIME

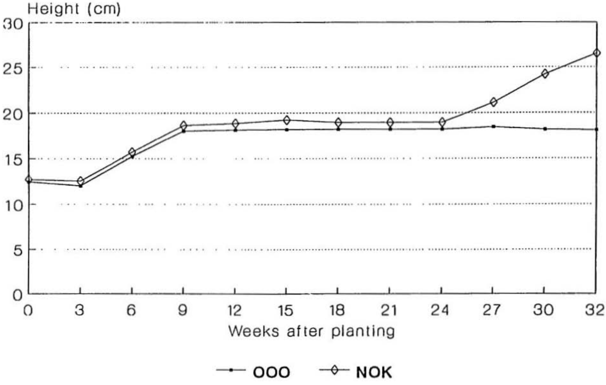


Figure 1.6
TREATMENT MEANS VS TIME

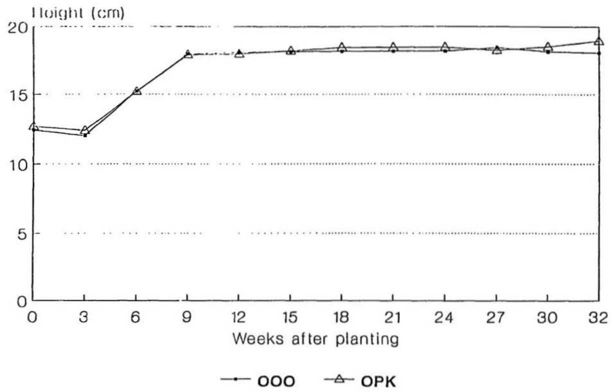


Figure 1.7
TREATMENT MEANS VS TIME

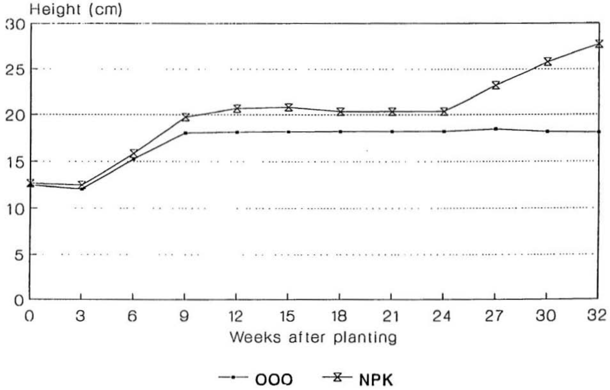


Figure 1.8
TREATMENT MEANS VS TIME

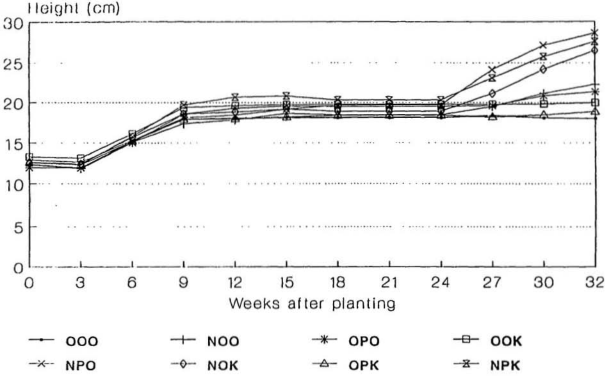
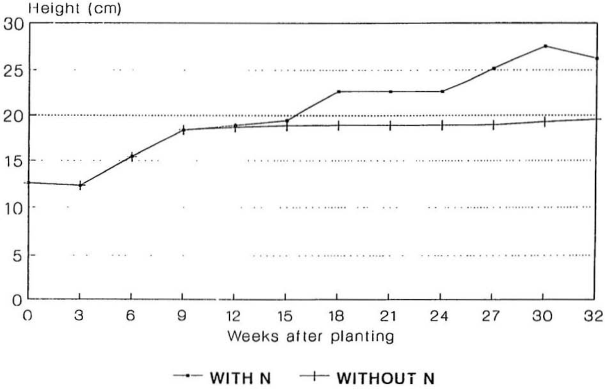


Figure 1.9
TREATMENT MEANS VS TIME



1.3 PLAN OF THE STUDY

Four questions were posed by Dr. Titus concerning this experiment. They were:

1. Is there any effect due to row positioning ?
2. For future studies are all or any of the covariates listed in Section 1.1 worth measuring ?
3. Is there any treatment effect ?
4. If the answer to question 3 is yes, which treatments are more effective in promoting tree growth ?

The answers to these four questions provides the framework for this project. Chapter 2 consists of regression analysis using the final height of the seedlings at the conclusion of the experiment as the dependent variable. Least squares will be used to provide answers to Dr. Titus' first three questions. Chapter 3 is concerned with multiple comparison procedures. These procedures are useful in determining which of the treatment effects are significantly different from each other. These procedures will be used only if the answer to Dr. Titus' third question is favourable. The final chapter will provide a non-parametric analysis of the data. It will deal with ANCOVA through the use of ranks.

CHAPTER 2

DATA ANALYSIS

2.1 INTRODUCTION

The analysis of covariance (ANCOVA) procedure may be viewed as a combination of two well known statistical techniques, analysis of variance (ANOVA) and regression analysis. This statement will become clear after the ANCOVA model is defined in 2.3. Some of the main reasons for using ANCOVA are given by Huitema (1980):

> "When the design involves the random assignment of subjects to treatments, the increase in power is the major pay off in selecting analysis of covariance. That is, the size of the error term is smaller with the use of ANCOVA rather than ANOVA if certain conditions are met. At the same time, the ANCOVA procedure includes an adjustment of treatment effect that reduces bias that may be caused by pretreatment differences between groups."[1]

By using ANCOVA we reduce pretreatment differences that may exist by reducing the error term. Therefore even before an experiment begins, i.e. before treatments are administered to the subjects, there may already exist differences between the groups under

[1]Huitema, Bradley E., 'Analysis of Covariance and its Alternatives', 1980, p.13

study. If group differences existed before the experiment started and one detects group differences after the experiment concludes, how can we distinguish between treatment and pretreatment effects? The following example is given to illustrate this point. Suppose we take one of our pre-experimental variables (covariates), say X_7, which is the initial height of the seedling before planting. One will agree that there will exist differences in the initial heights of the seedlings simply because the seedlings' heights are not uniform, thus implying that there are differences in the pretreatment group means. At the conclusion of this experiment one may find significant differences in the treatment group means by way of ANOVA. If this happens can one attribute the significant group means to treatment effects alone or does X_7 play a role? One has to take the possible covariate effect into consideration.

Analysis of covariance deals with this problem simply by eliminating the covariate effect and then proceeds to analyze the data to detect differences among the adjusted treatment group means. Adjusted treatment means are defined as the treatment means after they have been adjusted for the covariate effect, i.e. after covariate effect has been removed.

2.2 PRELIMINARY ANALYSIS

One should note that all of the seedlings in one of the 48 pots used in the experiment died. In order to correct for this, an estimate of the mean value for these five seedlings will be calculated. The idea is to calculate an estimate for the missing data point and use it throughout the analysis. This estimated mean value will restore balance to the experimental design, i.e. sample sizes will all be equal among the treatment groups. The only change in the analysis is that one degree of freedom from the error term is lost. The methodology used for this calculation is discussed by Hicks (1982). For more details on the calculation of this estimate see Appendix A.

Before any analysis on treatment effects can proceed one must provide answers to the first two questions listed in 1.3. First let us recall that Question 1 asks if there is any row effect present in the data. This simply means "does the placement of the pots used in the experiment in someway affect the final height of the seedlings ?". To find the answer to this question one may use a partial F test. This test consists of fitting two models to a set of data. The first is called a full model and contains the complete set variables (k - variables) under study, and the second is referred to as the reduced model, and contains a subset of these variables (g - variables). A partial F test determines whether or not the coefficients of the g + 1 to k parameters are equal to zero. The partial F test may be summarized as follows :

<u>COMPLETE MODEL</u>

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_g X_g + \beta_{g+1} X_{g+1} + \cdots + \beta_k X_k.$$

<u>REDUCED MODEL</u>

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_g X_g.$$

$$H_0 : \beta_{g+1} = \beta_{g+2} = \cdots = \beta_k = 0$$

$$H_A : \text{at least one of these} \neq 0$$

$$F = \frac{(SSE_1 - SSE_2) / (k - g)}{SSE_2 / n - (k + 1)}$$

where

SSE_1	=	sum of squared errors for the reduced model
SSE_2	=	sum of squared errors for the full model
$k - g$	=	the number of β parameters given by H_0
$k + 1$	=	the number of β parameters given by the complete model
n	=	the number of observations

The above F follows a F distribution with degrees of freedom equal to $\nu_1 = k - g$ and $\nu_2 = [n - (k + 1)]$.

One should note that the partial F test determines whether or not a group of coefficients associated with their respective variables are equal to zero or not. If the coefficients are indeed equal to zero, further investigation can be used through the use of sequential sum of squares. The group of g + 1 to k variables each have one degree of freedom and thus their individual contribution should be addressed. In order to determine an answer to this question, the following models were developed. The full model comprises of all the variables (see p. 3) under study and takes the following form :

COMPLETE MODEL

$$
\begin{aligned}
Y = \beta_0 &+ \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \\
&\beta_8 X_7 + \beta_9 X_1 + \beta_{10} X_{56} + \beta_{11} X_2 + \beta_{12} X_3 + \beta_{13} X_4 + \beta_{14} R_1 + \\
&\beta_{15} R_2 + \beta_{16} R_3 + \beta_{17} R_4 + \beta_{18} R_5 + \varepsilon .
\end{aligned}
$$

where

X_1	=	root volume
X_2	=	fresh weight of seedling
X_3	=	root length
X_4	=	stem length
X_{56}	=	root collar diameter, from average of X_5 and X_6
X_7	=	initial seedling height
ε	=	random error

$$R_i = \begin{cases} 1 & \text{if ith row} \\ 0 & \text{O/w} \end{cases} \qquad i = 1,2,3,4,5$$

$$T_r = \begin{cases} 1 & \text{if rth treatment} \\ 0 & \text{O/w} \end{cases} \qquad r = 1.2,3,4,5,6,7$$

Next let us consider the reduced model which contains a subset of the variables contained in the above model.

REDUCED MODEL

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 X_1 + \beta_9 X_2 + \beta_{10} X_3 + \beta_{11} X_4 + \beta_{12} X_{56} + \beta_{13} X_7 + \varepsilon .$$

By comparing these two models we are in fact testing the null hypothesis $H_0 : \beta_{14} = \beta_{15} = \beta_{16} = \beta_{17} = \beta_{18} = 0$. These β coefficients represent the row effect in the model. The ANOVA tables generated from fitting the two models by least squares is summarized in Table 2.1. From this table one should note that the row effect comprises of 5 degrees of freedom with sums of squares equal to 68.413. The partial F test proves to be non-significant and the further partitioning of this five degrees of freedom into five separate components reveals that the position of the 48 pots does not contribute to final seedlings' heights. This provides an answer to question number 1 in Section 1.1.

TABLE 2.1

ANOVA SUMMARY TABLE

Source	df	SS	MS	F
Regression (r)	13	845.86	65.07	
Regression (c)	18	914.637	50.793	
Error (r)	33	358.05	10.85	
Error (c)	28	289.637	10.344	
Row effect	5	68.413	13.683	1.32^{NS}
Total	46	1203.910		

NS = non-significant

r = reduced model

c = complete model

After eliminating the variables that represented the row effect we next bring our attention to the covariates. The second question that Dr. Titus wanted an answer to was to determine which, if any, of the covariates are important. For this we decided to test to see if the covariate effects are significantly different from zero. This question may also be answered through the method of a partial F test. To test this hypothesis, consider the following two models :

COMPLETE MODEL

$$Y = \beta_0 + \beta_1T_1 + \beta_2T_2 + \beta_3T_3 + \beta_4T_4 + \beta_5T_5 + \beta_6T_6 + \beta_7T_7 + \beta_8X_1 + \beta_9X_2 + \beta_{10}X_3 + \beta_{11}X_4 + \beta_{12}X_{56} + \beta_{13}X_7 + \varepsilon .$$

REDUCED MODEL

$$Y = \beta_0 + \beta_1T_1 + \beta_2T_2 + \beta_3T_3 + \beta_4T_4 + \beta_5T_5 + \beta_6T_6 + \beta_7T_7 + \varepsilon .$$

In this situation we are testing the hypothesis $H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = 0$. The results of running the above two models is summarized below in Table 2.2. The partial F test shows that the overall covariate effect is non-significant but further testing reveals that the covariate X_7 by itself is highly significant. Out of the seven covariates measured before planting (see p. 3), only X_7 (initial height) is worth keeping for further analysis of the data. For future experiments of this type one may only want to measure the initial height of the seedlings.

TABLE 2.2

ANOVA SUMMARY TABLE

Source	df	SS	MS	F
Regression (r)	7	703.09	100.44	
Regression (c)	13	845.86	65.07	
Error (r)	39	500.82	12.84	
Error (c)	33	358.05	10.85	
Covariate effect	6	142.77	23.795	2.19^{NS}
X_7	1	99.45	99.45	9.16^{**}
X_1	1	1.23	1.23	$< 1^{NS}$
X_{56}	1	1.85	1.85	$< 1^{NS}$
X_2	1	1.60	1.60	$< 1^{NS}$
X_3	1	38.25	38.25	3.52^{NS}
X_4	1	0.35	0.35	$< 1^{NS}$
Total	46	1203.91		

NS = non-significant

** = significant at $\alpha = 0.01$

r = reduced model

c = complete model

2.3 ANCOVA - A REGRESSION APPROACH

The one-way analysis of covariate model with one covariate is defined by:

$$Y_{ij} = \mu_{.} + \tau_{r} + \gamma(X - \bar{X}_{..}) + \varepsilon_{ij} ,$$

where

Y_{ij}	=	ith jth observation
$\mu_{.}$	=	overall mean
τ_r	=	r^{th} treatment effect
γ	=	regression coefficient for the covariate term
X	=	covariate of interest
$\bar{X}$	=	mean of covariate of interest
ε_{ij}	=	random error
$\Sigma\tau_r$	=	0 .

The usual one-way analysis of variance as is for the analysis of covariance is concerned with testing the null hypothesis $H_0 : \tau_1 = \tau_2 = \cdots = \tau_r = 0$ for r treatment groups.

As with any other statistical technique, certain assumptions must apply. The following are four assumptions that are associated with analysis of covariance. These assumptions will be presented here and discussed later in Section 2.4. The assumptions are (see Huitema (1980) and Neter and Wasserman (1974)) :

1. Error term has a normal distribution.
2. Treatment groups have equal variances.
3. Treatment groups have equal regression slopes.
4. Regression relationship is linear.

In order to test the hypothesis of equal treatment means a linear regression model was developed. Let us consider the transformation.

$$Z_{ij} = X_{ij} - \bar{X}_{..}.$$

Next let us use r - 1 indicator variables to describe the r treatment group effects.

$$T_1 = \begin{cases} 1 & \text{if 1st treatment is selected} \\ 0 & \text{O/W} \end{cases}$$

$$\vdots$$

$$T_{r-1} = \begin{cases} 1 & \text{if } (r-1)\text{th treatment is selected} \\ 0 & \text{O/W} \end{cases}$$

With these above modifications the one-way analysis of covariate model may be rewritten as

$$Y_{ij} = \beta_0 + \beta_1 T_1 + \cdots + \beta_{r-1} T_{r-1} + \beta_r Z_{ij} + \varepsilon_{ij}.$$

The relationship between these two model may be express by

$$\beta_0 = \mu_{.} + \tau_r$$

$$\beta_j = \tau_j - \tau_r \qquad j = 1, \dots, r-1$$

$$\beta_r = \gamma.$$

For our experiment we have eight treatment groups and one covariate, X_7. With this information the model that we are interested in takes the following form:

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 Z + \varepsilon.$$

where

$$T_1 = \begin{cases} 1 & \text{if 1st treatment is selected} \\ 0 & \text{O/W} \end{cases}$$

$$\vdots$$

$$T_7 = \begin{cases} 1 & \text{if 7th treatment is selected} \\ 0 & \text{O/W} \end{cases}$$

$$Z_{ij} = X_7 - \overline{X}_{..}$$

In order to test to see if the treatment effects are significant we simply test the null hypothesis that $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ for the above model. In order to test this hypothesis the following two models were constructed. The results of running a least squares regression procedure for the two models is summarized in Table 2.3.

a least squares regression procedure for the two models is summarized in Table 2.3.

COMPLETE MODEL

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 Z + \varepsilon .$$

REDUCED MODEL

$$Y = \beta_8 Z + \varepsilon .$$

TABLE 2.3

ANOVA SUMMARY TABLE

Source	df	SS	MS	F
Regression (r)	1	116.46	116.46	
Regression (c)	8	802.55	100.32	
Error (r)	45	1087.44	24.17	
Error (c)	38	401.36	10.56	
Treatment effect	7	686.08	98.01	9.28**
Total	46	1203.910		

** = significant at $\alpha = 0.01$

r = reduced model

c = complete model

Since the F value is significant we conclude that β_1, β_2, β_3, β_4, β_5, β_6, β_7 are not equal to zero. This implies that there are indeed significant treatment effects.

Another way to analyze the data is through the use of a 2^3 factorial design. Table 2.4 presents a detail breakdown of the three main treatment nutrients (N, P and K). Also present in the table is the contribution of the covariate, X_7. From this table it is clear (i) the nutrients of N and P prove to be significant and (ii) the covariate X_7 is highly significant.

In order to determine which of these treatment effects significantly differ from each other, multiple comparisons tests will be used. As previously noted this topic will be discussed and illustrated in Chapter 3. Thus Chapter 3 will provide an answer to Dr. Titus' final question.

TABLE 2.4

ANOVA TABLE FOR 2^3

FACTORIAL EXPERIMENT

Source	df		SS	MS	F
Main effects	3		602.58	200.86	19.02^{**}
N		1	530.14	530.14	50.20^{**}
P		1	67.45	67.45	6.38^{*}
K		1	4.99	4.99	0.47^{NS}
Two-way effects	3		99.87	33.29	3.15^{*}
NxP		1	19.69	19.69	1.86^{NS}
NxK		1	9.69	9.61	0.91^{NS}
PxK		1	70.52	70.52	6.67^{**}
Three-way effects					
NxPxK	1		0.70	0.70	0.06^{NS}
Covariate					
X_7	1		99.39	99.39	9.40^{**}
Error	38		401.43	10.56	9.40
Total	46		1203.91		

NS = non-significant

** = sigificant at $\alpha = 0.01$

* = significant at $\alpha = 0.05$

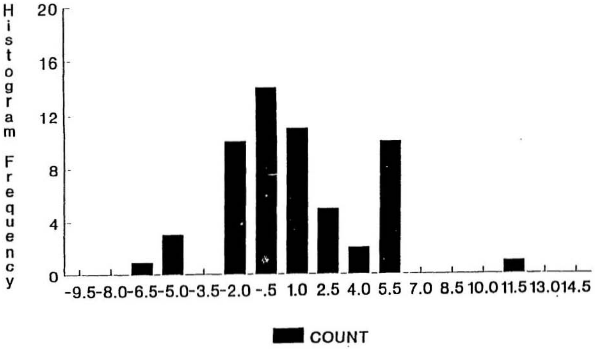
2.4 ASSUMPTION TESTING

This section is concerned with the verification of the assumptions stated in Section 2.3. It is important to verify the assumptions associated with a statistical test in order to validate the statistical analysis. If we find any departures of these assumptions we should evaluate its effect on our statistical analysis.

2.4.1 Error Term is Normal

This assumption may be verified in several ways either through a graphical display of the residuals or a more formal procedure. The histogram of the residuals associated with the fitted model (Figure 2.1) does appear to be normal. Neter and Wasserman (1977) suggest that one may use a goodness of fit test to determine whether or not the error term has a normal distribution. One may either perform a chi-square or a Kolmogorov-Smirnov (K-S) test on the residuals to check this assumption. A K-S test based on the residuals yielded a p-value of 0.438, which is large enough to indicate that the assumption of normally has not been violated.

FIGURE 2.1
FREQUENCY HISTOGRAM
RESIDUALS



2.4.2 HOMOGENEITY OF TREATMENT GROUP VARIANCES

Huitema (1980) describes a methodology for testing the assumption of equal treatment group variances. The test is based upon Bonferroni's F_B distribution. Huitemas' procedure for this test consists of four steps :

1. The residual sum of squares by group around the pooled within-group slope is computed. The formula for this quantity is :

$$\text{jth group SSres} = \left(1 - r_W^2\right)\sum y_j^2$$

where

$$r_W^2 = \frac{\sum xy_W}{\sqrt{\left(\sum x_W^2\right)\left(\sum y_W^2\right)}}$$

$$\sum xy_W = \sum xy_1 + \sum xy_2 + \cdots + \sum xy_r,$$

where

$$\sum xy_j = \sum XY - \frac{\left(\sum X_j\right)\left(\sum Y_j\right)}{n_j} \qquad \text{for } j = 1, \cdots, r$$

$$\sum x_W = \sum x_1^2 + \sum x_2^2 + \cdots + \sum x_r^2,$$

where

$$\sum x_j = \sum X_j - \frac{\left(\sum X_j\right)^2}{n_j} \qquad \text{for } j = i, \cdots, r$$

$$\sum y_w = \sum y_1^2 + \sum y_2^2 + \cdots + \sum y_r^2 ,$$

where

$$\sum y_j = \sum Y_j - \frac{\left(\sum Y_j\right)^2}{n_j} \qquad \text{for } j = 1, \cdots, r .$$

2. $S^2_{y_j \mid x}$ is calculated next, which is the estimation of the conditional variance for each of the r groups. For the jth group, $S^2_{y_j \mid x}$ is found by dividing the residual sum of squares by its degrees of freedom $n_j - 1 - c$. The quantities n_j and c denote the sample size for the jth group and the number of covariates respectively.

3. The F ratio, F_B, is calculated by dividing the largest variance estimate $S^2_{y \mid x_{largest}}$ by the smallest value $S^2_{y \mid x_{smallest}}$ founded in Step 2.

4. The F value found in 3 is compared with a Bonferroni F_B value equal to

$$F_{B\,(\alpha/2,\; c,\; n_{largest} - 1 - c,\; n_{smallest} - 1 - c)}$$

where $c = [r(r - 1)]/2$.

Complete details for this test are given in Appendix A. The value of F_B is found to be equal to 38.14. If this value is compared with $F_{B(\alpha/2,\, c,\, n_{largest} - 1 - c,\, n_{smallest} - 1 - c)}$, where $c = r(r - 1)/2 = 28$ we will find that it is less than $F_{B(.05,\, 28,\, 4,\, 4)} \approx 41.09$ and thus one may conclude that the assumption of equal conditional variances has been validated at the ten percent level.

Neter and Wasserman (1977) suggest that the assumptions of parallel slopes and linearity may also be tested by the use of the partial F test, which was discussed in third section of Chapter 2.

2.4.3. TEST OF PARALLEL SLOPES

This assumption is concern with testing to see if the slopes of the regression lines that represent the treatment groups are parallel. This is equivalent to testing to see if their is any interaction effect present in the model. If we use a partial F first to test this assumption we must first determine the complete and reduced models.

COMPLETE MODEL

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 Z_8 + \beta_9 T_1 Z + \beta_{10} T_2 Z + \beta_{11} T_3 Z + \beta_{12} T_4 Z + \beta_{13} T_5 Z + \beta_{14} T_6 Z + \beta_{15} T_7 Z + \varepsilon .$$

REDUCED MODEL

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 Z_8 + \varepsilon .$$

If we compare these two models we are testing that $H_O : \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = \beta_{15} = 0$. This hypothesis is in fact testing the assumption of parallel slopes. The results obtained from running regression analysis on these two models are summarized in Table 2.5.

TABLE 2.5

ANOVA SUMMARY TABLE

Source	df	SS	MS	F
Regression (r)	8	802.55	100.32	
Regression (c)	15	[illegible]2.844	60.856	
Error (r)	38	401.36	10.56211	
Error (r)	31	291.066	9.38921	
Interaction effect	7	110.294	15.756	1.67^{NS}
T_1Z	1	0.979	0.979	< 1
T_2Z	1	4.058	4.058	< 1
T_3Z	1	67.501	67.501	7.20^{*}
T_4Z	1	0.067	0.067	< 1
T_5Z	1	10.201	10.201	1.08
T_6Z	1	3.682	3.682	< 1
T_7Z	1	23.746	23.746	2.53
Total	46	1203.90980		

NS = non-significant

* = significant at $\alpha = 0.05$

r = reduced model

c = complete model

From the table it is clear that assumption of parallel slope has not been violated.

2.4.4. LINEARITY OF REGRESSION

The assumption of linearity of regression is concern with testing to see if there is a presence of curvature in the model. This test is in fact used to see if the curvature coefficient which is represented by β_9 contained in the complete model is zero.

<u>COMPLETE MODEL</u>

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 Z + \beta_9 Z^2 + \varepsilon .$$

<u>REDUCED MODEL</u>

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_3 T_3 + \beta_4 T_4 + \beta_5 T_5 + \beta_6 T_6 + \beta_7 T_7 + \beta_8 Z_8 + \varepsilon .$$

From Table 2.5 it is evident that the coefficient that represents possible curvature in the model is equal to zero.

TABLE 2.6

ANOVA SUMMARY TABLE

Source	DF	SS	MS	F
Regression (R)	8	802.55	100.32	
Regression (C)	9	825.588	91.732	
Error	38	401.36	10.56	
Error	37	378.322	10.225	
Quadratic effect	1	23.039	23.039	2.25^{NS}
Total	46	1203.90980		

NS = non-significant

** = significant at $\alpha = 0.01$

* = significant at $\alpha = 0.05$

CHAPTER 3

MULTIPLE COMPARISON PROCEDURES

3.1 INTRODUCTION

When the null hypothesis of equal treatment means has been rejected, we must conclude that at least two of the treatment means differ. One way to determine which means differ is through the use of a multiple comparison procedure.

3.2 MULTIPLE COMPARISONS

This section will present different tests along with their associated simultaneous confidence intervals that may be used to compare treatment means in ANCOVA. Four such tests are outlined here. Huitema (1980) discusses the following four procedures:

1. Fisher's least significant difference procedure.
2. Bryant-Paulson generalization of Tukey's honestly significant difference.

3. Dunn-Bonferoni test.

4. Scheffé test.

All of the above methods may be used for pairwise comparisons. A pairwise comparison simply compares two group means to see if they are differ from each other, and in fact tests the following hypotheses:

$$H_0 : \tau_{i\ adj} = \tau_{j\ adj}$$
$$H_1 : \tau_{i\ adj} \neq \tau_{j\ adj}$$

The final two methods (Dunn - Bonferoni, and Scheffé) may extend beyond simple comparisons of two means to more complex comparisons of group means. They can be used to explore linear combinations of treatment means.

Huitema (1980) suggests that the choice of which procedure to use depends upon two factors - (i) the type of comparisons, and (ii) whether or not simultaneous confidence intervals are of interest.

If simultaneous confidence intervals are not of interest but the main concern is some or all pairwise comparisons, then one should use the LSD procedure. The Bryant-Paulson generalization of Tukey's HSD procedure will be chosen if all pairwise comparisons and simultaneous confidence intervals are of interest to the experimenter. The Dunn -

Bonferroni procedure is useful if the number of planned pairwise comparisons is small in number. These planned comparisons may be simple or complex in nature. Finally, the Scheffé method on the other hand should be employed if the number of planned or unplanned comparisons, regardless of complexity, is large.

3.2.1 Fisher's LSD

The following test statistic has a t distribution with N - r - 1 degrees of freedom. $Y_{i\,adj}$ and $Y_{j\,adj}$ are considered significantly different if the calculated value of t is greater than the absolute value of a t distribution with its associated degrees of freedom for a given α level:

$$t = \frac{\bar{Y}_{i\,adj} - \bar{Y}_{j\,adj}}{S_{\bar{Y}_{i\,adj} - \bar{Y}_{j\,adj}}}$$

where

$$S_{\bar{Y}_{i\,adj} - \bar{Y}_{j\,adj}} = \sqrt{MSres_w \left[\left[\frac{1}{n_i} + \frac{1}{n_j} \right] + \frac{\left(\bar{X}_i - \bar{X}_j\right)^2}{SS_{W_x}} \right]}$$

$MSres_W$	=	ANCOVA mean square error
n_i, n_j	=	sample sizes for ith and jth groups
X_i, X_j	=	covariate means for the ith and jth groups

SS_{w_X} = sum of squares within groups for covariate variable

The associated simultaneous confidence interval for this test is:

$$\overline{Y}_{i\,adj.} - \overline{Y}_{j\,adj} \pm S_{\overline{Y}_{i\,adj} - \overline{Y}_{j\,adj}} [t_{\alpha\ (N-3)}]$$

where $S_{\overline{Y}_{i\,adj} - \overline{Y}_{j\,adj}}$ is given as above.

3.2.2 Bryant - Paulson generalization of Tukey's HSD

The Bryant-Paulson test uses the test statistic Q_P, which is known as the generalized studentized range statistic:

$$Q_P = \frac{\overline{Y}_{i\,adj} - \overline{Y}_{j\,adj}}{\sqrt{MSres_w\ [1 + (MS_{b_x} / SS_{w_x})] / n}}$$

where $MSres_W$ = ANCOVA mean square error

MS_{b_X} = mean square between groups for X (ANOVA on covariate)

SS_{w_X} = sum of squares within groups for X (ANOVA on covariate)

The critical value for this test is $Q_{p(\alpha, c, r, N-r-c)}$, where c is the number of covariates under study and r represents the number of treatment groups.

Simultaneous confidence intervals for the Bryant-Paulson procedure may be calculated using the formula:

$$\overline{Y}_{i\ adj} - \overline{Y}_{j\ adj} \pm \sqrt{MSres_{w}\left[1 + \frac{MS_{b_x}}{SS_{w_x}}\right] / n}$$

3.2.3 The Dunn - Bonferoni Test

This test is concerned with planned comparisons. Before the experiment is conducted the researcher may be interested in simple or complex mean comparisons. The test statistic for the Dunn-Bonferoni test is:

$$t_{DB} = \frac{c_1\left(\overline{Y}_{1\ adj}\right) + c_2\left(\overline{Y}_{2\ adj}\right) + \cdots + c_r\left(\overline{Y}_{r\ adj}\right)}{\sqrt{MSres_{W}\left[1 + \frac{MS_{b_x}}{SS_{W_x}}\right]\left[\frac{(c_1)^2}{n_1} + \frac{(c_2)^2}{n_2} + \cdots + \frac{(c_r)^2}{n_r}\right]}}$$

where

$c_1, c_2, \ldots, c_r$	=	pre-experimental contrasts
$Y_{i\ adj}, \ldots, Y_{r\ adj}$	=	adjusted treatment means
$n_1, \ldots, n_r$	=	sample size for each of the r groups
$MSres_W$	=	ANCOVA error term

MS_{b_X} = mean square between groups on X (ANOVA on covariate variable)

SS_{w_X} = sum of squares within groups on X (ANOVA on covariate variable)

Once the absolute value of t_{DB} is computed it is then compared with a critical value of $t_{DB(\alpha, k, N-r-1)}$, where k is the number of planned comparisons.

Simultaneous confidence intervals for the Dunn-Bonferoni procedure may be calculated from the formula:

$$c_1\left(\overline{Y}_{1\,adj}\right) + c_2\left(\overline{Y}_{2\,adj}\right) + \cdots + c_r\left(\overline{Y}_{r\,adj}\right) \pm t_{DB(\alpha, k, N-r-1)} \times$$

$$\sqrt{MSres_W\left[1 + \frac{MS_{b_x}}{SS_{w_x}}\right]\left[\frac{(c_1)^2}{n_1} + \frac{(c_2)^2}{n_2} + \cdots + \frac{(c_r)^2}{n_r}\right]}$$

3.2.4 The Scheffé Test

The test statistic for this test is:

$$F' = \frac{c_1\left(\overline{Y}_{1\,adj}\right) + c_2\left(\overline{Y}_{2\,adj}\right) + \cdots + c_r\left(\overline{Y}_{r\,adj}\right)}{\sqrt{MSres_W\left[1 + \frac{MS_{b_x}}{SS_{w_x}}\right]\left[\frac{(c_1)^2}{n_1} + \frac{(c_2)^2}{n_2} + \cdots + \frac{(c_r)^2}{n_r}\right]}}$$

where

$c_1, c_2, \ldots, c_r$	=	pre-experimental contrasts
$Y_{1\,adj}, \ldots, Y_{r\,adj}$	=	adjusted treatment means
$n_1, \ldots, n_r$	=	sample size for each of the r groups
$MSres_W$	=	ANCOVA error term
MS_{b_X}	=	mean square between groups on X (ANOVA on covariate variable)
SS_{w_X}	=	sum of squares within groups on X (ANOVA on covariate variable)

The critical value for this test is $\sqrt{(r-1)\,F_{(\alpha,r-1,N-r-1)}}$.

The associated simultaneous confidence intervals for Scheffé test may be obtained by using the following:

$$c_1\left(\overline{Y}_{1\,adj}\right) + c_2\left(\overline{Y}_{2\,adj}\right) + \cdots + c_r\left(\overline{Y}_{r\,adj}\right) \pm \sqrt{(r-1)F_{\alpha,r-1,N-r-1)}} \times$$

$$\sqrt{MSres_W\left[1 + \frac{MS_{b_x}}{SS_{w_x}}\right]\left[\frac{(c_1)^2}{n_1} + \frac{(c_2)^2}{n_2} + \cdots + \frac{(c_r)^2}{n_r}\right]}$$

3.3 Illustration

The purpose of this section is to illustrate one of the four procedures discussed in the previous section. The method that will be viewed here is the LSD procedure. In our case we are concerned with all possible pairwise comparisons regardless of their associated intervals.

Table 2.4 (see p. 31) presented a detailed breakdown of the three main treatment nutrients (N, P and K). From this table we concluded that the nutrients N and P are significant.

The adjustment means for the three nutrients groups are:

$$\bar{Y}_{N\ adj} = 22.55$$

$$\bar{Y}_{P\ adj} = 21.96$$

$$\bar{Y}_{K\ adj} = 19.24$$

Before we can compare the adjusted means we need the values for the quantities $MSres_W$, S_{W_X}. The value of $MSres_W$ is 10.56 which may be obtained from Table 3.1. The value of S_{W_X} is found by performing an ANOVA over the treatment groups and has a value of 98.37.

A summary of the findings are as follows:

1. N and P are significantly differ from each other.
2. N and K are significantly differ from each other.
3. P and K are significantly differ from each other.

CHAPTER 4

NONPARAMETRIC

ANALYSIS OF COVARIANCE

4.1 INTRODUCTION

Quade (1967) presents a method to perform a non-parametric analysis of covariance. Ranks are individually assigned to the X and Y data regardless of group membership. These associated ranks are then used to determine if the r groups under study have identical conditional population distributions. One should note that this method may be considered if (i) one is in doubt that the assumptions associated with a regular parametric ANCOVA have been strongly violated, or (ii) one may want to analyze data that take the form of ranks.

Let us recall that in Chapter 2 the assumption concerning equal group variances was significant at the 10 percent level. With this in mind one may use a non-parametric test for further analysis.

4.2 RANK ANCOVA

Non-parametric ANCOVA is concerned with testing the hypothesis that the conditional population distribution of Y given X are the same for all the r treatment populations.

Huitema (1980) presents a twelve step procedure for calculating the following summary table.

TABLE 4.1

NONPARAMETRIC ANCOVA

SUMMARY TABLE

Source	df	SS	MS	F
Treatment	r - 1	$\sum_{i=1}^{r}\left[\left(\sum_{j=1}^{n} Z_{ij}\right)^2 / n_j\right]$	$\frac{SStr}{r-1}$	$\frac{MStr}{MSE}$
Error	N - r	$\sum_{i=1}^{r}\sum_{j=1}^{n} Z_{ij}^2 - \sum_{i=1}^{r}\left[\left(\sum_{j=1}^{n} z_{ij}\right)^2 / n_j\right]$	$\frac{SSE}{N-r}$	
Total	N - 1	$\sum_{i=1}^{r}\sum_{j=1}^{n} Z_{ij}^2$		

The eight step procedure is as follows:

STEP 1 Rank the X data regardless of group membership. Arrange the X data in ascending order and assign a rank of one to the smallest value of X, a rank of two to the next smallest and continue assigning ranks to each of the remaining observations. If two or more observation are equal an average rank may be assigned. Once the X observations have been ranked proceed with the Y values.

STEP 2 Calculate the deviation ranks of X and Y by :

$$X_{rank} - \bar{X}_{rank} = x_{rank} \qquad Y_{rank} - \bar{Y}_{rank} = y_{rank}$$

STEP 3 Use the x_{rank}'s and y_{rank}'s found in Step 2 to calculate a Spearman rank-correlation coefficient r_S. This is equivalent to finding a Pearson correlation substituting x_{rank}'s and y_{rank}'s for the original data.

STEP 4 An estimated deviation rank on Y $(\hat{y}_{rank})$ is determined by multiplying r_S by x_{rank}

$$\hat{y}_{rank} = r_s \, (x_{rank})$$

STEP 5 If we then subtract y_{rank} from we $(\hat{y}_{rank})$ will create a residual called Z.

STEP 6 Treatment sum of squares may be calculated by:

$$\sum_{i=1}^{r}\left[\left(\sum_{j=1}^{n} Z_{ij}\right)^{2} / n_j\right]$$

STEP 7 The error sum of squares is obtained by the following formula:

$$\sum_{i=1}^{r}\sum_{j=1}^{n} Z_{ij}^{2} - \sum_{i=1}^{r}\left[\left(\sum_{j=1}^{n} Z_{ij}\right)^{2} / n_j\right]$$

STEP 8 Finally we take the ratio of

$$\frac{\text{Treatment sum of squares} / r - 1}{\text{Error sum of squares} / N - r}$$

to give the F statistic.

The F statistic is then compared with F values with degrees of freedom r - 1 and N - r. If the observed F statistic exceeds this critical value we would conclude that the conditional distribution of Y given X is not the same for all of the r treatment populations. One should note that this procedure may be shorten by performing an analysis of variance on the Z observations obtained in Step 5. A one-way ANOVA on Z by treatment group will produce a summary table equivalent to the above table.

4.3 ILLUSTRATION

The data in Table A.4 (Appendix A) will be analyzed in order to illustrate Quade's method. Table A.5 (Appendix A) shows the rankings of the original data founded in Table A.4. With the transformed data we may calculate y_{rank}s and x_{rank}s for the observations using the following :

$$x_{rank} = X_{rank} - X_{rank} = X_{rank} - 24.5$$

$$y_{rank} = Y_{rank} - Y_{rank} = Y_{rank} - 24.5$$

This information is given in Table A6 (Appendix A). With y_{rank} and x_{rank} calculated we next find the value of the Spearman rank-order correlation coefficient, r_S. Using the SPSS/PC+ statistical package, the value of r_S is .3068. Table A7 (Appendix A) summarizes the observed y_{rank}s, y_{rank}s and the residuals Z by group membership. From this table a one-way analysis of variance using a computer yielded the following summary table :

TABLE 4.2

Summary table

Source	df	SS	MS	F
Treatment	7	5998.1940	856.8849	14.42*
Error	39	2346.8407	60.1754	
Total	46	8345.0347		

From this table the F statistic is highly significant, indicating that the conditional distribution . Y given X differs over the treatment groups.

CHAPTER 5

SUMMARY AND CONCLUSIONS

The data for this experiment were collected through a greenhouse experiment conducted by Forestry Canada. The experiment was set up to evaluate the effects of various fertilizers had on black spruce in the presence of a shrub know as *Kalmia*.

Partial F tests were used in Chapter 2 to provide answers to those questions concerning the significance of the covariates and treatment factors. Of the seven covariates that were measured, only X_7, initial height of the seedling proved significant. Also within the chapter a one-way analysis of covariance (ANCOVA) model was developed. This model was used to determine which if any of the treatment fertilizers contributed to the growth of the seedlings. Multiple regression based on least squares method showed that at least two of the treatment groups significantly differed from each other. The last part of the chapter was concern with the validation of the four assumptions that are associated with ANCOVA. All four were checked and appeared not to have been violated.

Since it was discovered in Chapter 2 that significant differences between the treatment groups exists, four multiple comparison procedures which can be used to evaluate which treatment groups differ was presented in Chapter 3. One of the four, procedures, Fisher's LSD, test was illustrated and it was discovered that treatment fertilizers pairs of N and P, N and K, P and K significantly differed from each other.

Chapter 4 was concerned with a non-parametric approach to analysis of covariance. By using this type of analysis it was determined that the conditional distributions of Y given X were significantly different for treatment groups.

REFERENCES

Bryant, J. L. and Paulson, A. S. (1976). An Extension of Tukey's Method of Multiple Comparisons to Experimental Designs with Random Concomitant Variables, Biometrika, 63, 631-638.

Duncan, D. B. (1955). Multiple Range and Multiple F Tests, Biometric, 11, 1-42.

Dunn, O. J. and Clark, V. A. (1987). Applied Statistics:Analysis of Variance and Regression, New York: John Wiley and Sons, Inc.

Draper, Norman and Smith, Harry (1981). Applied Regression Analysis, 3rd edition, New York: John Wiley and Sons, Inc.

Edwards, Allen L. (1979). Multiple Regression and the Analysis of Variance and Covariance, San Francisco: W. H. Freeman and Company.

Everard, J. E. (1982). Fertilisers in the Establishment of Conifers in Wales and Southern England, Forestry Commision Booklet 41, London: Her Majesty's Stationery Office.

Hall, Ivan V., Jackson, Lloyd P. and Everett, C. Fred (1973). The Biology of Canadian Weeds 1. *Kalmia angustifolia L.* Can. J. Plant Sci. 53, 865-873.

Halperin, M. and Greenhouse, S. W. (1958). Note on Multiple Comparisons for adujted Means in the Analysis of Covariance, Biometrika, 45, 256-259.

Hicks, Charles R. (1982). Fundamental Concepts in the Design of Experiments, 3rd edtion, New York: CBS College Publishing.

Huitema, Bradley E. (1980). The Analysis of Covariance and Alternatives, New York: John Wily and Sons, Inc.

Jöreskog, K. G. (1970b). A General Method for Analysis of Covariance Structures, Biometrika, 57, 239-251.

Mallik, A. U. (1987). Allelopathic potential of Kalmia angustifolia to black spruce (*Picea mariana*), For. Ecol. Manage., 20: 43-51.

Minitab Inc. (1988). Minitab Reference Manual, Release 6.1.

Montgomery, Douglas C. (1984). Design and Analysis of Experiments, New York: John Wiley and Sons, Inc.

Neter, John and Wasserman, William (1977). Applied Linear Statistical Models, Illinois, Richard D. Irwin, Inc.

Peterson, Roger G. (1985). Design and Analysis of Experiments, New York: Marcel Dekker, Inc.

Quade, D. (1967). Rank Analysis of Covariance, Journal of the American Statistical Association, 62, 1187-1200.

Robinson, J. (1973). The Analysis of Covariance under a Randomization Model, Journal of the Royal Statistical Society, 35, 368-376.

SPSS Inc. (1988). SPSS-X User's Guide, New York.

SPSS Inc. (1988). SPSS/PC+ V3.0 Manual, Illinois.

Tilman, David (1988). Plant Strategies and the Dynamics and Structure of Plant Communities, Princeton, New Jeresy: Princeton University Press.

Walstad, John D. and Kuch, Peter J. (1987). Forest Vegetation Management for Conifer Production, New York: John Wiley and Sons, Inc.

Winer, B. J. (1971). Statistical Principles in Experimental Design, New York: McGraw Hill, Inc.

Wildt, Albert R. and Ahtola, Olli T. (1978). Analysis of Covariance, California: Sage Publications, Inc.

APPENDIX A

Table A.1 contain the mean heights of each of the 47 pots and Y, the missing observation.

Table A.1

SEEDLING HEIGHTS

AT 32 WEEKS

	Treatment								
REP	OOO	NOO	OPO	OOK	NPO	NOK	OPK	NPK	Total
I	20.54	24.16	28.98	20.84	22.23	24.70	22.38	27.76	191.59
II	21.62	24.25	23.62	21.60	31.36	22.65	18.36	29.32	192.78
III	17.64	23.75	22.92	19.14	24.18	29.40	19.38	Y*	156.41
IV	14.88	17.78	15.88	18.68	28.27	27.68	17.24	27.12	167.53
V	14.98	24.00	19.10	20.28	40.83	26.73	21.06	25.86	222.11
VI	18.82	19.92	17.62	19.54	24.89	27.93	15.10	28.06	171.88
Total	108.48	133.86	128.12	120.08	171.76	188.36	113.52	138.12	1102.30

* missing observation

$$Y_{ij}^{*} = \frac{nT_{i.} + JT_{.j} - T_{..}^{*}}{(n - 1)(J - 1)}$$

where $T_{i.}^{*}$, $T_{.j}^{*}$, and $T_{..}^{*}$ denote the row, column and overall total respectively excluding the missing observation, Y. Thus the estimate of Y_{38} is

$$Y_{38}^{*} = \frac{6(156.41) + 8(138.12) - 1102.30}{(5)(7)}$$

$$Y_{38}^{*} = 26.89$$

The following are the calculations associated which testing the assumption of equal treatment group variances discussed in Section 2.4.2 :

$$\sum y_j^2 = \sum Y_j^2 - \frac{\left(\sum Y_j\right)^2}{n_j}$$

$\Sigma y_1^2 =$	39.17	$\Sigma y_5^2 =$	231.17
$\Sigma y_2^2 =$	38.35	$\Sigma y_6^2 =$	29.96
$\Sigma y_3^2 =$	114.73	$\Sigma y_7^2 =$	34.49
$\Sigma y_4^2 =$	6.04	$\Sigma y_8^2 =$	6.90

$$\therefore \sum y_W^2 = 39.17 + 38.35 + \cdots + 6.90 = 500.81$$

$$\sum x_j^2 = \sum X_j^2 - \frac{\left(\sum X_j\right)^2}{n_j}$$

$\Sigma x^2_1 =$	25.26	$\Sigma x^2_5 =$	6.09
$\Sigma x^2_2 =$	9.28	$\Sigma x^2_6 =$	18.56
$\Sigma x^2_3 =$	6.38	$\Sigma x^2_7 =$	12.19
$\Sigma x^2_4 =$	2.60	$\Sigma x^2_8 =$	11.42

$$\therefore \ \sum x_w^2 = 25.27 + 9.28 + \cdots + 11.42 = 91.78$$

$$\sum xy_j = \sum XY_j - \frac{\left(\sum X_j\right)\left(\sum Y_j\right)}{n_j}$$

$\Sigma xy_1 =$	30.53	$\Sigma xy_5 =$	10.32
$\Sigma xy_2 =$	14.77	$\Sigma xy_6 =$	-0.43
$\Sigma xy_3 =$	25.16	$\Sigma xy_7 =$	18.81
$\Sigma xy_4 =$	1.68	$\Sigma xy_8 =$	-5.30

$$\therefore \ \sum xy_w = 30.53 + 14.77 + \cdots - 5.30 = 95.54$$

$$r_w^2 = \frac{\sum xy_w}{\sqrt{\left(\sum x_w^2\right)\left(\sum y_w^2\right)}}$$

$$r_w^2 = \frac{95.54}{\sqrt{(91.78)(500.81)}}$$

$$r_w^2 = 0.4456$$

From this we can find the residual sum of squares for the jth group by using the formula $\left(1 - r_w^2\right)\sum y_j^2$.

TABLE A.2

RESIDUAL SS

BY GROUP

Group	$(1 - r^2_w)\Sigma y^2_j$		
1	(1 - 0.4456)39.17	=	21.71
2	(1 - 0.4456)38.35	=	21.26
3	(1 - 0.4456)114.73	=	63.60
4	(1 - 0.4456)6.04	=	3.35
5	(1 - 0.4456)231.17	=	128.16
6	(1 - 0.4456)29.96	=	16.61
7	(1 - 0.4456)34.50	=	19.13
8	(1 - 0.4456)6.90	=	3.82

The next step is to calculate the conditional variances from each of the eight groups.

TABLE A.3

CONDITIONAL VARIANCES

BY GROUP

Group	$(1 - r^2_w)\Sigma y^2_j$ / n_j - 1 - c		
1	21.72/4	=	5.43
2	21.26/4	=	5.32
3	63.60/4	=	15.90
4	3.35/4	=	0.84
5	128.16/4	=	32.04
6	16.61/4	=	4.15
7	19.13/4	=	4.78
8	3.82/4	=	0.96

From Table A3 the F ratio, which is the largest divided by the smallest of the quantities is

$$F = \frac{32.04}{0.84}$$

The following pages illustrate the method of rank analysis of covariance.

TABLE A.4

RAW DATA BY

TREATMENT GROUP

OOO		NOO		OPO		OOK	
Y	X_7	Y	X_7	Y	X_7	Y	X_7
14.95	9.48	19.92	11.74	23.62	12.84	21.60	13.16
18.82	12.80	17.78	10.40	28.98	13.72	19.14	12.90
17.64	12.54	24.16	13.64	19.10	11.56	19.54	12.44
14.88	10.12	24.25	12.82	17.62	10.56	18.68	13.28
20.54	15.02	23.75	14.00	15.88	11.39	20.28	13.80
21.62	14.52	24.00	11.64	22.92	12.10	20.84	14.50

NPO		NOK		OPK		NPK	
Y	X_7	Y	X_7	Y	X_7	Y	X_7
31.36	12.56	27.68	11.58	15.10	10.74	25.86	13.52
22.23	13.16	27.93	12.32	17.24	11.06	27.12	11.32
24.89	10.92	29.40	12.30	21.06	14.20	29.32	10.18
24.18	13.86	24.70	16.12	22.38	14.40	27.76	14.00
40.53	13.72	22.65	10.46	18.36	13.28	26.89	13.50
28.27	13.54	26.73	13.24	19.38	12.26	28.06	13.16

TABLE A.5

TRANSFORMED DATA BY

TREATMENT GROUP

OOO		NOO		OPO		OOK	
Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}
2	1	16	15	27	25	21	28
11	23	8	4	44	37.5	13	26
7	21	30	36	12	12	15	20
1	2	32	24	6	6	10	31.5
18	47	28	41	4	11	17	39
22	46	29	14	26	16	19	45

NPO		NOK		OPK		NPK	
Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}
47	22	39	13	3	7	35	34
23	28	41	19	5	9	38	10
34	8	46	18	20	43	45	3
31	40	33	48	24	44	40	41
48	37.5	25	5	9	31.5	37	33
43	35	36	30	14	17	32	28

TABLE A.6

DEVIATIONS OF RANKS BY

TREATMENT GROUP

OOO		NOO		OPO		OOK	
Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}
-22.5	-23.5	- 8.5	- 9.5	2.5	.5	- 3.5	3.5
-13.5	- 1.5	-16.5	-20.5	19.5	13.0	-11.5	1.5
-17.5	- 3.5	5.5	11.5	-12.5	-12.5	- 9.5	- 4.5
-23.5	-22.5	7.5	- 0.5	-18.5	-18.5	-14.5	7.0
- 6.5	22.5	3.5	17.0	-20.5	-13.5	- 7.5	14.5
- 2.5	21.5	4.5	-10.5	1.5	- 8.5	- 5.5	20.5

NPO		NOK		OPK		NPK	
Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}	Y_{rank}	X_{rank}
22.5	- 2.5	14.5	-11.5	-21.5	-17.5	10.5	9.5
- 1.5	3.5	16.5	- 5.5	-19.5	-15.5	13.5	-14.5
9.5	-10.5	21.5	- 6.5	- 4.5	18.5	20.5	-21.5
6.5	15.5	8.5	23.5	- 0.5	19.5	15.5	17.0
23.5	13.0	0.5	-19.5	-15.5	7.0	12.5	8.5
18.5	10.5	11.5	5.5	-10.5	- 7.5	17.5	3.5

TABLE A.7

SUMMARY TABLE

	Obs.	Observed y_{rank}	Estimated y_{rank} $= 0.3068(x_{rank})$	Residual Z
	1	-22.5	-7.21	-15.29
	2	-13.5	-0.46	-13.04
Group 1	3	-17.5	-1.07	-16.43
	4	-23.5	-6.90	-16.60
	5	- 6.5	6.90	-13.40
	6	- 2.5	6.60	- 9.10
	7	- 8.5	-2.91	- 5.59
	8	-16.5	-6.29	-10.21
Group 2	9	5.5	3.53	1.97
	10	7.5	-0.15	7.65
	11	3.5	5.22	- 1.72
	12	4.5	-3.22	7.72
	13	2.5	0.15	2.35
	14	19.5	3.99	15.51
Group 3	15	-12.5	-3.89	- 8.66
	16	-18.5	5.68	-12.82
	17	-20.5	-4.14	-16.36
	18	1.5	-2.61	4.11
	19	- 3.5	1.07	- 4.57
	20	-11.5	0.46	-11.96
	21	- 9.5	-1.38	- 8.12
Group 4	22	-14.5	2.15	-16.65
	23	- 7.5	4.45	-11.95
	24	- 5.5	6.29	-11.79
	25	22.5	-0.77	23.27
	26	- 1.5	1.07	- 2.57
	27	9.5	-5.06	14.56
Group 5	28	6.5	4.76	1.74
	29	23.5	3.99	19.51
	30	18.5	3.22	15.28

TABLE A.7 (con't)

	Obs.	Observed y_{rank}	Estimated y_{rank} $= 0.3068(x_{rank})$	Residual Z
	31	14.5	-3.53	18.03
	32	16.5	-1.69	18.19
	33	21.5	-1.99	23.49
Group 6	34	8.5	7.21	1.29
	35	0.5	-5.98	6.48
	36	11.5	1.69	9.81
	37	-21.5	-5.37	-16.13
	38	-19.5	-4.76	-14.74
	39	- 4.5	5.68	-10.18
Group 7	40	- 0.5	5.98	- 6.48
	41	-15.5	2.15	-17.65
	42	-10.5	-2.30	- 8.20
	43	10.5	2.91	7.59
	44	13.5	-4.45	17.95
	45	20.5	-6.60	27.10
Group 8	46	15.5	5.22	10.28
	47	12.5	2.61	9.89
	48	17.5	1.07	16.43