A COMPARISON OF METHODS IN USE FOR EVALUATING CONTRASTS AMONG MEANS

A Thesis
Presented to
THE DEPARTMENT OF ACTUARIAL MATHEMATICS
AND STATISTICS
THE UNIVERSITY OF MANITOBA

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
BRIAN DOUGLAS MACPHERSON
April 1963

TABLE OF CONTENTS

CHAPTER		PAGE
I	THE PROBLEM AND DEFINITIONS OF TERMS USED	1
	The problem	1
	Statement of the problem.	1
	Importance of the study	2
	Definitions of terms used	2
	Population.	2
	Sample.	2
	Statistic	3
	Arithmetic mean	3
	Variance or Mean Square	3
	Standard deviation.	4
	Standard error.	4
	Range	4
	Standardized range.	4
	Studentized range	4
	Degrees of freedom.	5
	Analysis of variance.	5
	Contrast.	9
	Orthogonal contrasts.	9
II	DESCRIPTION OF METHODS.	12
	Fixed critical values	12
	Least Significant Difference (LSD).	12
	Fisher's Modified technique	13
	Tukey's Allowance procedure	14
	Scheffé's test.	15
	Dunnett's test.	16
	Multiple Critical Values.	17
	Student-Newman-Keuls procedure.	17
	Tukey's 1953 procedure.	18
	Duncan's New Multiple Range Test.	19
	Scheffé's Modified technique.	21
	Table of Critical Factors	21

CHAPTER		PAGE
III	CONSTRUCTION OF TABLES.	23
	Distribution of the Range	23
	Tables of the Studentized Range	25
	Tables for Duncan's New Multiple Range Test .	29
	Tables for Dunnett's Procedure	32
IV	COMPARISON OF METHODS	33
	Treatments versus Control	33
	Analysis of all possible contrasts.	34
	Confidence limits and tests of significance .	36
	A priori and *a posteriori* comparisons	37
	Effect of a Prior F-test.	39
	Application of Methods	40
V	CONCLUSIONS	55
BIBLIOGRAPHY	. . . .	57

LIST OF TABLES

TABLE		PAGE
I	Percentage of Barley Kernels Dehulled During an Abrasive Test.	5
II	Analysis of Variance of the Data in Table I.	7
III	Comparison of Critical Value Forms and Factors for Various Test Procedures	22
IV	Comparison of Critical Range Factors for 5% Level Tests of 25 means with f = 96	44
V	Co-operative Wheat Test (Foremost, Alberta - 1960) Experiment 1., Critical Range Value for 5% Level Tests of 25 means with f = 96 and $S_{\bar{x}}$ = 1.23 Bushels.	45
VI	Co-operative Wheat Test (Evansburg, Alberta - 1960) Experiment 2., Critical Range Values for 5% Level Tests of 25 means with f = 96 and $S_{\bar{x}}$ = 1.62 Bushels.	46
VII	Co-operative Wheat Test (Regina, Saskatchewan - 1960), Experiment 3., Critical Range Values for 5% Level Tests of 25 means with f = 96 and $S_{\bar{x}}$ = 1.10 Bushels.	47
VIII	Co-operative Wheat Test (Morden, Manitoba - 1960) Experiment 4., Critical Range Value for 5% Level Tests of 25 means with f = 96 and $S_{\bar{x}}$ = 1.16 Bushels.	48
IX	Co-operative Wheat Test (Foremost, Alberta - 1960) Experiment 1. Results.	49
X	Co-operative Wheat Test (Evansburg, Alberta - 1960) Experiment 2. Results.	50
XI	Co-operative Wheat Test (Regina, Saskatchewan - 1960), Experiment 3. Results	51
XII	Co-operative Wheat Test (Morden, Manitoba - 1960) Experiment 4. Results.	52
XIII	The number of Significant Differences Per Test Detected in the Experimental Series.	53

ABSTRACT

In the analysis of experimental data, the experimenter is faced with the problem of attempting to isolate the particular effect due to an individual treatment or combination of treatments. Methods have been proposed to accomplish this purpose but the various methods are known to give widely differing results in many instances. This study considers these methods and their properties in order to recommend which method is the best to use (1) in a general situation and (2) when particular situations are encountered either in the design of the experiment or the philosophy of the experimenter.

The study includes an outline of the following procedures,

(a) Least Significant Difference (LSD)
(b) Fisher's Modified technique
(c) Tukey's Allowance procedure
(d) Scheffé's test
(e) Dunnett's test
(f) Student-Newman-Keuls procedure
(g) Tukey's 1953 procedure
(h) Duncan's New Multiple Range Test
(i) Scheffé's Modified technique

with particular emphasis given to the method of application and critical values used in each. As an aid to the understanding of the differences between the critical values of the different procedures, a description of the distribution of the range and the tables of the range, studentized range, and special tables for Duncan's New Multiple Range Test and Dunnett's test is given.

The procedures are compared on the basis of several experimental situations and statistical techniques. These situations and techniques are such that should they be encountered in an experiment, the multiple comparison technique to be used in analysing the experiment is indicated. The effect on the choice of the test procedure by the inclusion of a control treatment in the experiment and by the *a priori* specification of the treatments to be tested is considered. The statistical techniques of confidence limits, all possible contrast and an F-test in the analysis of variance are discussed. As a guide in the application of the methods and an illustration of the topics considered, a series of four experiments is analysed and presented.

It is found that in the analysis of an experiment with a control treatment, Dunnett's procedure should be used. If the tests are specified *a priori*, the LSD procedure is valid but this is the only situation in which it may be applied. Scheffé's method is very insensitive but should be applied to the testing of contrasts involving more than two means. Tukey's Allowance procedure is recommended if confidence limits are to be placed about the true treatment mean difference. In the general situation, a strong recommendation is given to the Student-Newman-Keuls procedure.

CHAPTER I

THE PROBLEM AND DEFINITIONS OF TERMS USED

In the statistical analysis of experimental material, one of the most common questions to which an answer is desired is whether observed results are caused by particular conditions imposed upon the material by the experimenter or by the operation of chance factors only. The decision as to whether or not the treatment or variety effect in general is due to chance, offers very little difficulty to the statistician. When an attempt is made to isolate the particular causal effect or effects, increased complications are encountered.

I. THE PROBLEM

Statement of the Problem

In recent years, statistical literature has contained many articles dealing with this particular problem of determining the individual effects. As a result, the statistician is faced with a wide variety of differing procedures all purporting to do the same work.

In spite of the claims of the various authors, it has been noted with great frequency, that the application of each of the methods to the same data will result in a host of different conclusions. Since the methods do not give identical results, it should be possible to find one which is superior or certain situations in which a particular method is the best to use.

It is the purpose of this study to examine these proposed methods to determine (1) whether for a given set of experimental conditions there exists a best method, and (2) whether the characteristics of a method are particularly suitable or unsuitable for various types of experimental conditions.

Importance of the Study

The ability to isolate a particular treatment or variety as being superior to other treatments or varieties is becoming increasingly important. Larger and more complex experimental designs are being used in many scientific fields such as agriculture, chemistry, and medicine for the purpose of putting a new material or group of materials to a discriminating statistical analysis. Many of these experiments are easily analyzed to the stage where the next step is the actual isolation of the causal effects that are statistically important. If, at this stage, a procedure is used which is not valid, is too sensitive, or is not sensitive enough, not only will misleading results be obtained, but much valuable scientific work will have been lost. It is very important that some guidance be given to those people who may be faced with this problem so that the information available in the data will be fully utilized.

II. DEFINITIONS OF TERMS AND PROCEDURES USED

Throughout this study several terms and procedures are used freely. The following is a listing of several of these terms together with their definitions and symbolic representation.

Population

A population consists of all possible values of a variable. It is usually desired that the parameters of the population be obtained so that the population may be fully described. By virtue of the fact that the population may be infinite or finite, it may or may not be possible to obtain the actual parameter values.

Sample

A sample is a part of a population. The usual situation encountered is that it is impossible to observe the entire population at any one time. For the purpose of studying the characteristics of the population a representative part of it is obtained and inferences about the population are made on the basis of the sample results.

Statistic

A value obtained from a sample with a view to characterizing a parameter of the population from which the sample is obtained.

Arithmetic Mean

The sum of the values of a number of variables divided by the total number of variables. Standard notation refers to the mean as ($\bar{x}$) if obtained from a sample, or (μ) if obtained from a population. The arithmetic mean is calculated by the use of the formula:

$$\bar{x} = \frac{\sum x_i}{n}$$

where $\sum$ denotes the summation over the i sample values. The arithmetic mean is a measure of location or central tendency.

Variance, or Mean Square

The population variance is the average value of the squared deviations of the individual variables from the population mean. Symbolically it is represented by σ^2 and written:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

The sample variance is an unbiased estimate of the population variance and is represented symbolically by S^2 where

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

The variance is a measure of the dispersion of the variables about their central value.

Standard Deviation

This is defined as the square root of the variance and is denoted by (σ) for a population and (S) for a sample.

Standard Error

Since it is possible to obtain a large number of samples from a population, it is to be expected that statistics calculated from these samples will themselves be subject to random variation. A measure of this variation is the standard deviation or standard error of the sample statistic. The standard error referred to most frequently in this study is the standard error of the mean, denoted by ($S_{\bar{x}}$) for the sample values. The relation between the standard deviation of the variables and the standard error of a mean is given by:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Range

The range of a sample is defined as the difference between the maximum and minimum values in the sample. It may be written symbolically as:

$$w = x_{max.} - x_{min.}$$

Standard Range

The ratio of the range of a sample to the standard deviation of the population from which the sample is drawn.

Studentized Range

The ratio of the range of a sample to the sample value of the standard deviation. It may be written as:

$$q = \frac{w}{S}$$

The values of the range and the standard deviation are obtained from independent samples.

Degrees of Freedom

A term which is used to denote the number of independent comparisons that can be made among the members of a sample. Throughout this study, the degrees of freedom will be denoted by "f".

Analysis of Variance

A procedure by which the total variation contained in a set of observations may be separated into components readily associated with defined sources of variation used to classify the observations. The variation is measured as the sum of squares of the deviations from the mean. To illustrate the procedures and terms closely associated with the analysis of variance, an example is presented.

The observations in Table I are percentages of barley kernels dehulled during an abrasive test. The object of the experiment, arranged as a randomized complete block with six complete replications, is to test for differences among hull characteristics of four barley varieties.

TABLE I

PERCENTAGES OF BARLEY KERNELS DEHULLED DURING AN ABRASIVE TEST

Replications	Varieties				
	Parkland	MC247	ND_{B16}	ND_{B117}	Total
I	32.0	11.2	11.2	27.2	81.6
II	29.6	8.4	19.2	36.8	94.0
III	28.8	5.2	18.4	42.8	95.2
IV	34.0	10.0	16.8	50.0	110.8
V	29.6	5.2	11.2	34.8	80.8
VI	24.8	3.6	7.2	35.6	71.2
Total	178.8	43.6	84.0	227.2	533.6

It is possible to obtain a sum of squares identifiable with the variation among the total observations and to subdivide this into effects due to the varieties, the replications, and unexplained experimental error variation.

The total sum of squares is obtained by summing the squares of the individual observations. Since the desired sum of squares is to be in terms of deviations from the mean, correction of this total sum of squares for the contribution due to the mean is necessary. The correction factor is calculated as the square of the grand total divided by the number of observations. The total sum of squares is thus found to be:

$$(32.0)^2 + (29.6)^2 + (28.8)^2 + \ldots\ldots + (34.8)^2 + (35.6)^2 - \frac{(533.6)^2}{24} = 4072.61$$

The sum of squares due to replications is:

$$\frac{(81.6)^2 + (94.0)^2 + \ldots\ldots + (71.2)^2}{4} - \frac{(533.6)^2}{24} = 244.37$$

where the observed replicate totals are obtained from Table I and the divisor four is the number of observations per replicate.

The sum of squares due to varieties is obtained in a similar manner:

$$\frac{(178.8)^2 + (43.6)^2 + (84.0)^2 + (227.2)^2}{6} - \frac{(533.6)^2}{24} = 3560.66$$

where the variety totals are obtained from Table I and the divisor six is the number of replications.

The sum of squares for experimental error can now be obtained by subtracting both the replicate sum of squares and the variety sum of squares from the total sum of squares.

These sum of squares results may be summarized in the standard analysis of variance table. Table II gives the results as follows:

TABLE II

ANALYSIS OF VARIANCE OF THE DATA IN TABLE I

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Varieties	3	3560.66	1186.89
Replications	5	244.37	48.87
Error	15	267.58	17.84
Total	23	4072.61	

The column, degrees of freedom, in the analysis of variance table is obtained by the application of the definition given for this quantity. Among the twenty-four observations, twenty-three independent comparisons can be made and as a result, the total number of degrees of freedom for this experiment will be twenty-three. In accordance with the three subdivisions of the total sum of squares, the twenty-three degrees of freedom are divided into three parts representing the number of independent comparisons among varieties, among replicates and among the experimental error. The number of degrees of freedom among varieties and among replicates is one less than the number of items which may be compared within each classification. After removing the five degrees of freedom for replicates and the three degrees of freedom for varieties, the remaining fifteen degrees of freedom are attributed to experimental error.

The column headed "mean square" is found simply by dividing the sum of squares for each classification by the degrees of freedom appropriate to it. The value of the mean square for error provides an unbiased estimate of the error variance σ^2. This estimate, S^2, enables the experimenter to obtain the standard error of a variety mean. It is found by calculating:

$$S_{\bar{x}} = \sqrt{\frac{S^2}{r}}$$

where "r" represents the number of replications. In this example the value obtained is:

$$S_{\bar{x}} = \sqrt{\frac{17.84}{6}}$$
$$= 1.72$$

Having completed the analysis of variance table, the next step in the analysis of the experiment is to test the hypothesis that each of the four varieties has shown equal results. The test is achieved by the use of the ratio of the mean square for varieties to the mean square for error. This ratio follows the F-distribution and its value may thus be compared with a critical value obtained from the tables of the distribution of F in order to judge its significance. The observed value of F in this case is found to be:

$$\frac{1186.89}{17.84} = 66.53$$

Since the mean square for varieties carries three degrees of freedom and the mean square error has fifteen degrees of freedom, this calculated F-statistic must be compared with the tabular F-value with three and fifteen degrees of freedom in the numerator and denominator respectively. The critical value applicable in this case is obtained as:

$$F_{3,15,(0.05)} = 3.29$$

at the 0.05 probability level and,

$$F_{3,15,(0.01)} = 5.42$$

at the 0.01 probability level. Since the calculated value exceeds the tabulated value at probability level 0.01, it can be stated that the probability that the calculated value could have reached the value 66.53 by chance is less than 0.01. The calculated F-statistic is thus declared to be highly significant. Had the F-ratio exceeded the 0.05 critical value but not the 0.01 value, the ratio would be declared significant. It is thus necessary to reject the claim that the varieties are all alike and conclude that there is evidence to support a claim that at least one of the varieties is different from the rest.

Contrast

A contrast among variety totals is defined as any linear function of the variety totals of the form:

$$C_i = c_{i1}V_1 + c_{i2}V_2 + \ldots\ldots + c_{ik}V_k$$

such that,

$$c_{i1} + c_{i2} + \ldots\ldots\ldots + c_{ik} = 0$$

The contrast carries with it a component of the variety sum of squares with one degree of freedom. The value of the component of the sum of squares is calculated as:

$$\frac{C_i^2}{r\ D_i}$$

where

$$D_i = c_{i1}^{\ 2} + c_{i2}^{\ 2} + \ldots\ldots + c_{ik}^{\ 2}$$

and "r" is the number of replications.

Orthogonal Contrasts

Two contrasts of the form:

$$C_1 = c_{11}V_1 + c_{12}V_2 + \ldots\ldots + c_{1k}V_k$$
$$C_2 = c_{21}V_1 + c_{22}V_2 + \ldots\ldots + c_{2k}V_k$$

are said to be orthogonal if the sum of the products of corresponding coefficients of the V_i totals, is equal to zero. This may be represented as:

$$c_{11}c_{21} + c_{12}c_{22} + c_{13}c_{23} + \ldots\ldots + c_{1k}c_{2k} = 0$$

If in a set of contrasts, every pair of contrasts is orthogonal, the set is said to be an orthogonal set of contrasts. If any pair of contrasts is not orthogonal, the set of contrasts is said to be non-orthogonal.

An important property of an orthogonal set of contrasts is that the sum of the component sums of squares will be equal to the sum of squares for the over-all variety effect. Since each contrast carries one degree of freedom, it is possible to have as many contrasts as there are degrees of freedom in the variety sum of squares.

To illustrate the contrast concepts, use is made of the variety totals obtained from the data in Table I.

PARKLAND	MC247	ND_{B16}	ND_{B117}
178.8	43.6	84.0	227.2

Since the sum of squares for varieties contains three degrees of freedom, it is possible to obtain three contrasts each carrying one degree of freedom. It might be of interest to divide the varieties into two groups, one containing varieties Parkland and MC 247 and the other containing ND_{B16} and ND_{B117}. Component sums of squares may thus be obtained representing the difference between the groups and also the differences between the varieties within each of the groups. These effects may be represented in terms of combinations of the coefficients of the variety totals as follows:

1.	+1	+1	-1	-1
2.	+1	-1	0	0
3.	0	0	+1	-1

It is noted that each of these three effects is a contrast since the sum of the coefficients for each effect is zero. They also form an orthogonal set since the sum of the cross products total zero for every pair of contrasts.

The numerical value for these contrasts will be obtained as follows:

1. $\frac{(178.8 + 43.6 - 84.0 - 227.2)^2}{(6)(4)} = 328.56$

2. $\frac{(178.8 - 43.6)^2}{(6)(2)} = 1523.25$

3. $\frac{(84.0 - 227.2)^2}{(6)(2)} = 1708.85$

The sum of these component sums of squares is equal to 3560.66 which is identical with the value given in Table II for the sum of squares due to varieties.

Since each of these components is a sum of squares with one degree of freedom, the particular effect may be tested by obtaining the ratio of the value of the component to the mean square for error. This ratio is an F-ratio and may be compared with tabular value from the tables of the F-distribution with the appropriate degrees of freedom.

An example of a non-orthogonal set of contrasts may be represented in terms of the coefficients of the variety totals as:

1.	+3	-1	-1	-1
2.	+1	+1	-1	-1
3.	0	0	+1	-1

Consideration of these three effects individually reveals that in each case the sum of the coefficients is zero. As dictated by the definition, each effect is thus a contrast. The sum of products of the corresponding variety total coefficients in the three paired combinations of effects is not zero in every case and as a result the set is non-orthogonal.

The calculated component sum of squares for these contrasts are

1. 458.04
2. 328.56
3. 1708.85

The sum of these component effects is equal to 2495.45 which is not equal to the sum of squares for varieties.

CHAPTER II

DESCRIPTION OF METHODS

The multiple comparison techniques that are discussed here differ among themselves in three ways:

a. their purpose

b. their application

c. their results

In spite of these differences, similarities among the tests make it possible to group the methods according to the critical values used. In this chapter, each test is presented, giving where applicable, a brief historical and theoretical account of its development, its method of application and the determination of the critical values used.

The following is a list of the methods presented in this chapter, grouped according to the form of the critical value or values used:

I. Fixed Critical Value

- a. least significant difference
- b. Fisher's modified technique
- c. Tukey's allowance procedure
- d. Scheffé's test
- e. Dunnett's test

II. Multiple Critical Values

- a. Student-Newman-Keuls procedure
- b. Tukey's 1953 procedure
- c. Duncan's New Multiple Range test
- d. Scheffé's modified technique

I. FIXED CRITICAL VALUE

Least Significant Difference (LSD)

The LSD method of making comparisons among means, is perhaps the first method that was used in making additional