QUEEN MARY UNIVERSITY OF LONDON

The Analysis of Designed Experiments with Multivariate Data

MSc in MATHEMATICS

MTHM038

23 March 2016

Author:

Emre ARI

Supervisor:

Dr. D. Steve COAD

Abstract

This project involves a detailed study of multivariate analysis of variance (MANOVA), the main statistical technique used to analyse data from a designed experiment in which the data consists of observations on more than one variable. In its simplest form, MANOVA can be used to compare several mean vectors or treatments effects vectors, given random samples from multivariate normal distributions, each with the same covariance matrix. This method is known as One-Way MANOVA.

First, we will introduce a completely randomised design analysis of variance (One-Way ANOVA) method. The One-Way ANOVA method tests one independent variable (i.e. the treatment effects) on one dependent variable. We will briefly introduce the randomised block design analysis of variance (RBD ANOVA) in our study. RBD ANOVA analysis blocks both effects and treatment effects.

We then come to examine the main topic of this project, which is how a multivariate analysis of variance with one factor level (One-Way MANOVA) tests one independent variable on more than one dependent variable. The One-Way MANOVA method will use a special test called Wilks' Lamba to test the effect of the treatments vector. We later give some idea of the RBD MANOVA method, which tests the blocks effects vector in addition to the treatments effect vectors. After this section, we will analyse whether there are any differences between the types of school apparent in the GenStat statistical programme.

This study will, thus, analyse three different types of school (Private, Science and State High Schools) and how they affect students' net scores in multiple choice exams by using the One-Way MANOVA method. This exam includes questions about three lessons which are Maths, Turkish language and English language. Under the One-Way MANOVA, those types of school will be analysed using the Genstat statistical programme.

Contents

1 Introduction **1**

2 A Review of One-Way ANOVA **3**

2.1 One-Way ANOVA Table and F-test 4

2.2 One-Way ANOVA Example 6

2.3 Pairwise Comparisons 8

3 Randomised Block Design for ANOVA **9**

3.1 ANOVA Table and F-test for Randomised Block Design 11

3.2 A Numerical Example 13

4 One-Way MANOVA **15**

4.1 The Wilks' Lambda Test 18

4.2 The Wilks' Lambda Test for One-Way MANOVA 19

4.3 Bartlett's Approximation to One-Way MANOVA 23

4.4 Find Pooled Sample Variance-Covariance Matrix 24

4.5 Box's M Test for Covariance Assumptions 27

4.6 The Bonferroni Approach to Treatment Effects 29

5 Randomised Block Design MANOVA **31**

5.1 Wilks' Lambda Test for (RBD) MANOVA 33

5.2 Bartlett's Correction to (RBD) MANOVA 34

5.3 The Bonferroni Approach to (RBD) MANOVA 34

5.4 A Numerical Example 35

6 A GenStat Application **38**

6.1 Bonferroni Approach 40

6.2 Assumption Tests 41

6.2.1 Box M Test 41

6.2.2 Normality 42

7 Conclusion **43**

8 References **44**

9 Appendix **45**

9.1 GenStat 45

9.2 Excel 49

1 Introduction

Multivariate Analysis of Variance (MANOVA) models have been useful and implementable in a wide range of fields ranging from economics to biostatistics. Moreover, the models are essential tools for designing experiments in which the data consists of observations regarding more than one variable. MANOVA simply tests the difference between two or more means of treatment vectors.

The MANOVA method is usually used to test whether there are differences between mean vectors or treatment vectors since MANOVA uses more than one dependent variable. This means that MANOVA is a generalisation of the Analysis of Variance (ANOVA) method which tests one dependent variable if there are differences between the mean effects or if there are differences between different treatments. Zetterberg mentioned in his dissertation (2013) that the ANOVA method was first introduced by Sir Ronald A. Fisher in 1920s as a method for testing biological and agricultural data.

The objective of this study is mainly going to concentrate on more than one dependant variable with one independent variable; this method is known as the One-Way MANOVA. The One-Way MANOVA method, which has a similar relationship between ANOVA and MANOVA, is a generalisation of the One-Way ANOVA. That is why the One-Way MANOVA method will be introduced first in this thesis. This dissertation's sections, therefore, will be given in the following way:

In the section 2, initially, theory lying behind the One-Way ANOVA method, which is known as a completely randomised design analysis of variance, will be given. The One-Way ANOVA method will test the null hypothesis of whether there are different treatment effects by F-test statistics. This section will show an example under a practical algorithm calculation by facilitating the construction of a One-Way ANOVA Table. If the null hypothesis is rejected, a pairwise comparison test will enable one to find which treatments differs from the others.

The section 3 will, in turn, present the Randomised Block Design Analysis of Variance (RBD ANOVA) method. That method analyses block effects in addition to treatment effects. Because the block effect test has more sensitive test results, this model will be illustrated by a numerical example.

In the section 4, under some assumptions, the theory underlying the Multivariate Analysis of Variance (i.e. that more than one dependent variable will be tested against one independent variable); One-Way MANOVA will be presented and the method will be constructed. Apart from One-Way ANOVA, this method will analyse more than one dependant variable and one independent variable which has been built to test the mean effect vectors or treatment effect vectors, which is different from the One-Way ANOVA. At this stage, the Wilks' Lambda test method will be given in order to provide an opportunity to test the treatment effect vectors. The Wilks' Lambda will be used both for the F-test as well as Barlett's Approximation. In the situation that where there are a large numbers of samples, the Bartlett's Approximation test method will be used to analyse the data instead of the F-test. If the null hypothesis is rejected, in order to distinguish between the different treatment effect vectors, the Bonferroni Approach will be adopted in order to enable one to find out simultaneous confidence intervals. The following part of this section will enable us to express the sum of residual squares and products (W); furthermore, it will enable one to obtain Sample Variance-Covariance Matrices. On the other hand, in order to apply the One-Way MANOVA method, the Box's M test theory will be used to analyse the assumption of covariance equality.

In the section 5, the Randomised Block Design Multivariate Analysis of Variance (RBD MANOVA) method will be introduced briefly. This model will test differences for blocks effect vectors in addition to treatment effect vectors. A similar theory will be expressed in this method and the block effect vectors theory will also consequently be defined. A numerical example will be illustrated.

In the section 6, the data taken from the private exam centre in Turkey will be plugged into the GenStat statistical programme for the One-Way MANOVA model. Then, the data which will have been obtained by GenStat will be commented on. This section will also show covariance matrices as well as the results of the assumptions tests.

And finally, in the seventh and last section, a conclusion will be provided to this master thesis. This conclusion will summarise what has been done throughout the whole paper.

2 A Review of One-Way ANOVA

The One-Way ANOVA method is a univariate analysis of variance which is also known as Completely Randomised Designed (CRD). Using this method, we would like to analyse t number of treatments which are the levels of a single factor. We assume that n_i units take treatment i for $i = 1, 2, ..., t$, $\sum_{i=1}^{t} n_i = n$; furthermore, we also suppose that all values X_{i1},X_{i2},...,X_{in_i} are random samples from $N(\mu_i, \sigma^2)$, where $i = 1, 2, ..., t$, $\mu_i = \mu + \alpha_i$ is $i - th$ the population mean and that the random samples are independent. Here, we want to measure a response x_{ij} on the $j - th$ unit of the $i - th$ treatment. In this way, the main One-Way ANOVA model will be

$$X_{ij} = \mu + \alpha_i + e_{ij}, \tag{1}$$

where μ is the overall mean, α_i is the $i - th$ population (treatment) effect, and that e_{ij}, identically independently distributed as $N(0, \sigma^2)$, has random errors with least square estimates and with constraint condition of $\sum_{i=1}^{t} n_i\alpha_i = 0$. We want to test whether the null hypothesis is different from its treatments,

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_t$$

contra the alternative hypothesis H_1 : that there is at least one of the treatments differ. With the construction of equation (1), the analysis of the variance model will be based on a corresponding, separated form of the observations data,

$$x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i). \tag{2}$$

In this newly founded equation, the definition of the components which, x_{ij}, are observations, $\bar{x}$ is the overall sample mean, which is an estimate of μ, $(\bar{x}_i - \bar{x})$ is the estimated treatment effect and which is an estimate of α_i, and $(x_{ij} - \bar{x}_i)$ are the estimated errors (residuals) that is an estimate of e_{ij}.

The decomposition of sums of squares (SS) apportions variability onto the combined samples in the mean, treatment, and errors components. An analysis of variance proceeds by the comparison of the relative size of S_T is the sum of squares for treatment effects. S_E is the sum of squares for residuals. If the null hypothesis, H_0, is true, the variance computed from the S_T and S_E is needed to be approximately equal. If $\bar{x}$ is subtracted from equation (2) and it is both sides are squared, it is going to give us the following new equation, which it will be yielded

$$(x_{ij} - \bar{x})^2 = (\bar{x}_i - \bar{x})^2 + (x_{ij} - \bar{x}_i)^2 + 2(\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) \tag{3}$$

when both sides of equation (3) are summed over j, since $\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i) = 0$ will give

$$\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = n_i(\bar{x}_i - \bar{x})^2 + \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2. \tag{4}$$

And then, summing both sides of equation (4) over i will give us the last form of the following equation:

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = \sum_{i=1}^{t} n_i(\bar{x}_i - \bar{x})^2 + \sum_{i=1}^{t}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 \tag{5}$$

$$S_G = S_T + S_E$$

S_G is the Total (Corrected) SS, S_T is the Between (Samples) SS, and S_E is the Within (Samples) SS.

2.1 One-Way ANOVA Table and F-test

For the One-Way ANOVA Table, the following observations assume that
$n_1 = n_2 = ... = n_t$

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n_1} \\ x_{21} & x_{22} & \dots & x_{2n_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \dots & x_{tn_t} \end{bmatrix}.$$

We will use the following algorithm under these conditions:

1-) $S_T = \sum_{i=1}^{t} n_i(\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{t} \frac{T_i^2}{n_i} - \frac{G^2}{n}$

2-) $S_G = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{G^2}{n}$

3-) $S_E = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{t} \frac{T_i^2}{n_i}$

The calculations were as follows:

1-) Treatment Totals $T_1 = \sum_{j=1}^{n_1} x_{1j}, ..., T_t = \sum_{j=1}^{n_i} x_{tj}$

2-) Grand Total $G = \sum_{i=1}^{t} T_i$

3-) Correction Factor is $\frac{G^2}{n}$

4-) Treatment Sum of Squares $S_T = \sum_{i=1}^{t} \frac{T_i^2}{n_i} - \frac{G^2}{n}$

5-) Total Sum of Squares (Corrected) $S_G = \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{G^2}{n}$

6-) Residual (Error) Sum of Squares $S_E = S_G - S_T = \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ij}^2 - \sum_{i=1}^{t} \frac{T_i^2}{n_i}$.

To decide whether the treatments generated different effects, the One-Way Anova Table needs to be applied and checked against the F-test value. Johnson and Wichern (2010) have shown that the sum of squares of treatments S_T has $t-1$ df, the sum of squares of residuals (errors) S_E has $(n_1 + n_2 + ... + n_t) - t = (n - t)$ df, and, lastly, that the total number of degrees of freedom for the sum of squares of the grand total of the corrected form S_G is equal to $(n-1)$. When univariate distribution theory is considered, these degrees of freedom for chi-square distributions are related to the corresponding sums of squares. The calculations of sums of squares and the associated degrees of freedom are sensibly obtained by an ANOVA Table. In this table, we have the means squares for treatments (M_T) and for the errors (M_E).

One-Way ANOVA Table				
Source of Variation	SS	df	MS	F
Treatment	S_T	$t-1$	$M_T = \frac{S_T}{t-1}$	$F_T = \frac{M_T}{M_E}$
Residual (Error)	S_E	$n-t$	$M_E = \frac{S_E}{n-t}$	
Total (Corrected)	S_G	$n-1$		

Table-1

From this table, the usual $F - test$ rejects the
$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_t$ in favour of the alternative hypothesis H_1 : that, if at least one of the treatments is different, it will be accepted at level α if,

$$F_T = \frac{M_T}{M_E} = \frac{\frac{S_T}{t-1}}{\frac{S_E}{n-t}} > F_{(t-1),(n-t),(\alpha)},$$

where $F_{(t-1),(n-t),(\alpha)}$ is the upper $100(\alpha) - th$ percentile of the F distribution with $t-1$ and $n-t$ degrees of freedom. This is equivalent to rejecting H_0 for large values of $\frac{S_T}{S_E}$ or for large values of $1 + \frac{S_T}{S_E}$. The statistic appropriate for a multivariate generalisation rejects H_0 and accepts H_1 for small values of the reciprocal,

$$\frac{1}{1 + \frac{S_T}{S_E}} = \frac{S_E}{S_E + S_T}. \quad (6)$$

2.2 One-Way ANOVA Example

Example -1)

Group 1 = 9, 6, 9

Group 2 = 0, 2

Group 3 = 3, 1, 2

Our model for this question is equation (1), but here we will use equation (2) in order to obtain a random sample. Therefore, the model equation will be

$$x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i),$$

where $i = 1, 2, 3$ and $j = 1, 2, ..., n_i$ which are $n_1 = 3$, $n_2 = 2$ and $n_3 = 3$. $\bar{x}$ is an overall sample mean, $(\bar{x}_i - \bar{x})$ is an estimated treatment effect, and $(x_{ij} - \bar{x}_i)$

are estimated random sample errors which are assumed identically independently distributed as $N(0, \sigma^2)$. Now, we apply the algorithm respectively:

1-) Treatment Totals $T_1 = \sum_{j=1}^{3} x_{1j} = 9 + 6 + 9 = 24, T_2 = \sum_{j=1}^{2} x_{2j} = 0 + 2 = 2,$
$T_3 = \sum_{j=1}^{3} x_{3j} = 3 + 1 + 2 = 6$
2-) Grand Total $G = \sum_{i=1}^{3} T_i = 24 + 2 + 6 = 32$
3-) Correction Factor is $\frac{G^2}{n} = \frac{32^2}{8} = 128$
4-) Treatment Sum of Squares
$S_T = \sum_{i=1}^{3} \frac{T_i^2}{n_i} - \frac{G^2}{n} = \frac{T_1^2}{3} + \frac{T_2^2}{2} + \frac{T_3^2}{3} - 128 = \frac{32^2}{3} + \frac{2^2}{2} + \frac{6^2}{3} - 128 = 78$
5-) Total Sum of Squares (Corrected)
$S_G = \sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{G^2}{n} = (9^2 + 6^2 + 9^2 + 0^2 + 2^2 + 3^2 + 1^2 + 2^2) - 128 = 88$
6-) Residual(Error) Sum of Squares $S_E = S_G - S_T = 88 - 78 = 10$.

To test the null hypothesis, we will use the One-Way ANOVA Table and apply an F-Test in order to decide whether we should reject the null hypothesis,

One-Way ANOVA Table				
Source of Variation	SS	df	MS	F
Treatment	78	2	$78/2 = 39$	$39/2 = 19.5$
Residual (Error)	10	5	$10/5 = 2$	
Total (Corrected)	88	7		

Table-2

We test for differences between the treatments of the null hypothesis
$H_0 : \alpha_1 = \alpha_2 = \alpha_3$ against that of the alternative hypothesis H_1 : that at least one of treatments differs from the other treatments. For the test at the 1% level of significance, we obtained the F value of $F_T = 19.5$. Therefore, with respect to the F-test statistic,

$$F_T = 19.5 > F_{2,5,(0.01)} = 13.27.$$

With this result, we rejected the null hypothesis $H_0 : \alpha_1 = \alpha_2 = \alpha_3$ and accept the alternative hypothesis H_1: that at least one of the treatments has different effect at the 0.01 significance level.

2.3 Pairwise Comparisons

After rejecting the null hypothesis, when we want to distinguish between the treatments, this form of analysis can be used to test treatments in pairs. For this, we have assumptions such that there are t treatments with replicate with n_i and which stipulate that, in order to use this method, one should firstly reject the null hypothesis. For two groups, we will have i and j treatments and we will have $n-t$ residual degrees of freedom and also $M_E = \frac{S_E}{n-t}$, which is called a residual mean square. This can be derived from the One-Way ANOVA Table. Under these circumstances, the null hypothesis yields $\alpha_i = \alpha_j$ or $\alpha_i - \alpha_j = 0$ for any pairs of treatment i and j; thus, we will reject the null hypothesis in favour of the alternative hypothesis $H_1 : \alpha_i - \alpha_j \neq 0$ if

$$|T| = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{M_E(\frac{1}{n_i} + \frac{1}{n_j})}} > t_{n-t,\frac{\alpha}{2}}$$

.

$\bar{x}_1 = 24/3 = 8$, $\bar{x}_2 = 2/2 = 1$ and $\bar{x}_3 = 6/3 = 2$; $n_1 = 3$, $n_2 = 2$ and $n_3 = 3$; $t = 3$, $n = 8$ and $M_E = 2$ derived from the One-Way ANOVA Table.

For Groups 1 and 2, the pairwise comparison which will be utilised is

$$|T| = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{M_E(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{|8-1|}{\sqrt{2(\frac{1}{3} + \frac{1}{2})}} = 5.422.$$

For Groups 1 and 3, the pairwise comparison which will be utilised is

$$|T| = \frac{|\bar{x}_1 - \bar{x}_3|}{\sqrt{M_E(\frac{1}{n_1} + \frac{1}{n_3})}} = \frac{|8-2|}{\sqrt{2(\frac{1}{3} + \frac{1}{3})}} = 5.196.$$

And, for Groups 2 and 3, the pairwise comparison which will be utilised is

$$|T| = \frac{|\bar{x}_2 - \bar{x}_3|}{\sqrt{M_E(\frac{1}{n_2} + \frac{1}{n_3})}} = \frac{|1-2|}{\sqrt{2(\frac{1}{2} + \frac{1}{3})}} = 0.774.$$

We will test $t_{n-t,\frac{\alpha}{2}} = t_{5,\frac{0.01}{2}} = 4.032$ at the 1 percent level of significance for components 1,2 and 1,3 since they are different treatments. There is no different effect, however, for components 2,3.

3 Randomised Block Design for ANOVA

Randomised Block Design is another method of ANOVA. It is also known as Two-Way ANOVA without interactions. Using this method, we will test t treatments in a similar way to One-Way ANOVA. In addition, we will also test whether the number of blocks has an effect on the block factor. The blocks will enable us to make a more sensitive analysis in terms of the differences between the treatments. For the randomised block design, we suppose that there are b blocks for each size t. Thus, the total number of observations is $n = tb$. The t treatments in each block were selected randomly. That is why treatment i replicates as follows: $n_i = b$. Randomised Block Design yields the following model:

$$X_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \tag{7}$$

for $i = 1, 2, ..., t$ and $j = 1, 2, ..., b$. This equation shares similar assumptions with the One-Way ANOVA. Besides these assumptions, blocks have been added. Here, x_{ij} are the responses to the $i - th$ treatment for the $j - th$ block, μ is the overall mean, α_i is the $i - th$ population (treatment) effect, β_j is the effect of $j - th$ block, and e_{ij} which are both identically and independently distributed as $N(0, \sigma^2)$, are random errors with least square estimates with constraint conditions $\sum_{i=1}^{t} \alpha_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$. In this model, we would like to test two null hypotheses. The first of them is similar to the One-Way ANOVA treatments effect; i.e.

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_t.$$

Here too, though, we judged in favour of the alternative hypothesis H_1 : that at least one of the treatment effects differ from α_i. The second of these null hypotheses effected blocks such that

$$H_0' : \beta_1 = \beta_2 = ... = \beta_b,$$

whereas we also had to judge in favour of the alternative hypothesis H_1' : that at least one of the blocks' effects differ from β_j. Concentrating on the decomposition (7), the method of randomised block design is based on a corresponding breakdown of the assumptions held regarding the treatments and blocks:

$$x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (\bar{x}_j - \bar{x}) + (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}) \tag{8}$$

in which x_{ij} are observations, $\bar{x}$ is the overall sample mean, which is an estimate of μ, $(\bar{x}_i - \bar{x})$ is the estimated treatment effect, which is an estimate of α_i, $(\bar{x}_j - \bar{x})$ is the estimated block effect, which is an estimate of β_j, and $(x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})$ are the estimated errors (residuals), which are an estimate of e_{ij}. Analogously to the One-Way ANOVA, we subtract $\bar{x}$ from equation (8) and square both sides, thereby obtaining:

$$(x_{ij} - \bar{x})^2 = (\bar{x}_i - \bar{x})^2 + (\bar{x}_j - \bar{x})^2 + (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 +$$

$$2[(\bar{x}_i - \bar{x})(\bar{x}_j - \bar{x}) + (\bar{x}_i - \bar{x})(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}) + (\bar{x}_j - \bar{x})(x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})].$$

We sum both sides over i and j, using $\sum_{i=1}^{t}(\bar{x}_i - \bar{x}) = 0$ and $\sum_{j=1}^{b}(\bar{x}_j - \bar{x}) = 0$. Similarly to the One-Way ANOVA calculations and assumptions, the decomposition of the sums of the squares apportions variability to the combined samples in the mean, treatments, blocks and residual (error) components. An analysis of the variance proceeds by comparing the relative sizes of S_T with S_E. If the null hypothesis H_0 is true, the variance computed from S_T and S_E need to be approximately equal. If the null hypothesis H_0' is true, the variance computed from $S_{Blocks(B)}$ and S_E need to be approximately equal. Subtracting $\bar{x}$ from equation (8) and squaring both sides give us the following equation:

$$\sum_{i=1}^{t}\sum_{j=1}^{b}(x_{ij} - \bar{x})^2 = \sum_{i=1}^{t} b(\bar{x}_i - \bar{x})^2 + \sum_{j=1}^{b} t(\bar{x}_j - \bar{x})^2 + \sum_{i=1}^{t}\sum_{j=1}^{b}(x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2. \tag{9}$$

$$S_G = S_T + S_B + S_E$$

S_G is the Total (Corrected) SS, S_T is the Between (Treatments) SS,
S_B is the Between (Blocks) SS and S_E is the Within (Treatments and Blocks) SS.

3.1 ANOVA Table and F-test for Randomised Block Design

The Randomised Block Design ANOVA is similar to the One-Way ANOVA Table and F test. We will also add block components to the Table. The null hypothesis for the effects of the blocks on the treatments effect will be tested. This is done in order to determine whether the treatments and blocks have different effects. The ANOVA Table for Randomised Block Design needs to be applied and checked for F-test value.

Blocks						
		1	2	...	b	Total
	1	x_{11}	x_{12}	...	x_{1b}	T_1
	2	x_{21}	x_{22}	...	x_{2b}	T_2
Treatments	.	.	.	.	.	.
	.	.	.	.	.	.
	t	x_{t1}	x_{t2}	...	x_{tb}	T_t
Total		B_1	B_2	...	B_b	G

Table-3

For the Randomised Block Design, the ANOVA Table will be defined as an algorithm similar to the One-Way ANOVA algorithm. This will enable us to make easier calculations. We will use the following algorithm with the following conditions:

1-) $S_T = \sum_{i=1}^{t} b(\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{t} \frac{T_i^2}{b} - \frac{G^2}{n}$

2-) $S_B = \sum_{j=1}^{b} t(\bar{x}_j - \bar{x})^2 = \sum_{i=1}^{t} \frac{B_j^2}{t} - \frac{G^2}{n}$

3-) $S_G = \sum_{i=1}^{t}\sum_{j=1}^{b}(x_{ij} - \bar{x})^2 = \sum_{i=1}^{t}\sum_{j=1}^{b} x_{ij}^2 - \frac{G^2}{n}$

4-) $S_E = \sum_{i=1}^{t}\sum_{j=1}^{b}(x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = \sum_{i=1}^{t}\sum_{j=1}^{b} x_{ij}^2 - \sum_{i=1}^{t}\frac{T_i^2}{b} - \sum_{j=1}^{b}\frac{B_j^2}{t} + \frac{G^2}{n}.$

The calculations will be as follows:

1-) Treatment Totals $T_1 = \sum_{j=1}^{n_1} x_{1j}, ..., T_t = \sum_{j=1}^{n_t} x_{tj}$

2-) Block Totals $B_1 = \sum_{i=1}^{t} x_{i1}, ..., B_b = \sum_{i=1}^{t} x_{ib}$

3-) Grand Total $G = \sum_{i=1}^{t} T_i = \sum_{j=1}^{b} B_j$

4-) Correction Factor is $\frac{G^2}{n}$

5-) Treatment Sum of Squares $S_T = \sum_{i=1}^{t} \frac{T_i^2}{b} - \frac{G^2}{n}$

6-) Block Sum of Squares $S_B = \sum_{j=1}^{b} \frac{B_i^2}{t} - \frac{G^2}{n}$

7-) Total Sum of Squares (Corrected) $S_G = \sum_{i=1}^{t}\sum_{j=1}^{b} x_{ij}^2 - \frac{G^2}{n}$

8-) Residual (Error) Sum of Squares
$S_E = S_G - S_T - S_B = \sum_{i=1}^{t}\sum_{j=1}^{b} x_{ij}^2 - \sum_{i=1}^{t} \frac{T_i^2}{b} - \sum_{j=1}^{b} \frac{B_i^2}{t} + \frac{G^2}{n}.$

Analogously, S_T has $t-1$ df, S_B has $b-1$ df S_E has $(t-1)(b-1)$ df, and, lastly, the total (corrected) number of degrees of freedom is equal to $tb-1 = n-1$. When univariate distribution theory is considered, these degrees of freedom for the chi-square distributions relate to the corresponding sums of squares. The calculations of the sums of the squares and the associated degrees of freedom are sensibly obtained by the ANOVA Table.

Randomised Block Design ANOVA Table				
Source of Variation	SS	df	MS	F
Treatments	S_T	$t-1$	$M_T = \frac{S_T}{t-1}$	$F_T = \frac{M_T}{M_E}$
Blocks	S_B	$b-1$	$M_B = \frac{S_B}{b-1}$	$F_B = \frac{M_B}{M_E}$
Residual (Error)	S_E	$(t-1)(b-1)$	$M_E = \frac{S_E}{n-t}$	
Total (Corrected)	S_G	$n-1$		

Table-4

The usual $F-tests$ reject the null hypothesis $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_t$, at the α level if

$$F_T = \frac{M_T}{M_E} = \frac{\frac{S_T}{t-1}}{\frac{S_E}{(t-1)(b-1)}} > F_{(t-1),(t-1)(b-1),(\alpha)}$$

where $F_{(t-1),(t-1)(b-1),(\alpha)}$, where (α) is the upper $100(\alpha) - th$ percentile of the F distribution with $t-1$ df for the treatments and $(t-1)(b-1)$ df for the error terms.

This is equivalent to rejecting H_0 for large values of $\frac{S_T}{S_E}$. It can be concluded in favour of the alternative hypothesis H_1 that the treatment effect means that at least one of the α_i differs, in a similar way to the blocks effect. The $F-test$ rejects the null hypothesis $H_0' : \beta_1 = \beta_2 = ... = \beta_b$, at level α if

$$F_B = \frac{M_B}{M_E} = \frac{\frac{S_B}{b-1}}{\frac{S_E}{(t-1)(b-1)}} > F_{(b-1),(t-1)(b-1),(\alpha)}$$

where $F_{(t-1),(t-1)(b-1),(\alpha)}$, and where (α) is the upper $100(\alpha)-th$ percentile of the F distribution with $b-1$ df for the blocks and $(t-1)(b-1)$ df for the error terms. This is equivalent to rejecting H_0' for large values of $\frac{S_B}{S_E}$. Again, one can conclude in favour of the alternative hypothesis H_1' for the blocks effect means that at least one of the β_j's differ.

3.2 A Numerical Example

Example-2) In this example, we want to test the weight gain in bear cubs under five years of age's diets. A randomised block design was used in which there were four litters, with each of the five cubs in each litter being randomly assigned to a different diet. The data, together with the treatment and block totals, are given below.

Litters						
		1	2	3	4	Total
Diets	1	6	4	10	4	24
	2	8	3	11	2	24
	3	9	5	9	5	28
	4	5	2	8	1	16
	5	8	3	7	4	22
Total		36	17	45	16	114

Table-5

We have the treatment totals $T_1 = 24$, $T_2 = 24$, $T_3 = 28$, $T_4 = 16$ and $T_5 = 22$, the block totals $B_1 = 36$, $B_2 = 17$, $B_3 = 45$ and $B_4 = 16$, and the grand total $G = 114$. According to the algorithm, the Correction Factor is $\frac{G^2}{n} = \frac{114^2}{20}$

$S_T = \frac{24^2}{4} + \frac{24^2}{4} + \frac{28^2}{4} + \frac{16^2}{4} + \frac{22^2}{4} - \frac{114^2}{20} = 19.2$

$S_B = \frac{36^2}{5} + \frac{17^2}{5} + \frac{45^2}{5} + \frac{16^2}{5} - \frac{114^2}{20} = 123.4$

$S_G = 6^2 + 8^2 + ... + 4^2 - \frac{114^2}{20} = 160.2$

$S_E = S_G - S_T - S_B = 160.2 - 19.2 - 123.4 = 17.6.$

The ANOVA Table will be as follows:

Randomised Block Design ANOVA Table				
Source of Variation	SS	df	MS	F
Diets	19.2	4	4.8	3.27
Litters	123.4	3	41.13	28.05
Residual (Error)	17.6	12	1.467	
Total (Corrected)	160.2	19		

Table-6

In order to test for differences between the treatments at the 5% level of significance, the value of $F_T = 3.27$ is compared with $F_{4,12,0.05} = 3.259$, which can be looked up in Table 12(b) of the New Cambridge Statistical Tables. Since F_T is greater than $F_{4,12,0.05}$, we reject the null hypothesis of no differences at the 5% level. Similarly, since $F_B = 28.05$ is greater than $F_{3,12,0.05} = 3.490$, we also reject H_0' : of there being no effects due to there being litters at the 5% level of significance.

4 One-Way MANOVA

Comparing Several Multivariate Population Mean Vectors is called a One-Way MANOVA in which generally two or more variables can be compared. Johnson and Wichern (2010) proposed that random samples, collected from each of the t populations, should be arranged as follows:

$$\begin{bmatrix} Population-1: \underline{X}_{11}, \underline{X}_{12}, ..., \underline{X}_{1n_1} \\ Population-2: \underline{X}_{21}, \underline{X}_{22}, ..., \underline{X}_{2n_2} \\ ... \\ Population-t: \underline{X}_{t1}, \underline{X}_{t2}, ..., \underline{X}_{tn_t} \end{bmatrix}$$

First of all, the One-Way MANOVA is used to analyse whether the population mean vectors are the same and, if not, which mean components differ significantly. Assumptions regarding the Structure of the Data for One-Way MANOVA are as follows:

$1-)$ $\underline{X}_{i1}, \underline{X}_{i2}, ..., \underline{X}_{in_i}$, are random samples of size n_i from a population with a mean vector $\underline{\mu}_i$, where $i = 1, 2, ..., t$. The random samples from the different populations are independent.

$2-)$ All populations have a common covariance matrix Σ

$3-)$ Each population is multivariately normal.

Based upon these assumptions specifically, the third condition can be expanded by applying to it the central limit theorem when the sample size n_i is large. We can define the One-Way MANOVA method in a similar way to how we defined the One-Way ANOVA method. The differences which exist between the One-Way ANOVA and the One-Way MANOVA models are that the ANOVA components are scalar while the MANOVA terms are vectors. Thus, based upon this difference, all terms in MANOVA have $p \times 1$ vector columns. The aim of the One-Way MANOVA model is that of comparing the t population mean vector or treatment vectors in a similar way to the One-Way ANOVA model,

$$\underline{X}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{e}_{ij}, \tag{10}$$

where $i = 1, 2, ..., t,\ j = 1, 2, ..., n_i$ and $\underline{e}_{ij}$ are independent vector terms $N_p(0, \Sigma)$. Here, $\underline{\mu}$ is an overall vector mean, and $\underline{\alpha}_i$ represents the $i - th$ treatment effect vector's terms with $\sum_{i=1}^{t} n_i \underline{\alpha}_i = 0$. According to the model in (10), each term of the observation vector $\underline{X}_{ij}$ satisfies equation (1). The error vectors for the components of $\underline{X}_{ij}$ are correlated, but the covariance matrix Σ is the same for all populations. A vector's observations may be decomposed as suggested by the model. Thus,

$$\underline{x}_{ij} = \underline{\bar{x}} + (\underline{\bar{x}}_i - \underline{\bar{x}}) + (\underline{x}_{ij} - \underline{\bar{x}}_i) \tag{11}$$

$\underline{x}_{ij}$ are the observation vectors, $\underline{\bar{x}}$ is the overall sample mean vectors for $(\underline{\hat{\mu}})$, $(\underline{\bar{x}}_i - \underline{\bar{x}})$ are the estimated treatment effect vector $(\underline{\hat{\alpha}}_i)$, and $(\underline{x}_{ij} - \underline{\bar{x}}_i)$ are residual (error) vector terms $(\underline{\hat{e}}_{ij})$. The decomposition of equation (11) leads to the multivariate analogy of breakup of the univariate sum of the squares given in equation (3). First we note that the product

$$(\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})'$$

can be written as follows

$$(\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})' = [(\underline{x}_{ij} - \underline{\bar{x}}_i) + (\underline{\bar{x}}_i - \underline{\bar{x}})][(\underline{x}_{ij} - \underline{\bar{x}}_i) + (\underline{\bar{x}}_i - \underline{\bar{x}})]'$$

$$= (\underline{x}_{ij} - \underline{\bar{x}}_i)(\underline{x}_{ij} - \underline{\bar{x}}_i)' + (\underline{x}_{ij} - \underline{\bar{x}}_i)(\underline{\bar{x}}_i - \underline{\bar{x}})' + (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}}_i)' + (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})'.$$

The sum over j in the middle of the two expressions is the zero matrix, because $\sum_{j=1}^{n_i} (\underline{x}_{ij} - \underline{\bar{x}}_i) = 0$. Hence, summing the cross product over i and j yields

$$\sum_{i=1}^{t} \sum_{j=1}^{n_i} (\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})' =$$

$$= \sum_{i=1}^{t} n_i (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})' + \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\underline{x}_{ij} - \underline{\bar{x}}_i)(\underline{x}_{ij} - \underline{\bar{x}}_i)'. \tag{12}$$

$\sum_{i=1}^{t} \sum_{j=1}^{n_i} (\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})'$ is the Total (Corrected) Sum of Squares and Cross Products,
$\sum_{i=1}^{t} n_i (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})'$ is the Treatment (Between) Sum of Squares and Cross Products,
$\sum_{i=1}^{t} \sum_{j=1}^{n_i} (\underline{x}_{ij} - \underline{\bar{x}}_i)(\underline{x}_{ij} - \underline{\bar{x}}_i)'$ is the Residuals (Within) Sum of Squares and Cross Products. The within sum of squares and cross product matrix can be expressed by

$$W = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}}_i)(\underline{x}_{ij} - \bar{\underline{x}}_i)'$$
$$= (n_1 - 1)S_1 + (n_2 - 1)S_2 + ... + (n_t - 1)S_t, \quad (13)$$

where S_i is the sample covariance matrix for the $i-th$ sample. It plays a dominant role in testing for the presence of treatment effects. Analogous to the univariate result, the hypothesis of no treatment effects,

$$H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = ... = \underline{\alpha}_t$$

is tested by considering the relative sizes of the treatment and residual sums of the squares and cross products. Equivalently, we may consider the relative size of the residual and total (corrected) sum of squares and cross products. Formally, we summarise the calculations leading to the test statistic in a MANOVA table.

One-Way MANOVA Table		
Source of Variation	Matrix of sum of squares and Cross Products (SSCP)	df
Treatment	$B = \sum_{i=1}^{t} n_i(\bar{\underline{x}}_i - \bar{\underline{x}})(\bar{\underline{x}}_i - \bar{\underline{x}})'$	$t-1$
Residual (Error)	$W = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}}_i)(\underline{x}_{ij} - \bar{\underline{x}}_i)'$	$n-t$
Total (Corrected)	$B + W = \sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}})(\underline{x}_{ij} - \bar{\underline{x}})'$	$n-1$

Table-7

This table shares almost the same form as the One-Way ANOVA Table, term by term. The differences between the Tables are that the squares of scalars are changed by their vector representations. To obtain the this matrix form of this One-Way MANOVA Table, the following algorithm will be followed just like the one for the One-Way ANOVA. Firstly, we will calculate the sum of the squares for each $k-th$ dependent variable where $k = 1, 2, ..., p$ and find that the Calculations will be the terms for each independent variable. Here x_{ijk} are $k-th$ replicate for $i-th$ treatment and the $j-th$ block.

1-) Treatment Totals $T_{1k} = \sum_{j=1}^{n_1} x_{1jk}, ..., T_{tk} = \sum_{j=1}^{n_t} x_{tjk}$

2-) Grand Total $G_k = \sum_{i=1}^{t} T_{ik}$

3-) Correction Factor is $\frac{G_k^2}{n}$

4-) Treatment Sum of Squares $S_{T_k} = \sum_{i=1}^{t} \frac{T_{ik}^2}{n_i} - \frac{G_k^2}{n}$

5-) Total Sum of Squares (Corrected) $S_{G_k} = \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ijk}^2 - \frac{G_k^2}{n}$

6-) Residual (Error) Sum of Squares $S_{E_k} = S_{G_k} - S_{T_k} = \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ijk}^2 - \sum_{i=1}^{t} \frac{T_{ik}^2}{n_i}$.

Afterwards, we will find the rest of the calculations between the dependent variables, such as components k and k', to be different; such that:

1-) The Correction Factor is $\frac{(G_k)(G_{k'})}{n}$

2-) The Treatment Sum of the Products are $SP_{T_{kk'}} = \sum_{i=1}^{t} \frac{(T_{ik})(T_{ik'})}{n_i} - \frac{(G_k)(G_{k'})}{n}$

3-) The Total Sum of the Products (Corrected) $SP_{G_{kk'}} = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (x_{ijk})(x_{ijk'}) - \frac{(G_k)(G_{k'})}{n}$

4-) The Residual Sum of the Products are $SP_{E_{kk'}} = SP_{G_{kk'}} - SP_{T_{kk'}}$.

If these calculations are applied to all of the dependent variables for each pair, the One-Way MANOVA Table will be obtained as mentioned.

4.1 The Wilks' Lambda Test

According to Crichton (2000), the Wilks' Lambda (λ^*) is a test statistic used in multivariate analyses of variance (MANOVAs) to test whether there are differences between the means of the identified groups of subjects when combined to dependent variables. The Wilks' Lambda performs, in a multivariate setting, with a combination of dependent variables - the same role as the F-test performs in a one-way analysis of variance. The Wilks' Lambda is a direct measure of the proportion of variance in the combination of dependent variables that is unaccounted for by the independent variable (i.e. the grouping variable or factor). The Wilks' Lambda is obtained by applying the method of maximum likelihood. According to Johnson and Wichern (1982) have shown that the method of maximum likelihood can be expressed by

$$\lambda^* = (\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|})^{n/2} = (\frac{|\sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}}_i)(\underline{x}_{ij} - \bar{\underline{x}}_i)'|}{|\sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}})(\underline{x}_{ij} - \bar{\underline{x}})'|})^{n/2} < c_\alpha.$$

As a result of this equation, the null hypothesis will be accepted at the level α per cent significance level.

4.2 The Wilks' Lambda Test for One-Way MANOVA

In order to analyse the comparisons of several multivariate treatments of vector means, testing the null hypothesis $H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = ... = \underline{\alpha}_t$ involved generalising variances. The null hypothesis will be rejected when the proportion of generalised variances of λ^* is too small.

$$\lambda^* = \frac{|W|}{|B+W|} = \frac{|\sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}}_i)(\underline{x}_{ij} - \bar{\underline{x}}_i)'|}{|\sum_{i=1}^{t}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \bar{\underline{x}})(\underline{x}_{ij} - \bar{\underline{x}})'|}. \quad (14)$$

The value of λ^*, as proposed by Wilks, corresponds to an equivalent form of equation (14). For a different case, λ^* can be used according to the lists provided in Table-8.

$\lambda^* = \frac{\|W\|}{\|B+W\|}$, Distribution of The Wilks' Lambda		
Number of Variables	Number of Groups	Sampling distribution for Multivariate Normal Data
$p = 1$	$t \geq 2$	$(\frac{n-t}{t-1})(\frac{1-\lambda^*}{\lambda^*}) \sim F_{(t-1),(n-t)}$
$p = 2$	$t \geq 2$	$(\frac{n-t-1}{t-1})(\frac{1-\sqrt{\lambda^*}}{\sqrt{\lambda^*}}) \sim F_{2(t-1),2(n-t-1)}$
$p \geq 1$	$t = 2$	$(\frac{n-p-1}{p})(\frac{1-\lambda^*}{\lambda^*}) \sim F_{p,(n-p-1)}$
$p \geq 1$	$t = 3$	$(\frac{n-p-2}{p})(\frac{1-\sqrt{\lambda^*}}{\sqrt{\lambda^*}}) \sim F_{2p,2(n-p-2)}$

Table-8

The Wilks' Lambda can be obtained as a product of eigenvalues which can be obtained by the eigenvalues of the matrix of $W^{-1}B$ by following method

$$\lambda^* = \prod_{i=1}^{k}(\frac{1}{1+\hat{\lambda_i}}),$$

where $k = min(p, t-1)$ and the rank of the B matrix and the expression $\hat{\lambda_1}, \hat{\lambda_2}, ..., \hat{\lambda_k}$ are eigenvalues of the $W^{-1}B$ matrix.

Example-3 (Wilks' Lambda for analysing the equality of three mean vectors in a MANOVA Table)

When more variables are added to (Example-1) with the sample size, $n_1 = 3$, $n_2 = 2$ and $n_3 = 3$, arranging the observations, produce the following structure:

$$\begin{bmatrix} \begin{bmatrix} 9 \\ 3 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} \begin{bmatrix} 9 \\ 7 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 4 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 3 \\ 8 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \end{bmatrix} \begin{bmatrix} 2 \\ 7 \end{bmatrix} \end{bmatrix}.$$

We have already calculated all the steps of the algorithm for the first variable in the One-Way ANOVA section for Example 1. According to the calculations of the findings, we will do for the other variable. Between the first and the second calculations, and with the following results, we will add notation of the $k-th$ variable in order to define the difference between variables;

$$\begin{bmatrix} 9 & 6 & 9 \\ 0 & 2 & \\ 3 & 1 & 2 \end{bmatrix}$$

for the first variable.

1-) The Treatment Totals $T_{11} = 24, T_{21} = 2$ and $T_{31} = 6$

2-) The Grand Total $G_1 = 32$

3-) The Correction Factor is 128

4-) The Treatment Sum of Squares $S_{T_1} = 78$

5-) The Total Sum of Squares (Corrected) $S_{G_1} = 88$

6-) Residual (Error) Sum of Squares $S_{E_1} = S_{G_1} - S_{T_1} = 88 - 78 = 10.$

For the second component, when we applied a similar method to the ANOVA method, the following form will be obtained as

$$\begin{bmatrix} 3 & 2 & 7 \\ 4 & 0 & \\ 8 & 9 & 7 \end{bmatrix}$$

1-) The Treatment Totals $T_{12} = \sum_{j=1}^{3} x_{1j2} = 3+2+7 = 12, T_{22} = \sum_{j=1}^{2} x_{2j2} = 4+0 = 4,$ $T_{32} = \sum_{j=1}^{3} x_{3j2} = 8+9+7 = 24$

2-) The Grand Total $G_2 = \sum_{i=1}^{3} T_{i2} = 12 + 4 + 24 = 40$

3-) The Correction Factor is $\frac{G_2^2}{n} = \frac{40^2}{8} = 200$

4-) The Treatment Sum of Squares
$S_{T_2} = \sum_{i=1}^{3} \frac{T_{i2}^2}{n_i} - \frac{G_2^2}{n} = \frac{T_{12}^2}{3} + \frac{T_{22}^2}{2} + \frac{T_{32}^2}{3} - 200 = \frac{12^2}{3} + \frac{4^2}{2} + \frac{24^2}{3} - 200 = 48$

5-) The Total Sum of Squares (Corrected)
$S_{G_2} = \sum_{i=1}^{3} \sum_{j=1}^{n_i} x_{ij2}^2 - \frac{G_2^2}{n} = (3^2 + 2^2 + 7^2 + 4^2 + 0^2 + 8^2 + 9^2 + 7^2) - 200 = 72$

6-) The Residual (Error) Sum of Squares $S_{E_2} = S_{G_2} - S_{T_2} = 72 - 48 = 24.$

The results of the first and the second variables results will differ respectively. The Treatment Totals are $T_{11} = 24$, $T_{21} = 2$ and $T_{31} = 6$. The Grand Total $G_1 = 32$. The Treatment Totals $T_{12} = 12, T_{22} = 4$ and $T_{32} = 24$. The Grand Total $G_2 = 40$. We will apply these two components to the algorithm for the Sum of Products

1-) The Correction Factor is $\frac{(G_k)(G_{k'})}{n} = \frac{(32)(40)}{8} = 160$

2-) The Treatment Sum of Products are
$SP_{T_{kk'}} = \sum_{i=1}^{t} \frac{(T_{ik})(T_{ik'})}{n_i} - \frac{(G_k)(G_{k'})}{n} = \frac{(24)(12)}{3} + \frac{(2)(4)}{2} + \frac{(6)(24)}{3} - 160 = -12$

3-) The Total Sum of Products (Corrected)

$SP_{G_{kk'}} = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\sum_{k=1}^{p}(x_{ijk})(x_{ijk'}) - \frac{(G_k)(G_{k'})}{n} = [(9)(3)+(6)(2)+...+(2)(7)]-160 = -11$

4-) The Residual Sum of Products are $SP_{E_{kk'}} = SP_{G_{kk'}} - SP_{T_{kk'}} = -11-(-12) = 1$

After all of these results are figured, we will obtain the following One-Way MANOVA Table as:

Table for MANOVA which analyses Population Mean Vectors		
Source of Variation	Matrix of sum of squares and Cross Products (SSP)	df
Treatment	$B = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix}$	$3-1=2$
Residual (Error)	$W = \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}$	$8-3=5$
Total (Corrected)	$B+W = \begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix}$	$8-1=7$

Table-9

From equation (12), we can see that

$$\begin{bmatrix} 88 & -11 \\ -11 & 72 \end{bmatrix} = \begin{bmatrix} 78 & -12 \\ -12 & 48 \end{bmatrix} + \begin{bmatrix} 10 & 1 \\ 1 & 24 \end{bmatrix}.$$

In equation (14) can be found Wilks' Lambda is

$$\lambda^* = \frac{|W|}{|B+W|} = \frac{\begin{vmatrix} 10 & 1 \\ 1 & 24 \end{vmatrix}}{\begin{vmatrix} 88 & -11 \\ -11 & 72 \end{vmatrix}} = \frac{10(24)-1^2}{88(72)-(-11)^2} = \frac{239}{6215} = 0.0385.$$

To test if there is a treatment effect on this example when using the One-Way Manova, we will use the data in Table-8 to formulate the null hypothesis. Thus,

$$H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3.$$

This means that there is no treatment effect against H_1 : that there is at least one of the $\underline{\alpha}_i \neq 0$. In this example, we have $p = 2$ and $t = 3$ populations which, in turn, means that we will have to apply the second row of Wilks's Lambda formula at F-distribution and test it at a 1% level of significance.

$p = 2$, $t \geq 2$ and

$$F = (\frac{n-t-1}{t-1})(\frac{1-\sqrt{\lambda^*}}{\sqrt{\lambda^*}}) \sim F_{2(t-1),2(n-t-1)}.$$

The F-distribution has $v_1 = 2(t-1) = 2(3-1) = 4$ and $v_2 = 2(n-t-1) = 2(8-3-1) = 8$ degrees of freedom and

$$F = (\frac{8-3-1}{3-1})(\frac{1-\sqrt{0.0385}}{\sqrt{0.0385}}).$$

$F = 8.16$ and $F_{4,8,0.01} = 7.01$ means that $F = 8.16 > F_{4,8,0.01} = 7.01$. We reject the null hypothesis $H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3$ and accepted H_1 : that there is at least one of the $\underline{\alpha}_i \neq 0$.

4.3 Bartlett's Approximation to One-Way MANOVA

As in the previous section, we might have a chance to use another kind of test approximation introduced by Bartlett, M.S. (1954) using a chi-square test instead of an F-distribution test. Bartlett's test is a modification of the corresponding likelihood ratio test designed to make the approximation of the χ^2 distribution better at all stages. It will be run on the previous example. We will only use the following test:

$$-(n-1-\frac{p+t}{2})(\ln\lambda^*) \sim \chi^2_{p(t-1)}.$$

With this method of hypothesis, this test will enable us to decide if there is a different effect.

$$-(n-1-\frac{p+t}{2})(\ln\lambda^*) = -(8-1-\frac{2+3}{2})(\ln 0.0385) = 14.6569$$

$$\chi^2_{2(3-1),0.01} = \chi^2_{4,0.01} = 13.28$$

According to the result of this test, the null hypothesis was rejected in favour of the alternative hypothesis - i.e the same result which was obtained as the previous F-test.

4.4 Find Pooled Sample Variance-Covariance Matrix

To get the W matrix, equation (13) will need to be used, where S_i is the sample covariance matrix for the $i-th$ sample as below:

$$S_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \underline{\bar{x}}_i)(\underline{x}_{ij} - \underline{\bar{x}}_i)'. \tag{15}$$

The pooled sample covariance matrix can be written in terms of W divided by residual degrees of freedom

$$S = \frac{\sum\limits_{i=1}^{t}(n_i - 1)S_i}{\sum\limits_{i=1}^{t}(n_i - 1)} = \frac{W}{n - t} \tag{16}$$

Example-4)(A Multivariate Analysis of Wisconsin Nursing Home Data)

The aim of the Wisconsin Department of Health and Social Services is to analyse what impact ownership (private party, non-profit organisation, and government) has on the following four costs: the cost of nursing labour, the cost of dietary labour, the cost of plant operation and maintenance labour, and the cost of housekeeping and laundry labour. The total number of observations are 516, with $p = 4$ variables divided according to the ownership of those three groups.

Table for MANOVA which analyses Population Mean Vectors

Groups	Number of Observations	Sample Means Vectors
1-(private)	$n_1 = 271$	
2-(nonprofit)	$n_2 = 138$	$\underline{\bar{x}}_1 = \begin{bmatrix} 2.066 \\ 0.48 \\ 0.82 \\ 0.36 \end{bmatrix}, \underline{\bar{x}}_2 = \begin{bmatrix} 2.167 \\ 0.596 \\ 0.124 \\ 0.418 \end{bmatrix}, \underline{\bar{x}}_3 = \begin{bmatrix} 2.273 \\ 0.521 \\ 0.125 \\ 0.383 \end{bmatrix}$
3-(govern)	$n_3 = 107$	

Table-10

The sample covariance matrices are as follows:

$$S_1 = \begin{bmatrix} 0.291 & & & \\ -0.001 & 0.011 & & \\ 0.002 & 0.000 & 0.001 & \\ 0.01 & 0.003 & 0.000 & 0.010 \end{bmatrix}, S_2 = \begin{bmatrix} 0.561 & & & \\ 0.011 & 0.025 & & \\ 0.001 & 0.004 & 0.005 & \\ 0.037 & 0.007 & 0.002 & 0.019 \end{bmatrix},$$

and

$$S_3 = \begin{bmatrix} 0.261 & & & \\ 0.030 & 0.017 & & \\ 0.003 & -0.000 & 0.004 & \\ 0.018 & 0.006 & 0.001 & 0.013 \end{bmatrix}.$$

The method of pooling the sample covariance matrix from equation (13) will give W as

$$W = (n_1 - 1)S_1 + (n_2 - 1)S_2 + (n_3 - 1)S_3$$

$$W = \begin{bmatrix} 182.962 & & & \\ 4.408 & 8.200 & & \\ 1.695 & 0.633 & 1.484 & \\ 9.581 & 2.428 & 0.394 & 6.538 \end{bmatrix}.$$

We find that the overall mean vector is

$$\underline{\bar{x}} = \frac{n_1\underline{\bar{x}}_1 + n_2\underline{\bar{x}}_2 + n_3\underline{\bar{x}}_3}{n_1 + n_2 + n_3} = \begin{bmatrix} 2.136 \\ 0.519 \\ 0.102 \\ 0.380 \end{bmatrix}$$

and that B will be

$$B = \sum_{i=1}^{3} n_i(\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})' = \begin{bmatrix} 3.475 & & & \\ 1.111 & 1.225 & & \\ 0.821 & 0.453 & 0.235 & \\ 0.584 & 0.610 & 0.230 & 0.304 \end{bmatrix}.$$

Here, the null hypothesis for the treatment effect vectors is

$$H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3,$$

which will test whether there is no treatment effect against the alternative hypothesis H_1 : that there is at least one treatment that has different effects for $i \neq j$ at $\alpha_i \neq \alpha_j$

$$\lambda^* = \frac{|W|}{|B+W|} = 0.7714.$$

Since we have $p = 4$ variable and $t = 3$ populations, that means that we will apply the fourth row of the Wilks' Lambda formula at F-distribution and test at the 1% significance level. $p = 4$, $t = 3$ and

$$F = (\frac{n-p-2}{p})(\frac{1-\sqrt{\lambda^*}}{\sqrt{\lambda^*}}) \sim F_{2p,2(n-p-2)}.$$

F-distribution has $v_1 = 2p = 2(4) = 8$ and $v_2 = 2(n-p-2) = 2(516-4-2) = 1020$ degrees of freedom at

$$F = (\frac{516-4-2}{4})(\frac{1-\sqrt{0.7714}}{\sqrt{0.7714}}) \sim F_{8,1020}$$

$F = 17.67$ and $F_{8,1020,0.01} = 2.51$, which means $F = 17.67 > F_{8,1020,0.01} = 2.51$.
We reject the null hypothesis $H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3$ and accepted H_1 : that there is at least one of the $\underline{\alpha}_1$ or $\underline{\alpha}_2$ or $\underline{\alpha}_3$ treatments which differs for any of the costs. Barlett's Approximation has been applied to Example-4

$$-(n-1-\frac{p+t}{2})(\ln\lambda^*) = -(516-1-\frac{4+3}{2})(\ln 0.7714) = 132.76$$

$$\chi^2_{4(3-1),0.01} = \chi^2_{8,0.01} = 20.09.$$

With this result, the null hypothesis will be rejected in favour of the alternative hypothesis. This means that at least one of the costs will have a different effect on the ownership. Using this method, the same conclusion will be obtained based on different test hypotheses.

4.5 Box's M Test for Covariance Assumptions

The One-Way MANOVA method is based on three assumptions. One of these assumptions is defined in the One-Way MANOVA section which states that all populations should have a common covariance matrix $\sum$. According to Box (1949), before using the One-Way MANOVA method, it is necessary to demonstrate that this assumption ought to be true. For this, our null hypothesis is

$$H_0 : \sum\nolimits_1 = \sum\nolimits_2 = ... = \sum\nolimits_t = \sum, \tag{17}$$

where $\sum_i$ is the covariance matrix for the $i-th$ population, with $i = 1, 2, ..., t$ and $\sum$ is the common covariance matrix. The alternative hypothesis H_1 : that there is at least one covariance matrix that differs from the others. Under a likelihood ratio, the statistic condition for multivariate normal populations with which to test equation (17) of the null hypothesis is

$$\lambda = \prod_{i=1} (\frac{|S_i|}{|S|})^{(\frac{n_i-1}{2})}. \tag{18}$$

Here, n_i is the sample size for the $i-th$ group, S_i is the sample covariance matrix for the $i-th$ group, and S is the pooled sample covariance matrix as defined in (16), as follows:

$$S = \frac{1}{\sum_i (n_i - 1)}[(n_1 - 1)S_1 + (n_2 - 1)S_2 + ... + (n_t - 1)S_t].$$

The Box is tested based on this χ^2 approximation for the sampling distribution of $-2\ln\lambda = M$ is the Box's M statistics; these are stipulated as follows:

$$M = \sum\nolimits_i (n_i - 1)\ln|S| - \sum\nolimits_i (n_i - 1)\ln|S_i|, \tag{19}$$

The Box's test for equality amongst covariance matrices will use the Box's M statistics as follows:

$$D = [\sum\nolimits_i \frac{1}{n_i - 1} - \frac{1}{\sum_i (n_i - 1)}]\frac{2p^2 + 3p - 1}{6(p+1)(t-1)}. \tag{20}$$

Here, p is the number of dependant variables and t is the number of groups. The equation will be formed as follows:

$$E = (1 - D)M = (1 - D)[\sum_i (n_i - 1)] \ln |S| - \sum_i [(n_i - 1) \ln |S_i|], \qquad (21)$$

this equation will have an approximate χ^2 distribution with
$v = t\frac{1}{2}p(p+1) - \frac{1}{2}p(p+1) = \frac{1}{2}p(p+1)(t-1)$ degrees of freedom. The null hypothesis will be rejected at the α level of significance and if $E > \chi^2_{\frac{p(p+1)(t-1)}{2},(\alpha)}$. If the null hypothesis is not rejected, it can be used as a One-Way MANOVA to analyse the treatment effects. This approximation will be appropriate with p and t, which are not more than 5 each. For each n_i, however, 20 more are needed. Otherwise, the F approximation will be more appropriate for the sampling distribution of M. We use the results of example-4 in order to test the equality of the covariance matrices for the $p = 4$ variable and the $t = 3$ groups and we analyse

$$H_0 : \sum_1 = \sum_2 = \sum_3 = \sum.$$

We have $n_1 = 271$, $n_2 = 138$ and $n_3 = 107$; $|S_1| = 2.783\text{x}10^{-8}$, $|S_2| = 89.539\text{x}10^{-8}$, $|S_3| = 14.579\text{x}10^{-8}$ and $|S| = 17.398\text{x}10^{-8}$; then, the natural logarithm of the determinants will give us the $\ln |S_1| = -17.397$, $\ln |S_2| = -13.926$, $\ln |S_3| = -15.741$ and $\ln |S_{pooled}| = -15.564$. We calculate that

$D = (\frac{1}{270} + \frac{1}{137} + \frac{1}{106} - \frac{1}{270+137+106})\frac{2(4^2)+3(4)-1}{6(4+1)(3-1)} = 0.0133$
$M = (270 + 137 + 106)(-15.564) - (270(-17.397) + 137(-13.926) + 106(-15.741))$
$= 289.3$
$E = (1 - 0.0133)289.3 = 285.5$

the degree of freedom for E $v = \frac{1}{2}p(p + 1)(t - 1) = \frac{1}{2}4(4 + 1)(3 - 1) = 20$ with χ^2 distribution. From this result, it can be concluded that the null hypothesis will be rejected and at least one of the covariance matrices which differ from the others will be accepted for example-4. In light of this conclusion, it can be said that the covariance matrices' equality from the assumptions will not be met. That is why this method is not exactly appropriate for these data.

4.6 The Bonferroni Approach to Treatment Effects

Since the null hypothesis is rejected in example 4, we can test which treatments differ from one another in this part. In this section, we will use the Bonferroni approach to build simultaneous confidence intervals to test treatment in the pairs. For pairwise comparisons, the differences between the treatments will be used to analyse which is $(\underline{\alpha}_i - \underline{\alpha}_l)$ or $(\underline{\mu}_i - \underline{\mu}_l)$. We will test intervals at a shorter level than obtained for all contrasts. We assume that α_{ik} is the $k-th$ component of $\underline{\alpha}_i$ because $\underline{\hat{\alpha}}_i$ is estimated as $\underline{\hat{\alpha}}_i = \underline{\bar{x}}_i - \underline{\bar{x}}$

$$\hat{\alpha}_{ik} = \bar{x}_{ik} - \bar{x}_k. \tag{22}$$

Here, $\hat{\alpha}_{ik} - \hat{\alpha}_{lk}$ is the difference between the two independent sample means with

$$Var(\hat{\alpha}_{ik} - \hat{\alpha}_{lk}) = Var(\bar{X}_{ik} - \bar{X}_{lk}) = (\frac{1}{n_i} + \frac{1}{n_l})\sigma_{kk}.$$

Here, σ_{kk} is the $k-th$ diagonal term of covariance matrix $\sum$. From equation (16), $Var(\bar{X}_{ik} - \bar{X}_{lk})$ can be estimated by dividing within the sum of squares component (W) with its degrees of freedom. This will be as follows:

$$\widehat{Var}(\bar{X}_{ik} - \bar{X}_{lk}) = (\frac{1}{n_i} + \frac{1}{n_l})\frac{w_{kk}}{n-t}$$

the k-th diagonal element of W and will be expressed as s_k^2, which is the diagonal element of $\frac{W_{kk}}{n-t}$ and $n = n_1 + n_2 + ... + n_t$. We will have p variables and $\frac{t(t-1)}{2}$ pairwise differences. For each sample, two-sample t-interval will apply the critical value of the $t_{n-t}(\frac{\alpha}{2m})$, with $m = \frac{pt(t-1)}{2}$ being the number of simultaneous confidence intervals. Here, $\hat{\alpha}_{ik} - \hat{\alpha}_{lk}$ will belong to

$$\bar{x}_{ik} - \bar{x}_{lk} \pm t_{n-t}(\frac{\alpha}{pt(t-1)})\sqrt{s_k^2(\frac{1}{n_i} + \frac{1}{n_l})}, \tag{23}$$

where $k = 1, 2, ..., p$ for all variables and all difference terms will be $l < i = 1, 2, ..., t$, where w_{kk} is the diagonal element of W. If the simultaneous confidence interval includes a zero, we will accept that there is no treatment effect. Otherwise, if the interval does not include a zero, we will reject the null hypothesis that there is no treatment effect and will accept that there is a different treatment effect.

Examle-5 (This is continued from Example-4)

In this example, we would like to construct a simultaneous confidence interval for the treatment effects based on the data of example-4. To compare the X_3 variable, which is the cost of plant operation and maintenance labour between privately owned nursing homes and government-owned nursing homes, can be tested as $\hat{\alpha}_{13} - \hat{\alpha}_{33}$ by using (23) and a third component of diagonal W from the data from example-4. The data used in Example-4 will give us

$$\underline{\hat{\alpha}}_1 = (\underline{\bar{x}}_1 - \underline{\bar{x}}) = \begin{bmatrix} -0.070 \\ -0.039 \\ -0.020 \\ -0.020 \end{bmatrix}, \underline{\hat{\alpha}}_3 = (\underline{\bar{x}}_3 - \underline{\bar{x}}) = \begin{bmatrix} 0.137 \\ 0.002 \\ 0.023 \\ 0.003 \end{bmatrix}.$$

From these results, we can find $\hat{\alpha}_{13} - \hat{\alpha}_{33} = -0.020 - 0.023 = -0.043$ belongs to

$$\bar{x}_{13} - \bar{x}_{33} \pm t_{n-t}(\frac{\alpha}{pt(t-1)})\sqrt{s_3^2(\frac{1}{n_1} + \frac{1}{n_3})},$$

with $n = 516$, $p = 4$, and $t = 3$ for 95% of the simultaneous statements we had:

$$\bar{x}_{13} - \bar{x}_{33} \pm t_{513}(\frac{0.05}{4.3(2)})\sqrt{s_3^2(\frac{1}{n_1} + \frac{1}{n_3})}$$

$$-0.043 \pm 2.87(0.00614) = (-0.061, -0.025).$$

Based on these results, it can be concluded that the average maintenance and labour cost for government-owned nursing homes is higher from 0.025 to 0.061 hour per patient day than for privately owned ones, which means that the interval does not include zero. We conclude that there are different treatment effects. Based on the fact that $\hat{\alpha}_{13} - \hat{\alpha}_{23}$ belongs to $(-0.058, -0.026)$ and that it, therefore, does not include a zero, we conclude that there is a difference between the treatment effect. Furthermore, since $\hat{\alpha}_{23} - \hat{\alpha}_{33}$ belongs to $(-0.021, 0.019)$, and since this interval includes a zero, we can conclude that there does not exist any treatment effect. Therefore, it can be said that there is a difference in this cost between private and non-profit nursing homes (on the assumption that there is no difference between non-profit and government nursing homes).

5 Randomised Block Design MANOVA

In this chapter, we will show the method for designing a randomised block vector for multivariate populations. The randomised block design for ANOVA is similar to the One-Way MANOVA method. A Two-Way ANOVA (without interactions) with test blocks effect, in addition to treatments, populations as well. Here, we will call this model a Two-Way MANOVA (without interactions). It tests block effect vectors at the same time as treatment effect vectors. This will facilitate us to make a more sensitive analysis for identifying the differences between the treatment effect vectors. The Randomised Block Design model for MANOVA will be analogous to equation (8); namely:

$$\underline{X}_{ij} = \underline{\mu} + \underline{\alpha}_i + \underline{\beta}_j + \underline{e}_{ij}, \tag{24}$$

where $i = 1, 2, ..., t,\ j = 1, 2, ..., b,\ \underline{e}_{ij}$ are independent error vector terms $N_p(0, \Sigma)$ which are all identically independently distributed. $\underline{\mu}$ is an overall mean vector, $\underline{\alpha}_i$ is the $i - th$ treatment effect vector, $\underline{\beta}_j$ is the $j - th$ block effect vector component with conditions, including $\sum_{i=1}^{t} \underline{\alpha}_i = 0$, $\sum_{j=1}^{b} \underline{\beta}_j = 0$. From the model demonstrated in (24), each term of the observation vector $\underline{X}_{ij}$ satisfies, as usual, equation (7) in the univariate randomised block testing for ANOVA. The error vectors for the components of $\underline{X}_{ij}$ are correlated, but the covariance matrix Σ is the same for all populations. Decomposing the vector observations is suggested by the model. Thus,

$$\underline{x}_{ij} = \underline{\bar{x}} + (\underline{\bar{x}}_i - \underline{\bar{x}}) + (\underline{\bar{x}}_j - \underline{\bar{x}}) + (\underline{x}_{ij} - \underline{\bar{x}}_i - \underline{\bar{x}}_j + \underline{\bar{x}}). \tag{25}$$

The decomposition of equation (25) leads to the multivariate analogy of the univariate sum of squares breakup of equation (8). First, we note that the product,

$$(\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})'$$

can be written as follows:

$$(\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})' =$$

$$= [(\underline{\bar{x}}_i - \underline{\bar{x}}) + (\underline{\bar{x}}_j - \underline{\bar{x}}) + (\underline{x}_{ij} - \underline{\bar{x}}_i - \underline{\bar{x}}_j + \underline{\bar{x}})][(\underline{\bar{x}}_i - \underline{\bar{x}}) + (\underline{\bar{x}}_j - \underline{\bar{x}}) + (\underline{x}_{ij} - \underline{\bar{x}}_i - \underline{\bar{x}}_j + \underline{\bar{x}})]'$$

$$= (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})' + (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_j - \underline{\bar{x}})' + (\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}}_i - \underline{\bar{x}}_j + \underline{\bar{x}})' +$$

$$+(\underline{\bar{x}}_j-\underline{\bar{x}})(\underline{\bar{x}}_i-\underline{\bar{x}})'+(\underline{\bar{x}}_j-\underline{\bar{x}})(\underline{\bar{x}}_j-\underline{\bar{x}})'+(\underline{\bar{x}}_j-\underline{\bar{x}})(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})'$$

$$(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})(\underline{\bar{x}}_i-\underline{\bar{x}})'+(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})(\underline{\bar{x}}_j-\underline{\bar{x}})'+(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})'.$$

Summing the cross product over i and j yields the following conditions: $\sum_{i=1}^{t}(\underline{\bar{x}}_i-\underline{\bar{x}})=0$, $\sum_{j=1}^{b}(\underline{\bar{x}}_j-\underline{\bar{x}})=0$ will give equation,

$$\sum_{i=1}^{t}\sum_{j=1}^{b}(\underline{x}_{ij}-\underline{\bar{x}})(\underline{x}_{ij}-\underline{\bar{x}})'=\sum_{i=1}^{t}b(\underline{\bar{x}}_i-\underline{\bar{x}})(\underline{\bar{x}}_i-\underline{\bar{x}})'+\sum_{j=1}^{b}t(\underline{\bar{x}}_j-\underline{\bar{x}})(\underline{\bar{x}}_j-\underline{\bar{x}})'+$$

$$+\sum_{i=1}^{t}\sum_{j=1}^{b}(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})(\underline{x}_{ij}-\underline{\bar{x}}_i-\underline{\bar{x}}_j+\underline{\bar{x}})'. \tag{26}$$

It is by means of this method that we will test the null hypothesis for the treatment effect vectors in a similar way to the One-Way MANOVA. In addition to the treatments, we will also analyse the null hypothesis for block effect vectors. For the treatment effects, we will have

$$H_0:\underline{\alpha}_1=\underline{\alpha}_2=...=\underline{\alpha}_t.$$

For block effects,

$$H_0':\underline{\beta}_1=\underline{\beta}_2=...=\underline{\beta}_b$$

is tested by considering the relative sizes of the treatments, the blocks, and the residual sums of squares and cross products. We formally summarise the calculations leading to the test statistic in the MANOVA table for Two-Way MANOVA without interactions.

Randomised Block Design MANOVA Table		
Sources	SSCP	df
Treatments	$B = \sum_{i=1}^{t} b(\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})'$	$t-1$
Blocks	$C = \sum_{j=1}^{b} t(\underline{\bar{x}}_j - \underline{\bar{x}})(\underline{\bar{x}}_j - \underline{\bar{x}})'$	$b-1$
Residual	$W = \sum_{i=1}^{t}\sum_{j=1}^{b}(\underline{x}_{ij} - \underline{\bar{x}}_i - \underline{\bar{x}}_j + \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}}_i - \underline{\bar{x}}_j + \underline{\bar{x}})'$	$(t-1)(b-1)$
Total	$B+C+W = \sum_{i=1}^{t}\sum_{j=1}^{b}(\underline{x}_{ij} - \underline{\bar{x}})(\underline{x}_{ij} - \underline{\bar{x}})'$	$tb-1{=}n-1$

Table-11

Table-11 shares almost exactly the same form as the One-Way MANOVA Table, term by term. The differences between the Tables is the sum of squares and cross-product of block effect vectors which is then added to this Table.

5.1 Wilks' Lambda Test for (RBD) MANOVA

In this section, we would like to analyse the treatment effect vectors with regards to the One-Way MANOVA. We also test whether the block effect vectors differ. The structure of the analysis of the Randomised Block Design for MANOVA is similar to the One-Way MANOVA, although some specific differences do exist. Here, we will have two null hypotheses which are treatments and blocks. For the treatments, the null hypothesis

$$H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = ... = \underline{\alpha}_t$$

involves generalised variances, whereas, for the blocks, we will have

$$H_0' : \underline{\beta}_1 = \underline{\beta}_2 = ... = \underline{\beta}_b.$$

The null hypothesis will be rejected when the proportion of generalised variances of λ^* is too small for treatments and blocks. That is why we will have two different λ^*; i.e. λ_1^* for the treatments and λ_2^* for the blocks.

$$\lambda_1^* = \frac{|W|}{|B+W|}$$

$$\lambda_2^* = \frac{|W|}{|C+W|}$$

5.2 Bartlett's Correction to (RBD) MANOVA

For the test treatment effect vectors and block effect vectors, Bartlett's Correction technique will be used. The likelihood ratio test for the treatment effect vectors is as follows: λ_1^*,

$$-[tb - 1 - \frac{p+t}{2}] \ln \lambda_1^* > \chi^2_{p(t-1),(\alpha)}$$

where $\chi^2_{p(t-1)}(\alpha)$ is the upper $(100\alpha) - th$ percentile of the chi-square distribution with $(t-1)p$ degrees of freedom. The likelihood ratio test for the block effect vectors is as follows: λ_2^*,

$$-[tb - 1 - \frac{p+b}{2}] \ln \lambda_2^* > \chi^2_{p(b-1),(\alpha)}$$

where $\chi^2_{p(b-1)}(\alpha)$ is the upper $(100\alpha) - th$ percentile of chi-square distribution with $(b-1)p$ degrees of freedom.

5.3 The Bonferroni Approach to (RBD) MANOVA

In order to implement this method on the treatments and blocks, the null hypothesis should be rejected. Instead, the Bonferroni Approach can be used to construct simultaneous confidence intervals for pairwise comparisons for each treatment and block effects. A similar approach to the One-Way MANOVA method can be used to build simultaneous intervals $(\underline{\alpha}_i - \underline{\alpha}_l)$ from formula (23)

$$\bar{x}_{i.k} - \bar{x}_{l.k} \pm t_{(t-1)(b-1)}(\frac{\alpha}{pt(t-1)})\sqrt{\frac{w_{kk}}{(t-1)(b-1)}(\frac{1}{b}+\frac{1}{b})}.$$

$\bar{x}_{i.k}$ and $\bar{x}_{l.k}$ are the $k-th$ components for the mean of the treatment effect vectors. As for the block effect vectors, with the help of a similar methodology upon the treatment effect vectors, the block effect vector contrasts can be built like $(\underline{\beta}_i - \underline{\beta}_l)$;

however, there are some differences between the block effect vectors. Again, we have p variables, but this time there are $\frac{b(b-1)}{2}$ pairwise differences

$$\bar{x}_{.rk} - \bar{x}_{.sk} \pm t_{(t-1)(b-1)}(\frac{\alpha}{pb(b-1)})\sqrt{\frac{w_{kk}}{(t-1)(b-1)}(\frac{1}{t}+\frac{1}{t})}.$$

$\bar{x}_{.rk}$ and $\bar{x}_{.sk}$ are the $k-th$ components for the mean of the block effect vectors that are $s < r = 1, 2, ..., b$ and $n = tb$.

5.4 A Numerical Example

In this section, an example will be illustrated for the randomised block design method for MANOVA. The following example, which is provided by Chatfield and Collins (2010), to test the null hypothesis for the treatment effect vectors and block effect vectors. In this example (Example-6),

	Treatments			
Blocks	1	2	3	Total
1	13.3	13.6	14.2	41.1
	10.6	10.2	10.7	31.5
	21.2	21.0	21.1	63.3
2	13.4	13.2	13.9	40.5
	9.4	9.6	10.4	29.4
	21.0	20.1	19.8	60.9
3	12.9	12.2	13.9	39.0
	10.0	9.9	11.0	30.9
	20.5	20.7	19.1	60.3
Total	39.6	39.0	42.0	120.6
	30.0	29.7	32.1	91.8
	62.7	61.8	60.0	184.5

Table-12

Here, the sum of squares (SS) for component 1 will be calculated as follows:

SS (Total) $= 13.3^2 + 13.6^2 + ... + 13.9^2 - 120.6^2/9 = 2.92$

SS (Treatments) $= (39.6^2 + 39.0^2 + 42.0^2)/3 - 120.6^2/9 = 1.68$

SS (Blocks) $= (41.1^2 + 40.5^2 + 39.0^2)3 - 120.6^2/9 = 0.78$

SS (Residuals) = SS (Total) - SS (Treatments) - SS (Blocks)

$= 2.92 - 1.68 - 0.78 = 0.46.$

The Sum of Products (SP) will be obtained for components 1 and 2 as

SP (Total) $= (13.3)(10.6) + (13.6)(10.2) + ... + (13.9)(11.0) - (120.6)(91.8)/9 = 1.44$

SP (Treatmets) $= ((39.6)(30.0) + (39.0)(29.7) + (42.0)(32.1))/3 - (120.6)(91.8)/9$

$= 1.38$

SP (Blocks) $= ((41.1)(31.5) + (40.5)(29.4) + (39.0)(30.9))/3 - (120.6)(91.8)/9$

$= 1.38 = 1.44$

SP (Residuals) = SP (Total) - SP (Treatments) - SP (Blocks)

$= 1.44 - 1.38 - 0.03 = 0.03.$

When all of the calculations are done for all of the components with a similar methodology, the following table is generated.

Table-2		
Source	df	SSPM
Treatments	2	$B = \begin{bmatrix} 1.68 & 1.38 & -1.26 \\ 1.38 & 1.14 & -1.08 \\ -1.26 & -1.08 & 1.26 \end{bmatrix}$
Blocks	2	$C = \begin{bmatrix} 0.78 & 0.03 & 0.96 \\ 0.03 & 0.78 & 0.66 \\ 0.96 & 0.66 & 1.68 \end{bmatrix}$
Residuals	4	$W = \begin{bmatrix} 0.46 & 0.03 & -0.40 \\ 0.03 & 0.30 & -0.48 \\ -0.40 & -0.48 & 1.06 \end{bmatrix}$
Total (Corrected)	8	$B + C + W = \begin{bmatrix} 2.92 & 1.44 & -0.70 \\ 1.44 & 2.22 & -0.90 \\ -0.70 & -0.90 & 4.00 \end{bmatrix}$

Table-13

In this example, we will test the null hypotheses for treatment effect vectors

$$H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3$$

against the alternative hypothesis H_1 : that at least one of the treatments differs from the other treatments. As for the block effect vectors, we will test

$$H_0' : \underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}_3$$

against the alternative hypothesis H_1' : that there is at least one block that differs from the other blocks

$$\lambda_1^* = \frac{|W|}{|B+W|} = 0.0043$$

$$\lambda_2^* = \frac{|W|}{|C+W|} = 0.2015.$$

In this example, we have $p = 3$ variables, $t = 3$ treatments, $b = 3$ blocks and, in total, we have $n = 9$ values, with Bartlett's Correction criteria and a 95% percentile of the chi-square distribution

$$-[(3)(3) - 1 - \frac{3+1-(3-1)}{2}]\ln(0.0043) > \chi^2_{(3-1)3}(0.05)$$

$$22.0259 > 12.59.$$

This is for block effect vectors with λ_2^*

$$-[(3)(3) - 1 - \frac{3+1-(3-1)}{2}]\ln(0.2015) < \chi^2_{(3-1)3}(0.05)$$

$$11.2138 < 12.59.$$

Under these results, we rejected the null hypotheses for the treatments and accepted the alternative hypotheses, that means that at least one of the treatments differ from the other treatments. Since we accepted the null hypothesis for the block effect vectors, though, there is no difference between the blocks. At this stage, in order to distinguish which treatment has a significant effect, we would have to apply simultaneous confidence intervals using the Bonferroni Approach by means of using contrasts.

6 A GenStat Application

In this section, we will use the Genstat programme in order to test the three types of school's net scores in an exam which evaluates Maths, Turkish Language, and English Language Scores by means of the One-Way MANOVA model. Here, an exam was given to three types of school; viz., a private school, a science school and a state school. This exam includes 50 math questions, 30 Turkish language questions, and 20 English language questions. According to exam's regulations, all of the questions have 5 answer options, with each wrong answer cancelling a quarter of the correct answers. The data was taken from a private exam centre in Turkey. 59 student scores were chosen from a private school, 52 student scores were taken from a science school, and 60 student scores were taken from a state school, with a total number of 171 students being observed. We analysed the fixed data in this application because the schools were fixed effect factors, with our one level factor being the type of schools. Our variables were math, Turkish language, and English language scores. When we analysed the data using the One-Way Manova method using Genstat programme, we reached the results illustrated in the following table.

Table for MANOVA, which analyses Population Mean Vectors		
Groups	n	Sample Means Vectors
1-Private	$n_1 = 59$	$\bar{x}_1 = \begin{bmatrix} 32.1059 \\ 20.3559 \\ 14.1314 \end{bmatrix}, \bar{x}_2 = \begin{bmatrix} 43.2308 \\ 24.2308 \\ 15.8125 \end{bmatrix}, \bar{x}_3 = \begin{bmatrix} 23.2375 \\ 15.0500 \\ 11.1875 \end{bmatrix}$
2-Science	$n_2 = 52$	
3-State	$n_3 = 60$	

Table-10

Sample Covariance Matrices which are as following,

$$S_1 = \begin{bmatrix} 14.0974 & & \\ 6.7677 & 10.9573 & \\ 3.7940 & 4.1755 & 4.5288 \end{bmatrix}, S_2 = \begin{bmatrix} 6.4359 & & \\ 2.8501 & 8.2227 & \\ 1.6078 & 2.8027 & 3.6320 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 11.9416 & & \\ 3.8968 & 5.2136 & \\ 4.8212 & 3.6642 & 6.0501 \end{bmatrix}$$

The method of pooling the sample covariance matrix from equation (13) will give W as

$$W = (n_1 - 1)S_1 + (n_2 - 1)S_2 + (n_3 - 1)S_3$$

$$W = \begin{bmatrix} 1850.4 & & \\ 767.8 & 1362.5 & \\ 586.5 & 601.3 & 804.9 \end{bmatrix}.$$

We can find the overall mean vector as being

$$\underline{\bar{x}} = \frac{n_1\underline{\bar{x}}_1 + n_2\underline{\bar{x}}_2 + n_3\underline{\bar{x}}_3}{n_1 + n_2 + n_3} = \begin{bmatrix} 32.8581 \\ 19.8789 \\ 13.7105 \end{bmatrix}$$

with the B,

$$B = \sum_{i=1}^{3} n_i(\underline{\bar{x}}_i - \underline{\bar{x}})(\underline{\bar{x}}_i - \underline{\bar{x}})' = \begin{bmatrix} 11142 & & \\ 5097 & 2390 & \\ 2563 & 1215 & 620 \end{bmatrix}.$$

Here, the null hypothesis for the treatment effect vectors is

$$H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3,$$

which will test whether there exists no treatment effect against the alternative hypothesis H_1 : that there is at least one treatment that has a different effect for $i \neq j$ at $\alpha_i \neq \alpha_j$

$$\lambda^* = \frac{|W|}{|B + W|} = 0.1281.$$

Since we have $p = 3$ variable and $t = 3$ populations, this means that we will apply the fourth row of Wilks's Lambda formula at F-distribution and test at a 1% level of significance. $p = 3$, $t = 3$ and

$$F = (\frac{n-p-2}{p})(\frac{1-\sqrt{\lambda^*}}{\sqrt{\lambda^*}}) \sim F_{2p,2(n-p-2)}.$$

The F-distribution has $v_1 = 2p = 2(3) = 6$ and $v_2 = 2(n-p-2) = 2(171-3-2) = 332$ degrees of freedom at α significance level.

$$F = (\frac{171-3-2}{3})(\frac{1-\sqrt{0.1281}}{\sqrt{0.1281}}) \sim F_{6,332}$$

$F = 99.25$ and $F_{6,,332,0.01} = 2.802$, which means $F = 99.25 > F_{6,332,0.01} = 2.802$. With Bartlett's Approximation,

$$-(n-1-\frac{p+t}{2})(\ln\lambda^*) = -(171-1-\frac{3+3}{2})(\ln 0.1281) = 343.1757$$

$$\chi^2_{6,0.01} = 16.81$$

obtained the same results with the F-test. We reject the null hypothesis $H_0 : \underline{\alpha}_1 = \underline{\alpha}_2 = \underline{\alpha}_3$ and accepted H_1 : that at least one of the $\underline{\alpha}_1$ or $\underline{\alpha}_2$ or $\underline{\alpha}_3$ treatments are different. The test details were given in the Appendix section.

6.1 Bonferroni Approach

To distinguish which type of school differs from which, the Bonferroni Approach's simultaneous confidence intervals will be built for the GenStat application. X_2 will be used for this, $\hat{\alpha}_{12} - \hat{\alpha}_{22}$ using formula (23) and the second component of the diagonal of w_{kk} or s_k from GenStat, results in

$$\hat{\underline{\alpha}}_1 = (\underline{\bar{x}}_1 - \underline{\bar{x}}) = \begin{bmatrix} -0.7522 \\ 0.4770 \\ 0.4209 \end{bmatrix}, \hat{\underline{\alpha}}_2 = (\underline{\bar{x}}_2 - \underline{\bar{x}}) = \begin{bmatrix} 10.3727 \\ 4.3519 \\ 2.1020 \end{bmatrix}.$$

From these results, we can find $\hat{\alpha}_{12} - \hat{\alpha}_{22} = 0.4770 - 4.3519 = -3.8749$, which belongs to

$$\bar{x}_{12} - \bar{x}_{22} \pm t_{n-t}(\frac{\alpha}{pt(t-1)})\sqrt{s_2^2(\frac{1}{n_1}+\frac{1}{n_2})},$$

with $n = 171$, $p = 3$, and $t = 3$ for 99% of the simultaneous statements, such as

$$\bar{x}_{12} - \bar{x}_{22} \pm t_{168}(\frac{0.01}{3.3(2)})\sqrt{(8.1100)^2(\frac{1}{59}+\frac{1}{52})}$$

$$-3.8749 \pm 3.291(1.5426) = (-8.9516, 1.2018).$$

If the interval included a zero, it could be concluded that the Turkish scores did not differ from a private and science high schools.

Since the $\hat{\alpha}_{12} - \hat{\alpha}_{32}$ interval $(0.414, 10.1994)$ does not include zero, we have concluded that there is a difference between the private and state high school with relation to their Turkish scores.

Since the $\hat{\alpha}_{22} - \hat{\alpha}_{32}$ interval $(4.1239, 14.2377)$ also did not include a zero, we can conclude that there is a difference between the science and the state high school in terms of Turkish scores.

6.2 Assumption Tests

The One-Way MANOVA method requires three assumptions, which were mentioned in the One-Way MANOVA section:

1−) $\underline{X}_{i1}, \underline{X}_{i2}, ..., \underline{X}_{in_i}$, are random samples of size n_i from a population with mean $\underline{\mu}_i$, where $i = 1, 2, ..., t$. The random samples from different populations are independent.

2−) All populations have a common covariance matrix Σ.

3−) Each population is multivariately normal.

The first assumption is not required to be demonstrated. However, the second and third assumptions will be applied for the Box M Test and Q-Q plot approximations in the following subsections.

6.2.1 Box M Test

One of the assumptions applies the Box M test for the equality of covariance matrices. We have a null hypothesis

$$H_0 : \sum_1 = \sum_2 = \sum_3 = \sum,$$

with $D = 0.0259$, $M = 25.8638$ and $E = (1 - D)M = 25.1939$.

If $E > \chi^2_{\frac{p(p+1)(t-1)}{2},(\alpha)}$ then the null hypothesis will be rejected.
$\chi^2_{\frac{p(p+1)(t-1)}{2},(\alpha)} = \chi^2_{\frac{3(3+1)(3-1)}{2},(0.01)} = \chi^2_{12,(0.01)} = 26.22.$

We accepted the null hypothesis and concluded that the assumption for the equality of the covariance matrices was met.

6.2.2 Normality

For another assumption, we applied the Q-Q plot for multivariate normality in GenStat programme. It had the following results: for Math, the Turkish Language, and the English Language, the following results have been obtained by running the GenStat application. Subsequently, all of the data is mostly concentrated on a line, but in the beginning and the end, there were some deviations from the line. That is why the Maths, Turkish Language, and English Language data did not meet the normality assumption. All of the information for the data supplied to the GenStat application will be given in the Appendix.

7 Conclusion

This dissertation has introduced how to analyse one independent variable and more than one dependent variable regarding a treatment effect vector by using the One-Way MANOVA method. Appropriate data has been implemented into the GenStat statistical programme which has enabled us to test its results. The entire dissertation has attempted to distinguish whether at least one of the treatment effect vectors differentiate from the rest of them.

In the section 2, the One-Way ANOVA model gave me the opportunity to build a similar One-Way MANOVA model. The difference between the One-Way ANOVA and the One-Way MANOVA methods is the number of dependent variables.

In the section 3, an RBD for the ANOVA was undertaken in order to make more sensitive tests when compared to the One-Way ANOVA. The RBD ANOVA has block effects in addition to the One-Way ANOVA model.

The section 4, on the other hand, expressed the main idea of this Masters thesis. In this section, the One-Way MANOVA model was investigated for the similarities shared with ANOVA. In this section, the theory behind the One-Way MANOVA model was supplied with the help of Wilks's Lambda. Since the One-Way MANOVA can be applied, the Box M test's covariances' equality rules were introduced. Given the large number of the sample, Bartlett's Approximation was also given. After rejecting the null hypothesis for selecting which treatments differed from the others, simultaneous confidence intervals were built with the help of Bonferroni Approach.

In the section 5, an RBD MANOVA was also constructed. It incorporates block effect vectors in addition to treatment effect vectors.

Finally, in the section 6, three different schools' treatment effect vectors (in relation to Math, Turkish Language, and English Language) were carried out using the GenStat statistical programme. This programme found that, there are differences between the different schools which were examined.

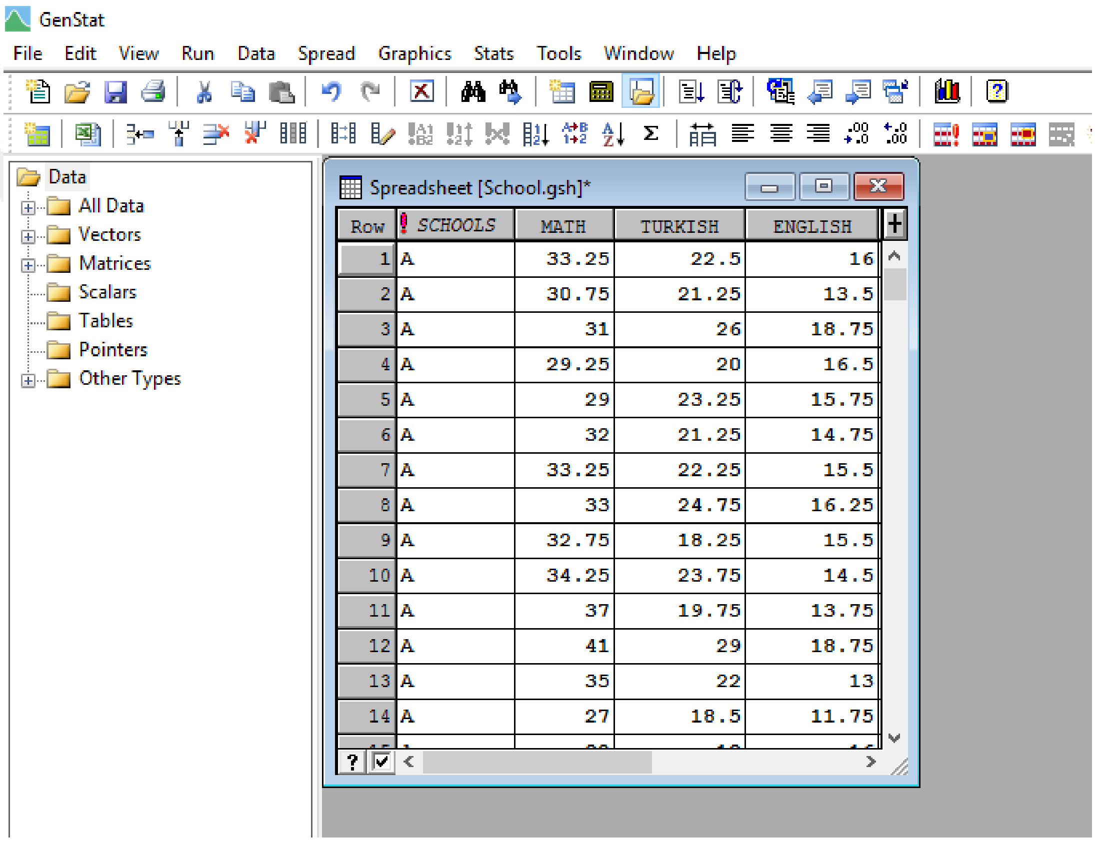
8 References

- Berry, D. A. and Lindgren, B. W. (1996). Statistics: Theory and methods. Duxbury Resource Center.
- Bartlett, M.S. (1954). A Note on the Multiplying Factors for Various chi square Approximations. Journal of the Royal Statistical Society. Series B (Methodological), 296-298.
- Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36, 317-346.
- Chatfield, C. and Collins, A. (1980). Introduction to Multivariate Analysis. Springer Science Business Media.
- Crichton, N. A. (2000). Defining and assessing competence: Issues and debates. Journal of Clinical Nursing, 9(3), 369-381.
- Everitt B.S. and Dunn G. (1991). Applied Multivariate Data Analysis. Edward Arnold. London. pp. 219-220.
- Johnson, R., and Wichern, D. (1982). Applied multivariate statistical analysis. Englewood Cliffs, N.J.: Prentice-Hall.
- Peck, R. and Devore, J. (2011). Statistics: The exploration and analysis of data. Cengage Learning.
- Tabachnick, B., and Fidell, L. (2007). Using multivariate statistics. Boston: Pearson/Allyn and Bacon.
- Zetterberg, P. (2013). Effects of Unbalancedness and Heteroscedasticity on Two-Way MANOVA tests (Master). Stockholm University.

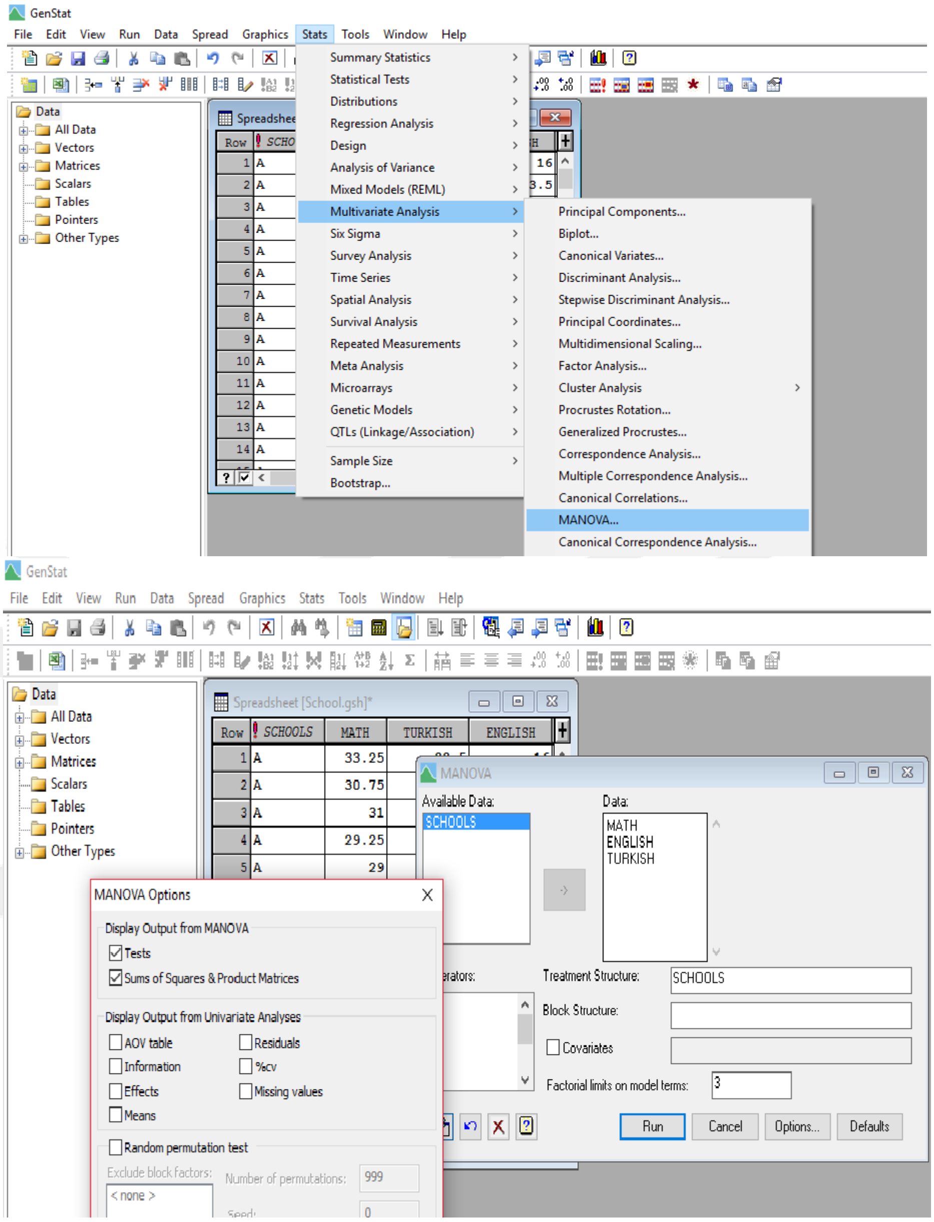
9 Appendix

9.1 GenStat

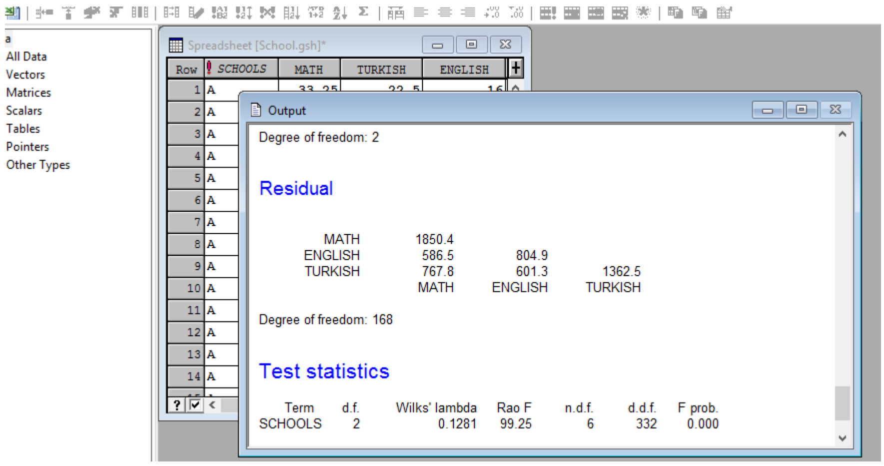
Firstly in GenStat programme all data are entered manually, Schools are converted to factor level clicking the right of the mouse on the Schools column and following steps are run as



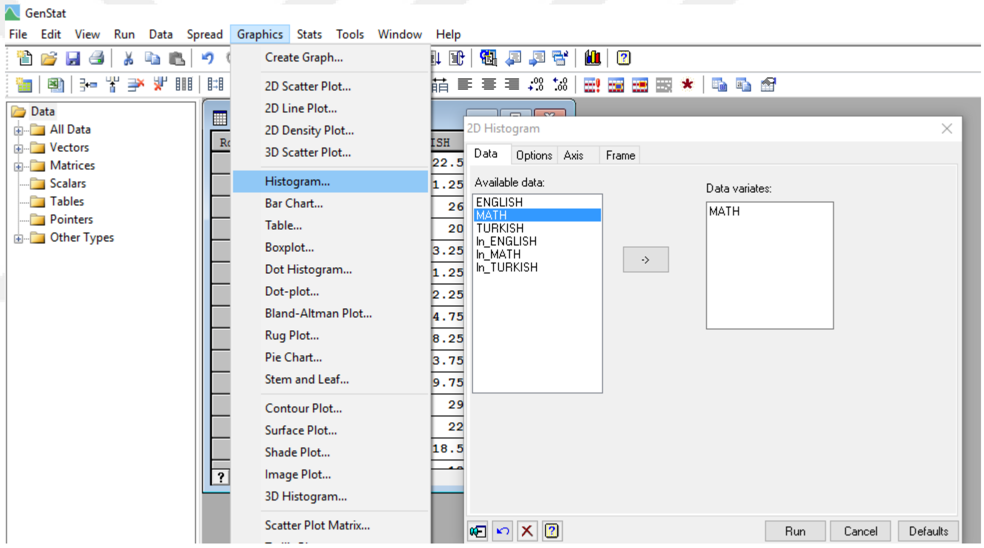
from stat section multivariate is chosen and subsection MANOVA has been clicked. It has been followed by MANOVA with choosing data and treatment structure from options second SSPM box is chosen

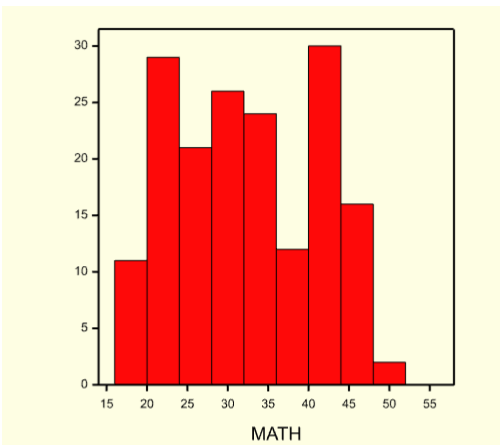


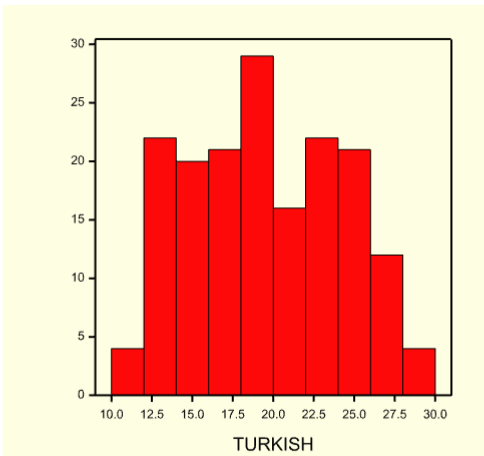
and lastly SSP matrices, Wilks' Lambda result and test results has been obtained by run GenStat program.

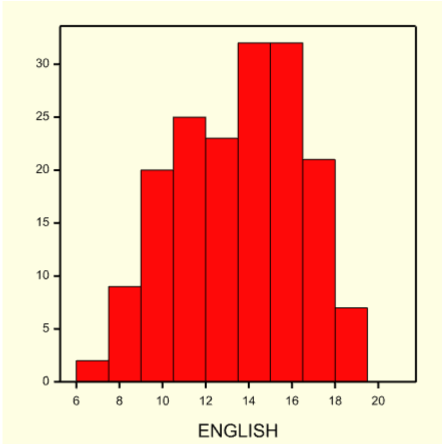


For one of the assumptions test is normality, we will apply histogram analysis and Q-Q plot test following,

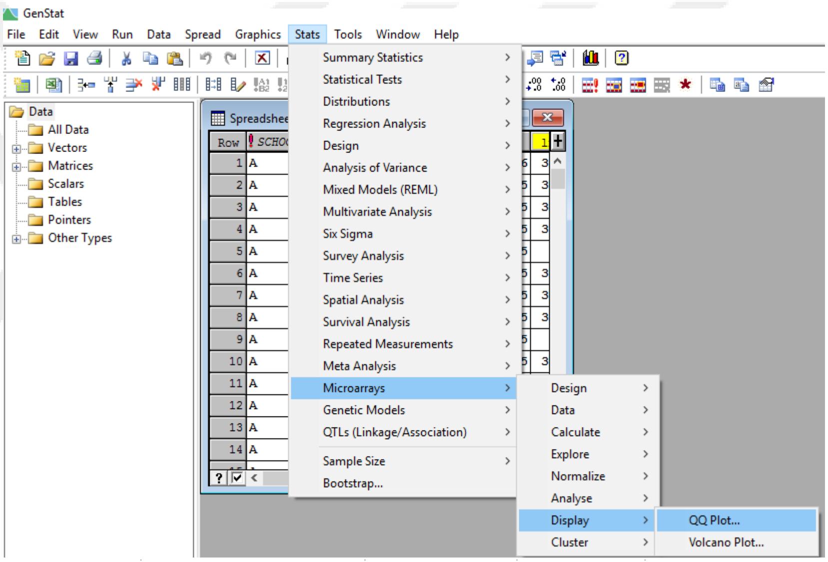


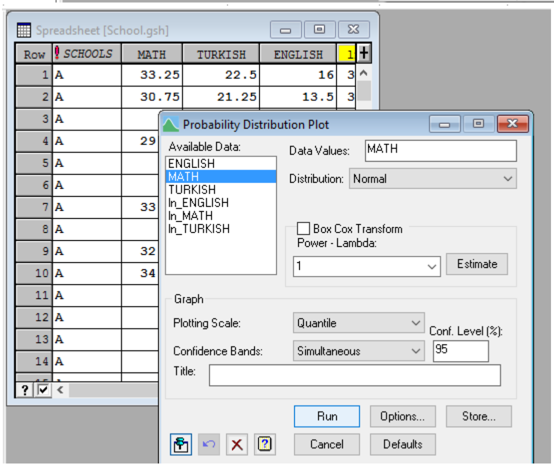


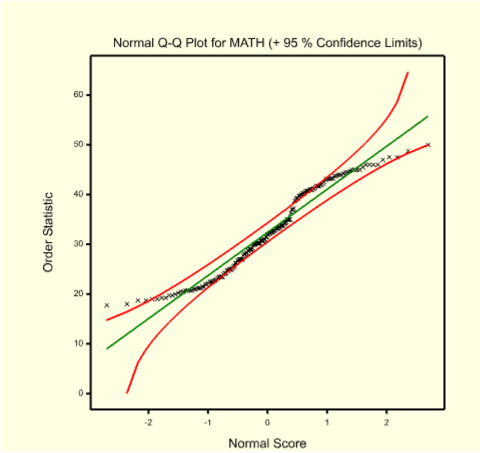


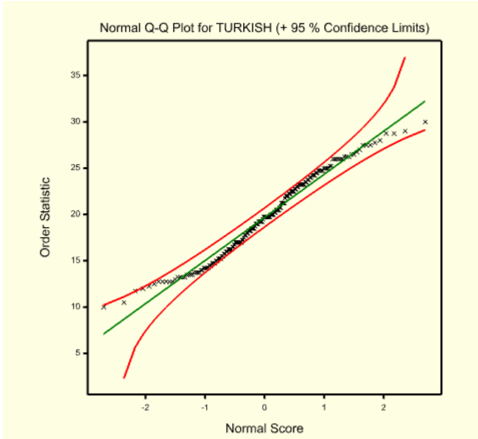


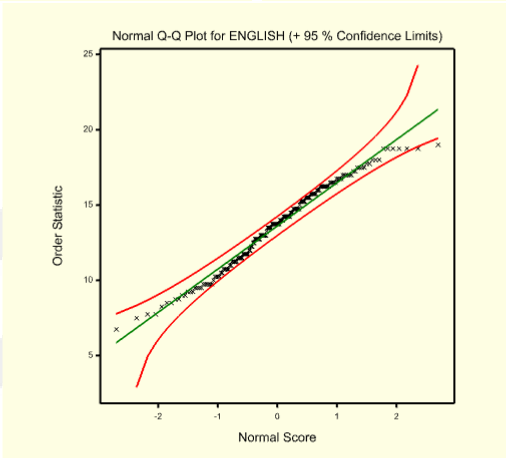
For Q-Q plot we will implement as,











9.2 Excel

Covariance matrices will be obtained by excel calculations following,

	14.0974211	6.76768	3.794035			817.6504	392.5254	220.054			1850.434	767.7937	586.5072
S1	6.76767972	10.95733	4.1755		58*S1	392.5254	635.5254	242.179		W	767.7937	1362.481	601.304
	3.79403492	4.1755	4.528784			220.054	242.179	262.6695			586.5072	601.304	804.857
	6.43589744	2.850113	1.607843			328.2308	145.3558	82			11.01449	4.570201	3.491114
S2	2.85011312	8.222662	2.802696		51*S2	145.3558	419.3558	142.9375		S	4.570201	8.110007	3.579191
	1.60784314	2.802696	3.632047			82	142.9375	185.2344			3.491114	3.579191	4.790815
											D	0.025917	
	11.9415784	3.896822	4.821239			704.5531	229.9125	284.4531			M	25.86377	
S3	3.89682203	5.213559	3.664195		59*S3	229.9125	307.6	216.1875			E	25.19347	
	4.82123941	3.664195	6.050053			284.4531	216.1875	356.9531					
				Averages									
			Math	Turkish	English				|S1|	303.0508		ln|S1|	5.7139
		A	32.10593	20.35593	14.13136	59			|S2|	116.5804		ln|S2|	4.758581
		B	43.23077	24.23077	15.8125	52			|S3|	140.9585		ln|S3|	4.948465
		C	23.2375	15.05	11.1875	60			|S|	202.1541		ln|S|	5.30903
			32.85807	19.8789	13.71045								