Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.365-373

Applied multivariate analysis of variance in experiment of randomized design

Nazik J. Sadik, Iqbal M. Alwan
University of Baghdad, Iraq

ABSTRACT

The general aim of an experimental design in this paper was to estimate the different treatments effects on the responses by statistical methods. The estimates must be averting biases and the random errors minimized as much as possible. We used multivariate analysis of variance (MANOVA) to analyze design of experiments for several responses. In this paper, we provided three fertilizers (mineral, humic, micro-elements) applied on Yellow Maize experiment. This experiment was conducted by completely randomized design (CRD). We tested four responses (Chlorophyll in paper, total ton / ha, paper area / cm^2 and plant height / cm) together to find significant test between them. The partial correlations are between Chlorophyll in paper and total ton of 0.77727. The difference between first fertilizers (mineral) and 3rd fertilizers (micro-elements) are significantly different for the total ton.

Keywords: MANOVA, completely randomized design (CRD), fertilizers

Corresponding Author:
Nazik J. Sadik,
University of Baghdad / Iraq
E-mail: dr.nazik@coadec.uobaghdad.edu.iq

1. Introduction

Multivariate analysis of variance (MANOVA) stands for basically the ANOVA with some dependent variables. ANOVA investigations for a change in means among dual or more collections, whereas MANOVA investigations for the difference in dual or more vectors of means. Several researchers in the literature had been discussed about MANOVA, like Anderson [1], Morrison [2], Timm [3], Rancher and Christensen [4]. The commonly multivariate statistics: Pillai's trace, Wilks' lambda, Lawley–Hoteling trace, which was also discussed by Johnstone and Wichern [5]. Asymptotically, Wilk's lambda, Pillai's trace, in addition to Lawley Hotelling trace have been identical. Nevertheless, their performance under several encroachments of a null theory and with minor samples has been dissimilar. No one of the three multivariate criteria seem to be the most influential in contradiction of all alternative hypotheses.

2. The aim of MANOVA

If we want to know the significant difference between means; as we compare two groups, ANOVA produces the same results as the T-test for independent samples. If the number of groups exceeds two, we use directly ANOVA. Nevertheless, if we have more response, or in other words, more than one dependent variable, so in this case we use MANOVA.

MANOVA can be used in the following cases and aims:

a) When there are many dependent variables (responses).
b) Samples was draw form same experiments.

c) Setting hypotheses to find the effect of independent variables on some response, dependent variables, in the experiments [6].

3. Assumptions of MANOVA

The implementation is to verify the hypotheses developed in order to get a particular decision. To achieve this, we must realize normality of data in which the responses and dependent variables must be typically distributed within groups in the experiments. There are linear relations between all pairs of dependent variables, with equal variance across the groups.

4. One-Way MANOVA

Let k be independent random samples of size n and the mathematical model for responses is:

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\ &= \mu_i \,| + \varepsilon_{ij} \qquad i = 1,2,\ldots,k; \quad j = 1,2,\ldots,n; \quad \ldots..(1) \end{aligned}$$

Rewriting model (1) in matrix form of the p variables as:

$$\begin{pmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijp} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{ip} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \vdots \\ \varepsilon_{ijp} \end{pmatrix}$$

$$= \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \\ \vdots \\ \varepsilon_{ijp} \end{pmatrix}$$

Consequently, a model for r_{th} variable (r = 1, 2,..., p) in every vector $\mathbf{y}_{ij}$ is:

$$\begin{aligned} y_{ijr} &= \mu_r + \alpha_{ir} + \varepsilon_{ijr} \\ &= \mu_{ir} + \varepsilon_{ijr} \end{aligned}$$

And the hypothesis is:

H_0: $\mu_{1r} = \mu_{2r} = \ldots\ldots.. = \mu_{kr}$ $r = 1,2,3,\ldots..,p$, in the case of dual means vary for only single variable, for instance, $\mu_{24} \neq \mu_{43}$ then H_O is false, and it is preferred to discard it [3, 7].

Thus, H_0 implies p sets of equalities:

$$\begin{aligned} &\mu_{11} = \mu_{21} = \ldots\ldots = \mu_{k1} \\ &\mu_{12} = \mu_{22} = \ldots\ldots = \mu_{k2} \\ &. \quad . \quad . \quad . \quad \ldots \quad . \\ &. \quad . \quad . \quad . \quad \ldots. \quad . \\ &\mu_{1p} = \mu_{2p} = \ldots\ldots = \mu_{kp} \end{aligned}$$

In the multivariate ANOVA, “between” and “within” matrices for **H** along with **E** are existing , expressed as:

$$\begin{aligned} H &= n\sum_{i=1}^{n} \left(\bar{y}_{i.} - \bar{y}_{..}\right)\left(\bar{y}_{i.} - \bar{y}_{..}\right)' \\ &= \sum_{i=1}^{k} \frac{1}{n} y_{i.} y_i' - \frac{1}{kn} y_{..} y_{..}' \qquad \ldots.(2) \end{aligned}$$

$$H = \begin{pmatrix} SSH_{11} & SPH_{12} & \dots & SPH_{1p} \\ SPH_{21} & SSH_{22} & \dots & SPH_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ SPH_{p1} & SPH_{p2} & \dots & SSH_{pp} \end{pmatrix} \quad ....(3)$$

Where, for instance:

$$SSH_{22} = n\sum_{i=1}^{k}(\bar{y}_{i.2} - \bar{y}_{..2})^2$$

$$= \sum_i \frac{y_{i.2}^2}{n} - \frac{y_{..2}^2}{kn},$$

$$SPH_{12} = n\sum_{i=1}^{k}(\bar{y}_{i.1} - \bar{y}_{..1})(\bar{y}_{i.2} - \bar{y}_{..2})$$

$$= \sum_i \frac{y_{i.1}y_{i.2}}{n} - \frac{y_{..1}y_{..2}}{kn}.$$

In above formulas, the subscript 1 or 2 specifies the 1st or 2nd variable.

$$\bar{y}_{i.} = \begin{pmatrix} \bar{y}_{i.1} \\ \bar{y}_{i.2} \\ \vdots \\ \bar{y}_{i.p} \end{pmatrix}$$

The matrix **E** is feasibly defined based on an arrangement equivalent to (3):

$$E = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})(y_{i.j} - \bar{y}_{i..})'$$

$$= \sum_{ij} y_{ij}y_{ij}' - \sum_i \frac{1}{n} y_{i.}y_{i..}' \quad ....(4)$$

$$E = \begin{pmatrix} SSE_{11} & SPE_{12} & \dots & SPE_{1p} \\ SPE_{21} & SSE_{22} & \dots & SPE_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ SPE_{p1} & SPE_{p2} & \dots & SSE_{pp} \end{pmatrix} \quad ....(5)$$

Where, for instance:

$$SSE_{22} = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij2} - \bar{y}_{i.2})^2$$

$$= \sum_{ij} y_{ij2}^2 - \sum_i \frac{y_{i.2}^2}{n}$$

$$SPE_{12} = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij1} - \bar{y}_{i.1})(y_{ij2} - \bar{y}_{i.2})$$

$$= \sum_{ij} y_{ij1}y_{ij2} - \sum_i \frac{y_{i.1}y_{i.2}}{n}$$

5. Test statistics

We will review the most important test statistics based on [8 , 9].

5.1. Wilks

The test Statistics is specified by:

$$\Lambda = \frac{|E|}{|E+H|} \quad ....(6)$$

H_0 is rejected in the case of $\Lambda \le \Lambda_{\alpha,p,V_H,V_E}$. Where the rejection is typically for small magnitudes. Exact critical magnitudes have been based on $\Lambda_{\alpha,p,V_H,V_E}$. The parameters in Wilks Λ have been:

p = variables (dimension) number,

ν_H = freedom degrees for hypothesis,

ν_E = freedom degrees for error.

$$df_1 = PVH, \qquad df_2 = wt - \frac{1}{2}(PV_H - 2),$$

$$w = V_E + V_H - \frac{1}{2}(P + V_H + 1), \qquad \text{t}=\sqrt{\frac{\text{P}^2 V_H^2 - 4}{P^2 V_H^2 - 5}}$$

5.2. Pillai

The test statistics are based on the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_s$ defined as:

$$V^{(S)} = tr\left[(E+H)^{-1} H\right] \Rightarrow \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} \quad ....(7)$$

Where, s = min(p,ν_H)

H_0 has been rejected for $V^{(S)} \ge V_\alpha^{(S)}$. The higher percentage points, $V_\alpha^{(s)}$($V_{\alpha,s,m,N}$) is for the following approximation using F-distribution for determining significance levels:

$$F_p = \frac{(2N+s+1)V^{(S)}}{(2m+s+1)(s-V^{(S)})}, \quad ....(8)$$

The s, m, and N parameters can be clarified by:

$$s = \min(V_H, P), \qquad \text{m} = \frac{1}{2}(|V_H - P| - 1), \qquad \text{N} = \frac{1}{2}(V_E - P - 1)$$

And F_p is roughly distributed as $F_{s(2m+s+1),s(2N+s+1)}$.

5.3. Lawley–Hotelling

The test statistics can be clarified by:

$$U^{(S)} = tr(E^{-1}H) \Rightarrow \sum_{i=1}^{S} \lambda_i \quad .....(9)$$

It has been as well-known as *Hotelling generalized* T^2*-statistics*, a rough *F*-statistic can be used as:

$$F_L = \frac{U^{(S)}}{c}, \quad ......(10)$$

Where s = min(p,ν_H)

And F_L is roughly distributed as $F_{a,b}$, in which:

$$a = PV_H, \qquad \text{b} = 4 + \frac{\text{a}+2}{\text{B-1}}, \qquad \text{c} = \frac{\text{a(b-2)}}{b(V_E - P - 1)}, \qquad \text{B} = \frac{(\text{V}_\text{E} + V_H - P - 1)(V_E - 1)}{(V_E - P - 3)(V_E - P)}.$$

6. Numerical example

In this part, we study three fertilizers (mineral, humic, micro-elements) applied on Yellow Maize experiment and the data was obtained from the College of Agriculture / University of Baghdad. Yellow Maize is one of the paramount cereal crops with different uses as food, fodder and industrial applications. As a food crop, it is a major source of nutrition for people in all over the world. It is also the main energy source used in livestock diets in most countries due to its high energy value, the presence of pigments and major fatty acids. Yellow varieties are distinguished as fodder for poultry because they are a rich source of carotene and cantophobia to give the yellow coloration to the coloring of egg yolks, fats and skin. Maize also has the highest amount of energy and has a high TDN ranging from 85 to 90%. By virtue of these advantages, maize is known as food grains. This experiment was conducted by completely randomized design (CRD). We test four responses (Chlorophyll in paper, total ton / ha, paper area / cm^2 and plant height / cm) together to find significant test between them. We used SPSS to get the results which to determine the best fertilizers.

Table 1. MANOVA for fertilizers and test statistics

Principle	Test Statistics	F	DF		P
			Num.	Denom.	
Wilks	0.61537	2.679	8	78	0.012
Lawley–Hotelling	0.57033	2.709	8	76	0.011
Pillais	0.41829	2.645	8	80	0.013

Table 1 represents p--values for Wilk's, Lawley--Hotelling, in addition to Pillai's test statistics for judging their significance evidence of model effects. These magnitudes are (0.012, 0.011, 0.013) for the fertilizers model. There has been significant evidence for fertilizers leading effects under levels higher than 0.013.

Table 2. SSCP matrix adopted for fertilizers

	Chlorophyll in paper	Total ton/ ha	Paper area in cm^2	Plant height in cm
Chlorophyll in paper	359.77	49.68	2251.9	685.5
Total ton/ ha	49.68	11.5	515.2	124.9
Paper area cm^2	2251.94	515.16	23112.2	5626.1
Plant height cm	685.51	124.9	5626.1	1503.9

We can use the sums of squares and cross products –SSCP- matrices for determining a partition of variability in similar way as univariate sums of squares. The matrix labeled H represents sums of squares and cross-products for Fertilizers, which represented in Table 2. The diagonal elements of H matrix, 359.77, 11.48, 23112.2, 1503.9, stand for univariate ANOVA sums of squares for a model fertilizers as the response variables are Chlorophyll in paper, total ton / ha, paper area cm^2 and plant height cm, as in Table 6,8,7,10. The off--diagonal constituents of H matrix stand for the cross products.

Table 3. SSCP matrix adopted for error

	Chlorophyll in paper	Total ton/ ha	Paper area in cm^2	Plant height in cm
Chlorophyll in paper	4280.9	340.68	17191	1869.4
Total ton/ ha	340.7	45	1387	175.8
Paper area cm^2	17191	1387.2	284416	27710
Plant height cm	1869.4	175.79	27710	7354.7

The matrix labeled E represents sums of squares and cross--products for Error, which represented in Table 3. The diagonal elements of E matrix, 4280.9, 44.88 , 284416 , 7354.7 , stand for the univariate ANOVA error sums of squares as the response variables are Chlorophyll in paper , total ton, paper area and plant height as in Tables 6,8,7, 10. The off-diagonal constituents of this matrix stand for the cross products.

Table 4. Partial correlations for the error SSCP matrix

	Chlorophyll in paper	Total ton/ ha	Paper area cm^2	Plant height cm
Chlorophyll in paper	1	0.77727	0.49267	0.33316

Total ton/ ha	0.77727	1	0.38829	0.30599
Paper area cm^2	0.49267	0.38829	1	0.60586
Plant height cm	0.33316	0.30599	0.60586	1

To determine the connection between the responses, we use matrix of partial correlations for the Error SSCP Matrix, which represented in Table 4. These are correlations between the residuals. The partial correlations are between Chlorophyll in paper and total ton / ha is 0.77727 and between plant height cm and paper area cm^2 of 0.60584 are large. The partial correlations between Chlorophyll in paper and paper area in cm^2, plant height in cm are 0.49267, 0.33316 respectively and between total ton / ha and paper area in cm^2, plant height in cm are 0.38829, 0.30599 respectively, that are not large. Since the correlation structure is not all weak, so we can use MANOVA procedure.

Table 5. Eigen analysis for fertilizers

Eigenvalue	0.4483	0.122	0	0
Proportion	0 .7861	0.2139	0	0
Cumulative	0 .7861	1.0	1.0	1.0
Eigenvector of Chlorophyll in paper	0.0135	0.0195	0.0099	0.0003
Eigenvector of Total ton/ ha	-0.1883	-0.1131	0.0256	0.0885
Eigenvector of Paper area cm^2	-0.0002	-0.0013	0.0009	-0.002
Eigenvector of Plant height cm	-0.0054	0.0094	-0.01	0

Eigen analysis represented in Table 5 was computed for the matrix of $E^{-1}H$ and we use it to calculate the four MANOVA tests.

Table 6. Investigation of variance for chlorophyll in paper, using employed S.S for tests

Source	df	SeqSS	AdjSS	AdjMS	F	P
Fertilizers	2	359.8	359.8	179.9	1.76	0.184
Error	42	4280.9	4280.9	101.9		
Total	44	4640.7				

The ANOVA chlorophyll in Table 6 shows the value of F statistics, which equals 1.76487, and the P-value has been bigger than 0.05. Hence, this signposts non-significance differences between the average chlorophyll of the three fertilizers.

Table 7. Investigation of variance for paper area cm^2, via adjusted SS for tests

Source	df	SeqSS	AdjSS	AdjMS	F	P
Fertilizers	2	23112	23112	11556	1.71	0.194
Error	42	284416	284416	6772		
Total	44	307528				

Table 7 shows the F statistics for Paper area in cm^2, which is equals to 1.7065 and the P-value greater than 0.05 and this explains a non-significance difference between the mean paper areas of fertilizers.

Table 8. Investigation of variance for total ton / ha, via adjusted SS for tests

Source	df	SeqSS	AdjSS	AdjMS	F	P
Fertilizers	2	11.485	11.485	5.742	5.37	0.008
Error	42	44.875	44.875	1.068		
Total	44	56.36				

The F statistics for Total ton / ha for ANOVA as in Table 8 equal to 5.37443. As the P-value has been smaller than 0.05, there has been a statistically significant difference amid the mean total ton of fertilizers. To determine which of the fertilizers means have been significantly different from each other's, we compute the multiple range tests.

Table 9. Multiple range tests for total ton by fertilizers

Contrast	Sig.	Difference	+/- Limits
1 – 2		-0.4798	0.761707
1 – 3	*	-1.22773	0.761707
2 – 3		-0.747933	0.761707

* symbolizes a statistically significant difference.

In Table 9, we see that the difference between first fertilizers (mineral) and third fertilizers (micro-elements) are significantly different for the total ton.

Table 10. Investigation of variance for plant height cm, by means of adopted SS for tests

Source	df	SeqSS	AdjSS	AdjMS	F	P
Fertilizers	2	1503.9	1503.9	752	4.29	0.02
Error	42	7354.7	7354.7	175.1		
Total	44	8858.6				

The ANOVA results in Table 10 show F statistics for plant height in cm equals to 4.29412, and the P-value is less than 0.05. As a result, there is a significance of difference between the mean plant heights from one level of fertilizers to another.

Table 11. Various range tests for plant height by fertilizers

Contrast	Sig.	Difference	+/- Limits
1 – 2		-1.158	9.75142
1 – 3	*	-12.8013	9.75142
2 – 3	*	-11.6433	9.75142

* symbolizes a statistically significant difference.

Table 11 employs multiple comparison procedures for determining means that are significantly different from others for plant height and we see that the difference between first fertilizers (mineral) and third fertilizers (micro-elements) are significantly different and the difference between second fertilizers (humic) with third fertilizers (micro-elements) are significantly different too.

Table 12. Multiple comparisons

<table>
<tr><th>Dependent Variable</th><th>Fertilizer_I</th><th>Fertilizer_J</th><th>Mean Difference (I-J)</th><th>Std. Error</th><th>Sig.</th></tr>
<tr><td rowspan="5">Chlorophyll</td><td rowspan="2">1</td><td>2</td><td>1.9753</td><td>3.68649</td><td>.595</td></tr>
<tr><td>3</td><td>-4.7613</td><td>3.68649</td><td>.204</td></tr>
<tr><td rowspan="2">2</td><td>1</td><td>-1.9753</td><td>3.68649</td><td>.595</td></tr>
<tr><td>3</td><td>-6.7367</td><td>3.68649</td><td>.075</td></tr>
<tr><td>3</td><td>1</td><td>4.7613</td><td>3.68649</td><td>.204</td></tr>
</table>

Dependent Variable	Fertilizer_I	Fertilizer_J	Mean Difference (I-J)	Std. Error	Sig.
		2	6.7367	3.68649	.075
Total_ton	1	2	-.4798	.37744	.211
		3	-1.2277*	.37744	.002
	2	1	.4798	.37744	.211
		3	-.7479	.37744	.054
	3	1	1.2277*	.37744	.002
		2	.7479	.37744	.054
Paper_area	1	2	-20.8687	30.04843	.491
		3	-54.9831	30.04843	.074
	2	1	20.8687	30.04843	.491
		3	-34.1145	30.04843	.263
	3	1	54.9831	30.04843	.074
		2	34.1145	30.04843	.263
Plant_height	1	2	-1.1580	4.83202	.812
		3	-12.8013*	4.83202	.011
	2	1	1.1580	4.83202	.812
		3	-11.6433*	4.83202	.020
	3	1	12.8013*	4.83202	.011
		2	11.6433*	4.83202	.020

Table 12 is based on apparent means. The error term is Mean Square (Error) = 175.113 and the star (*) in above table has been significant at the .05 level.

7. Conclusion

- The p-values of Wilks', Lawley- Hotelling, and Pillai's test statistic shows significant evidence for fertilizers main effects at levels greater than 0.013.
- The partial correlations between Chlorophyll in paper and total ton and between plant height and paper area are large. The partial correlations between Chlorophyll in paper and paper area, plant height and between total ton and paper area, plant height are not large. These agree with assumptions of MANOVA
- Statistically, there has been no significance of difference between the mean Chlorophyll from one level of fertilizers to another under 95.0% confidence level for both paper area because the P-value of the F-test has been bigger than or equivalent to 0.05. But plant height and Total ton are statistically significant difference because the P-value of the F-test has been smaller than 0.05.
- The difference between first fertilizers (mineral)and third fertilizers (micro-elements) are significantly different for total ton and the difference between first fertilizers (mineral) with there'd fertilizers (micro-elements) and the difference between second fertilizers (humic) with third fertilizers (micro-elements) are significantly different for plant height.

References

[1] T. W. Anderson, An introduction to multivariate statistical analysis (3rd ed.). New York: Wiley, 2003.

[2] D. F. Morrison, Multivariate Statistical Methods, (4th ed), Belmont, CA: Duxbury, 2005.

[3] N. H. Timm, Applied Multivariate Analysis, Springer Science & Business Media, 2007.

[4] A. C. Rencher and W. F. Christensen, Methods of Multivariate Analysis, (3rd ed). Hoboken, NJ: Wiley, 2012.

[5] R. A. Johnson, and D. W Wichern., Applied multivariate statistical analysis (6th ed.), England and Associated Companies throughout the world, 2014.

[6] G. Carey, Multivariate Analysis of Variance (MANOVA): I. Theory, 2011.

[7] F. G. Khamis, Ghaleb A. El-Refae, "Applying Multivariate and Univariate Analysis of Variance on Socioeconomic", Health, and Security Variables in Jordan, Statistics, Optimization and Information Computing, Vol. 8, June 2020, pp 386–402.

[8] A. A. Ameen, Aseel A. Jaaze," Robustness and Comparison of Wilks' Test Statistic for Two-Way MANOVA", Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.12(1) 2020, pp stat 1–23

[9] P. Sweta, C. D. Bhavsar, " Analysis of pharmacokinetic data by wilk's lambda (An important tool of manova)", International Journal of Pharmaceutical Science Invention, www.ijpsi.org Volume 2 Issue 1, January 2013, PP.36-44.