Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

MANCOVA Model For Heteroscedasticity And Unbalance Cell Size: Framework

Noor Abdul kareem Kadhim[a], Huda Zaki Naji[b]

[a] ***Department of Mathematics, College of Sciences, University of Basrah, Al- Basrah, Iraq***. *Email:* ***nawri.muhsin.sci@uobasrah.edu.iq***

[a] ***Department of Mathematics, College of Sciences, University of Basrah, Al- Basrah, Iraq***. *Email:* ***huda.naji@uobasrah.edu.iq***

ARTICLE INFO

Article history:
Received: 20 /02/2022
Rrevised form: 06 /03/2022
Accepted : 22 /03/2022
Available online: 23 /03/2022

Keywords:

MANCOVA

Heteroscedastic model

Modified MANOVA test

ABSTRACT

Multivariate analysis of covariance models are widely used in human societies, principally in health, ecology, environmental science and psychological perspective. As such, the researchers face the task of seeking for ways to improve the performance of these models to respond to the requirements which are treating. This paper proposes an extension of two way MANOVA to two way MANCOVA model under effect one covariate factor for heteroscedastic and unbalanced cell sizes. A type III sums squares of an influence is computed that adjusted for all other influences for suggested two way MANCOVA model, no matter in which order they are included. As classical tests could dangerously biased when homogeneity violates the assumption of cell covariance matrices, the suggested model implemented the modified tests obtained by Zhang & Xiao (2012) to test the hypothesis of model

MSC. 41A25; 41A35; 41A36.

https://doi.org/10.29304/jqcm.2022.14.1.885

1. INTRODUCTION

Generally, the analysis of covariance (ANCOVA) utilizes when considering the variance in a response variable that is explained by one or more covariates as well as the differences in the response variable between groups. (Huitema, (2011)). Many scientific, health and environmental disciplines rely on ANCOVA models as a statistical tool to obtain accurate results (see the references: Cooper, et al., (2018); Mochizuki, Amagai & Tani, (2018); Focht, et al. (2018); Kononova, McAlister & Oh, (2019)). Addition to default assumptions such as; independence, equal variances, moderation, ANCOVA models include some additional assumptions. Additional assumptions are that (i) there is a linear relationship between the response variable and a covariate, (ii) intergroup parallel regression, (iii) at each value of covariate the variance at the response variables for each group is equal, (iv) evenness of covariate between the groups, and (v) an independent variable is considered as fixed- influence that calculated errors (Cangür, Sungur & Ankaral, (2018)). Violation of one of these conditions threatens the validation of the experimental output of the ANCOVA models (Rheinheimer & Penfield, (2001)). Although, ANCOVA models are utilized to check the differences for intergroup response variable, considering the heterogeneity in response variables that explains one or a group of common variables. ANCOVA model exemplifies a collection from an analysis of variance and the linear regression. Whereas, these models are generated incorporating regression conditions into ANOVA models. ANCOVA models consider both the

*Corresponding author: Noor Abdul kareem Kadhim

Email addresses: ***nawri.muhsin.sci@uobasrah.edu.iq***.

Communicated by : Dr.Rana Jumaa Surayh aljanabi.

intergroup covariance and the regression variance as methodological components. Equation (1) below represents a one-way model for covariance:

$$Y_{ij} = \mu + \tau_j + +\beta\,(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij} \ , \quad i = 1,\ldots, k \quad j = 1,\ldots,n_i \tag{1}$$

Where, Yij represents the response of the i^{th} individual variable in a j^{th} level of factor, where μ shows overall mean, τ_j represents an effect of a j^{th} factor, β represents a coefficient vector of the linear regression for the variable Y over the variable X, Xij shows a value of the covariate for i^{th} in a factor level j^{th} , $\bar{X}_{..}$ represents a grand covariate mean addition to εij which represents the error term for i^{th} unit in j^{th} level of factor (Naji, (2008); Al-mouel, (2004); Naji, (2014)).

When all conditions are met, the analysis of covariance is a very useful statistical method in many scientific fields. All ANCOVA›s conditions are represented by the below points (Huitema, (2011)):

1. The conditional distribution of response variable Y is normally distributed with zero expected and σ^2 variance, when the variable X is given.
2. Within the group the slopes of the gradient must be homogeneous ($\beta^{\text{group 1}} = \beta^{\text{group 2}} = \cdots = \beta^{\text{group I}}$).
3. The conditional variances are homogeneous$\sigma^2_{Y_1|X} = \sigma^2_{Y_2|X} = \cdots = \sigma^2_{Y_I|X}$.
4. The errors in the model follow normal distributions with zero rate and variance σ^2, ($\varepsilon ij \sim N(0, \sigma^2)$).
5. Between the response variable and the covariate there is a relationship that must be linear.
6. Independence between the levels of the factor.
7. The common variable must be a constant variable and is calculated without errors.
8. Should be no robust relationship among the covariates if there is more than one covariate is utilized.

On the other hand, a Multivariate Analysis of Variance (MANOVA) expansions for Analysis of Variance (ANOVA) when influences of factors assessed on a linear combination of several response variables. Wilks, (1932) was the first suggested the multivariate popularization of the ANOVA model, for the time being a MANOVA methodologies are well decided and vastly utilized in various fields of science, ranging from biology to psychology (Zhang & Xiao (2012)). A MANOVA models have different features compared to ANOVA models:

- Using MANOVA models, the researcher can test the common hypotheses of all univariate ANOVA models and note the differences between factor levels (Where these models examine whether there are statistical differences between the levels of factors for response variables).
- Reduces the probability of type I error by using a single MANOVA model replace several ANOVA models. (Anderson, (2003); Kim & Timm, (2006)).
- Most ANOVA models do not consider the pattern of covariance between response variables. While the MANOVA model is sensitive to differences in the levels of factors also the difference between response variables. When the response variables are considered together, probably that these variables are associated with some extension and by performing several ANOVA analyze, this association will be lost.

In the classical ANOVA, ANCOVA and MANCOVA models where the covariance matrices are equal, the Wilkes Probability Ratio (WLR), follows the Lawley Hotelling (LHT), Pillai Bartlett trace (PBT) as well as Roy's largest root tests could be utilized (Watanabe, Hyodo & Nakagawa, (2020)). However, these tests might be seriously biased on the equal assumption for the covariance matrices are break. For realistic data, the well-known Box's test M is used to check the homogeneity of the assumption of covariance matrices [4]. Such, this defaults are seriously violated, some new tests need to be developed and the above tests should not be used for a two-way MANOVA. Unbalanced agricultural empirical data was first published in 1930 by scientist Frank Yates (Herr, (1986); Nelder & Lane, (1995)). Subsequently, several explorations that discussed unbalanced data appeared in the field of computer software and how to visualize data (Langsrud, (2003)). Often the articles published focus on the type of square sums used in the ANOVA table as well as the assumptions they make. As the three kinds of sum of squares differ, it often leads to a different interpretation of the statistical significance of different factors (Saber & Naji, (2010)).

Presenting study extended the two-way MANOVA for two-way MANCOVA included one covariate. Additionally, the suggested model presented the following points:

1. The present paper will develop a two-way MANCOVA model with un-equal cell sizes additionally un-equal cell of covariance matrices.
2. The modified tests presented by Zhang & Xiao, (2012) will be used for heteroscedastic two-way MANCOVA.
3. A type III sums squares of an influence is computed that adjusted for all other influences for the suggested two-way MANCOVA model, no matter in which order they are included.

The remainder of this paper is organized as follows. Section 2 presented the methodologies of this study. Section 3 presented the suggested two-way MANCOVA model. Conclusion was presented in section 4.

2. METHODOLOGIES OF THIS STUDY

2 .1. UNBALANCED TWO-WAY MANCOVA MODEL WITH INTERACTION

According to Zhang & Xiao (2012), imbalanced and non-homogeneous two-way MANCOVA model with interactions could be presented in this section. The notation and structure be expressed as follows:

$$\boldsymbol{Y_{ijk} = \mu + \vartheta_i + \psi_j + (\vartheta\psi)_{ij} + (Z_{ijk} - \bar{Z}_{\dots})\eta_1 + e_{ijk}} \tag{2}$$

Table 1 below summarizes of symbols that are used in the imbalanced and non-homogeneous two-way MANCOVA model.

Table 1 – summary of symbols that are used in the imbalanced and non-homogeneous two-way MANCOVA model.

symbols	Description
μ	Over all mean.
i	Experiment levels for factor A (indexed by a).
j	Experiment levels for factor B (indexed by b).
k	The index of experimental units to level (i, j) (indexed by n_{ij}).
ϑ_i	The i^{th} main influence of factor A.
ψ_j	The j^{th} main influence of factor B.
$\vartheta\psi_{ij}$	The/ $(i, j)^{th}$ interaction influence combine factor A and factor B.
Z	The value of covariate for unit *i* within groups (*j*, *k*).
$\bar{Z}$	The mean of covariate over all empirical units.
η	The slope corresponding to covariate Z.
e_{ijk}	Random error.

For the parameterization to be of full rank, we impose the following set of conditions:

$$\sum_{i=1}^{a}\vartheta_i = 0\,;\ \sum_{j=1}^{b}\psi_j = 0\,;\ \sum\nolimits_{i=1}^{a}\sum\nolimits_{j=1}^{b}(\vartheta\psi)_{ij} = 0$$

$$\sum\nolimits_{i=1}^{a}(\vartheta\psi)_{ij} = 0\,;\ j = 1,\dots,b-1 \quad,\quad \sum\nolimits_{j=1}^{b}(\vartheta\psi)_{ij} = 0\,;\ i = 1,\dots,a-1 \tag{3}$$

$$\boldsymbol{\sum_{i=1}^{a} Z_{ijk} = \sum_{j=1}^{b} Z_{ijk} = 0} \quad ; \quad \boldsymbol{e_{ijk} \sim N_P\,(0, \textstyle\sum_e)}$$

3. THE TWO-WAY MANCOVA MODEL

3.1. ESTIMATION IN THE TWO-WAY MANCOVA MODEL

According to the virtual balanced assumptions one may imposing restriction, as in the equation (1), on the over presented model (3) for obtaining decisions for problem of normal equations and single values for the estimator: $\hat{\vartheta}_\iota, \widehat{\psi_j}, \widehat{(\vartheta\psi)}_{\iota J}, \widehat{Z_{\iota jk}}$. These capabilities in the suggested model (3) can be written as follows:

$$\widehat{\vartheta_\iota} = \sum_{i=}^{a} n_{ij}(\overline{Y}^*_{i..} - \overline{Y}^*_{...})(\overline{Y}^*_{i..} - \overline{Y}^*_{...})' \quad (4)$$

$$\widehat{\psi_J} = \sum_{j=1}^{b} n_{ij}(\overline{Y}^*_{.j.} - \overline{Y}^*_{...})(\overline{Y}^*_{.j.} - \overline{Y}^*_{...})' \quad (5)$$

$$\widehat{(\vartheta\psi)}_{\iota J} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} n_{ij}\ (\overline{Y}^*_{ij.} - \overline{Y}^*_{i..} - \overline{Y}^*_{.j.} + \overline{Y}^*_{...})(\overline{Y}^*_{ij.} - \overline{Y}^*_{i..} - \overline{Y}^*_{.j.} + \overline{Y}^*_{...})' \quad (6)$$

$$\widehat{Z_{\iota Jk}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} \frac{1}{\eta^*}\left[(\overline{Y}^*_{ijk} - \overline{Y}^*_{jk})(\overline{Y}^*_{ijk} - \overline{Y}^*_{jk})'\right] \quad (7)$$

As can be seen in equations (4 - 7), expectations are.weighted.by n_{ij} that leads to that produced estimates are different than for the balanced model. For instance, $\overline{Y}_{...}$ no longer equals the overall mean of the sample as when data is balanced.

3.2. HYPOTHESIS TESTING

For an unbalanced two-way MANCOVA model by Types III (When a major factor (i.e. A and B), there is evaluated after all other factors (lines with interaction.) have been respected, are the following null hypotheses:

For no main effect of factor A:

$$H_{01} = \vartheta_1 = \vartheta_2 = \cdots = \vartheta_a = 0$$

For no main effect of factor B:

$$H_{02} = \psi_1 = \psi_2 = \cdots = \psi_b = 0$$

For no effects of the factor A:

$$H_{03}^{(A|B)} = \vartheta_1 + \vartheta_1\psi_1 = \cdots = \vartheta_1 + \vartheta_1\psi_b = \vartheta_2 + \vartheta_1\psi_1 = \cdots = \vartheta_2 + \vartheta_2\psi_b = \cdots = \vartheta_a + \vartheta_a\psi_b = 0$$

For no influences of the factor B:

$$H_{04}^{(B|A)} = \psi_1 + \vartheta_1\psi_1 = \cdots = \psi_1 + \vartheta_a\psi_1 = \psi_2 + \vartheta_1\psi_2 = \cdots = \psi_2 + \vartheta_a\psi_2 = \cdots = \psi_b + \vartheta_a\psi_b = 0$$

For no mutual influences of factors A and B.:

$$H_{05} = \vartheta_1\psi_1 = \vartheta_2\psi_1 = \cdots = \vartheta_a\psi_1 = \vartheta_1\psi_2 = \vartheta_2\psi_2 = \cdots = \vartheta_2\psi_b = \cdots = \vartheta_a\psi_b = 0$$

For no covariate effects:

$$H_{06} = (Z_{i1} - \bar{Z}_{...}) = (Z_{i2} - \bar{Z}_{...}) = \cdots = (Z_{ab} - \bar{Z}_{...}) = 0$$

With cellular heterogeneity matrices, homogeneity assumption is violated, the basically statistical tests cannot be used to test the above hypotheses (Cooper, et al., (2018)). According to Zhang & Xiao, we define a sum square and cross product (SSCP) for the imbalanced two-way MANCOVA as the following:

$$H_A = \frac{1}{a-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}^*_{i..} - \overline{Y}^*_{...})(\overline{Y}^*_{i..} - \overline{Y}^*_{...})' \quad (8)$$

$$H_B = \frac{1}{b-1}\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}^*_{.j.} - \overline{Y}^*_{...})(\overline{Y}^*_{.j.} - \overline{Y}^*_{...})' \quad (9)$$

$$H_{AB} = \frac{1}{(a-1)(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}^*_{ij.} - \overline{Y}^*_{i..} - \overline{Y}^*_{.j.} + \overline{Y}^*_{...})(\overline{Y}^*_{ij.} - \overline{Y}^*_{i..} - \overline{Y}^*_{.j.} + \overline{Y}^*_{...})' \quad (10)$$

$$H_{(A|B)} = \frac{1}{(a-1)b}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}^*_{ij.} - \overline{Y}^*_{.j.}\right)\left(\overline{Y}^*_{ij.} - \overline{Y}^*_{.j.}\right)' \quad (11)$$

$$H_{(B|A)} = \frac{1}{a(b-1)}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}^*_{.j.} - \overline{Y}^*_{...}\right)\left(\overline{Y}^*_{.j.} - \overline{Y}^*_{...}\right)' \quad (12)$$

$$H_Z = \sum_{i=1}^{a}\sum_{j=1}^{b}(\overline{Y}^*_{ijk} - \overline{Y}^*_{jk})(\overline{Y}^*_{ijk} - \overline{Y}^*_{jk})' \quad (13)$$

where H and G will be approximately Wishart distributed with unknown approximate degrees of freedom

(f_H, f_G) following respectively the distributions of H and G:

$$H \sim W\left(f_H, \frac{\Sigma}{f_H}\right),$$

$$G \sim W\left(f_G, \frac{\Sigma}{f_G}\right)$$

Unlike to the standard statistical tests, the modified tests be based on the unknown quantities Σ, f_H and f_G. According to (Zhang & Xiao, (2012)) to determine f_H and f_G for unbalanced two way MANCOVA model based on H and G respectively by replacing

(Σ by I_p , Σ_{ij} *by* $\Sigma^{-1/2}\,\Sigma_{ij}\,\Sigma^{-1/2}$ *and* $\Sigma_{\vartheta\psi}$ *by* $\Sigma^{-1/2}\,\Sigma_{\vartheta\psi}\,\Sigma^{-1/2}$) with $e_{ijk} \sim N_p\left(0, \frac{\Sigma^{-1/2}\,\Sigma_{ij}\,\Sigma^{-1/2}}{n_{ij}}\right)$ then obtain:

$$f_G = \frac{p\,(p+1)}{(ab)^{-2}\sum_{i,j}(n_{ij}-1)^{-1}\,(n_{ij})^{-2}\left\{ tr\left(\left[\Sigma_{ij}\,\Sigma^{-1}\right]^2 + tr^2\left[\Sigma_{ij}\,\Sigma^{-1}\right]\right)\right\}}$$

$$f_H = \frac{p\,(p+1)}{\sum_{i,j}\sum_{\alpha,\beta}\frac{c_{ij,\vartheta\psi}{}^2}{n_{ij}n_{\vartheta\psi}}\left\{ tr\left(\left[\Sigma_{ij}\,\Sigma^{-1}\,\Sigma_{\vartheta\psi}\,\Sigma^{-1}\right]^2 + tr\left[\Sigma_{ij}\,\Sigma^{-1}\right]\right) tr\left[\Sigma_{\vartheta\psi}\,\Sigma^{-1}\right]\right)\right\}}$$

where $c_{ij,\vartheta\psi}{}^2$ represented the weights for the hypothesis tested, while $\vartheta = ,1 \ldots, a$, $\psi = ,1 \ldots, b$ represented indices that used to convenience of calculation.

4. CONCLUSION

Classical statistical tests of MANOVA models become ineffective if one of their basic conditions is violated, and it becomes difficult to deal with them by adding real conditions. To address this serious problem, Zhang 2012 suggested the modified tests. The model differs from previous related work in that it takes into account the heterogeneity, unbalance of data as well as the influence of the covariance factor (MANCOVA). Also, this paper presented the two-way MANCOVA model under type III sums squares of an influence is computed that adjusted for all other influences for suggested two way MANCOVA model, no matter in which order they are included. The future work will be to solve our model by using real data that fits with the proposed model conditions.

References

[1] Huitema, B. The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments, and single-case studies, (2011), (Vol. 608). John Wiley & Sons.

[2] Cooper, D. J., Plewes, K., Grigg, M. J., Rajahram, G. S., Piera, K. A., William, T., ... & Barber, B. E. The effect of regularly dosed paracetamol versus no paracetamol on renal function in Plasmodium knowlesi malaria (PACKNOW): study protocol for a randomized controlled trial. *Trials*, (2018), *19*(1), 1-11.

[3] Mochizuki, T., Amagai, T., & Tani, A. Effects of soil water content and elevated CO2 concentration on the monoterpene emission rate of Cryptomeria japonica. *Science of the Total Environment*, (2018), *634*, 900-908.

[4] Focht, B. C., Lucas, A. R., Grainger, E., Simpson, C., Fairman, C. M., Thomas-Ahner, J. M., ... & Clinton, S. K. Effects of a group-mediated exercise and dietary intervention in the treatment of prostate cancer patients undergoing androgen deprivation therapy: results from the IDEA-P trial. *Annals of Behavioral Medicine*, (2018), *52*(5), 412-428.

[5] Kononova, A., McAlister, A., & Oh, H. J. Screen overload: Pleasant multitasking with screen devices leads to the choice of healthful over less healthful snacks when compared with unpleasant multitasking. *Computers in human behavior*, (2019), *80*, 1-11.

[6] CANGÜR, Ş., SUNGUR, M. A., & ANKARALI, H. The Methods Used in Nonparametric Covariance Analysis. *Duzce Medical Journal*, (2018), 20(1), 1-6.

[7] Rheinheimer, D. C., & Penfield, D. A. The effects of type I error rate and power of the ANCOVA F test and selected alternatives under non-normality and variance heterogeneity. *The Journal of Experimental Education*, (2001), *69*(4), 373-391.

[8] Naji, H. Z. *Covariates in the Analysis of Multivariate Repeated Measurements*, (2008), (M. Sc. Thesis, University of Basrah).

[9] AL-MOUEL, A. H. S. Nested Repeated measures model sufficiency and Estimation. *Chinese Journal of Applied Probability and Statistics,* (2004).

[10] Naji, H. Z. Two Random Covariates in Repeated Measurement Model. *Basrah journal of science (A)*, (2014), *32*(1), 130-140.

[11] Wilks, S. S. Certain generalizations in the analysis of variance. *Biometrika*, (1932), 471-494.

[12] Zhang, J. T., & Xiao, S. A note on the modified two-way MANOVA tests. *Statistics & Probability Letters*, (2012), *82*(3), 519-527.

[13] Kim, K., & Timm, N. *Univariate and multivariate general linear models: theory and applications with SAS*. CRC Press. (2006).

[14] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. 3rd edition, Wiley, New York, (2003).

[15] Watanabe, H., Hyodo, M., & Nakagawa, S. Two-way MANOVA with unequal cell sizes and unequal cell covariance matrices in high-dimensional settings. *Journal of Multivariate Analysis*, (2020), *179*, 104625.

[16] Herr, D. G. On the history of ANOVA in unbalanced, factorial designs: The first 30 years. *The American Statistician*, (1986), *40*(4), 265-270.

[17] Nelder, J. A., & Lane, P. W. The computer analysis of factorial experiments: In memoriam—Frank Yates. *The American Statistician*, (1995), *49*(4), 382-385.

[18] Langsrud. ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, (2003), *13*(2), 163-167.

[19] Saber, A. H., & zaki Naji, H. The Sphericity test for Multivariate Repeated Measurement Model has two Covariates. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, (2010), *2*(1).

[20] Naji, H. Z., Dawood, R., & Faik, R. R. Analysis of variance and covariance Repeated Measurement Model. *Journal of Basrah Researches*, (2011), *37*(1), 32-38.