Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 6 (2): 453-457, 2006
ISSN 1812-5654

Analysis of Repeated Measures by Using Multivariate Method

Suat Sahinler and Ozkan Gorgulu
Department of Animal Sciences, Faculty of Agriculture, Mustafa Kemal University, 31034, Hatay, Turkey

Abstract: The present study was carried out to explain the analysis of repeated measures by using the multivariate test statistic which is suitable when the assumptions validity for univariate methods for repeated measures is dubious. Special attention was given to the test of assumptions, illustration of the multivariate analysis of repeated measures and comparative discussion with univariate methods from the point of view that the effects on the statistical inferences of lacking or making wrong statistical procedures. Results showed that how much important the necessity of validation of the assumptions for univariate statistical test for repeated measures and how much their influences on test results and interpretation are great of importance.

Key words: Repeated measures, assumptions, multivariate analysis

INTRODUCTION

Repeated measurements are observations of the same characteristic, which are made several periods on the same experimental unit. In fact, repeated measurements are very frequent in all scientific fields where statistical models are used. The different methods were used for the analysis of repeated measures. Some of them are univariate variance analysis technique of repeated measures, non-parametric and multivariate analysis techniques. In univariate repeated for the measures of experimental designs, the following assumptions are very important and should be validated; I) the measurement errors are independent and identically normally distributed with mean 0 and the same variance. The variance of the difference between the estimated means for any two different factor levels will be the same. This property is called sphericity. A slightly more restrictive assumption is that the covariance between observations would be same for two different factor levels. This property is called compound symmetry. Compound symmetry is a special case of more general property sphericity. If compound symmetry exists, then spericity also exists, but it is possible for spericity to exist when compound symmetry does not[1]. ii) The subjects are considered to be a random sample from the subject population of interest, so that the subject effect can be random.

Because the measurements from the same subject in different periods will be positively non-constant correlated; the assumptions for univariate analysis technique will not validate. In many biological researches, due to be large tendency of variance based on the mean largeness, the variance-covariance matrix from t treatment groups is usually not homogeneous. In this case, for determining of the homogeneity of the variance assumption in repeated measurements experimental designs, the structure of the variance-covariance matrix have to be examined[2]. If the assumption is not validated, using the univariate analysis techniques may cause faulty results. For this reason, using multivariate analysis techniques in analysis of repeated measurements are more available when the assumptions validity is dubious. These include an ε adjustment procedure based on Geisser and Greenhouse[3] and a multivariate analysis using Hotelling's T^2 statistic[4].

In this study, illustration of main and interaction effects by using a multivariate test for repeated measures with two factors one of them contains repeated measures was explained and given the results and inferences for comparing treatments in a way that permit sensitive tests, often when the assumptions are violated. Special attention was given to the test of assumptions, illustration of the multivariate analysis of repeated measures and comparative discussion with univariate methods from the point of view that the effects on the statistical inferences of lacking or making wrong statistical procedures.

MATERIALS AND METHODS

Material: The data concerning the experiment aimed to determine variation on diastase activity and Hydroxymethylfurfural (HMF) in different honey origins during storage period was used. Three different honey origins which are orange, high plateau and cotton were used in this study. The data regarding diastase activity and Hydroxymethylfurfural (HMF) contents were measured at the four different periods.

Corresponding Author: Dr. Suat Sahinler, Department of Animal Sciences, Faculty of Agriculture, Mustafa Kemal University, 31034, Hatay, Turkey

The assumptions for univariate tests of repeated measures were tested and then both the univariate and multivariate tests were applied to data for comparing the results. The General Linear Model (GLM) procedures of SPSS package for repeated measures were used[5]. Duncan Multiple Range Test was used to separate the means among the Honey types. Bonferroni Multiple Range Test was used to separate the means among the periods.

The mathematical model of the experiment in the material was

$$y_{ijk} = \mu + \alpha_i + \gamma_{(i)k} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{(i)jk}\,, \quad i = 1, 2, 3. ..., n,\ j = 1, 2, 3,..., p,\ k = 1, 2, 3,..., r\ (n.q) \quad (1)$$

and used to test Between Subject Factor (BSF) and Within Subject Factor (WSF) and BSF X WSF interaction effect. BSF is a non-repeated or grouping factor, such as race or experimental group, for which subjects will appear in only one level (Honey origins). WSF is repeated factors for which subjects will participate in each level e.g., subjects participate in both experimental conditions, albeit at different times (Periods). Where μ is overall mean, α_i is the effect of ith level of BSF (Honey origins), β_j is the effect of jth level of WSF (Periods), $\alpha\beta_{ij}$ is interaction effect of ith level of BSF and jth level of WSF, $\gamma_{(i)k}$ is the effect of kth experimental unit in ith level of the BSF (Error 1) and ε_{ij} is error term (Error 2).

Method: All data for all characteristics were analyzed using the following way.

Testing of the assumptions: In univariate repeated measures experimental designs, the following assumptions are very important and should be validated for acceptability of the results of these statistical analyses.

The hypothesis which means the variance-covariance structure is homogenous in each level of BSF is tested. There are two cases for testing the Ho hypothesis

$$H_0 = \Sigma_1 = \Sigma_2 = ... = \Sigma_n \quad (2)$$

Deal with variance-covariance structure. These are;

$$\text{If n and } p \leq 5 \text{ then } q = g_3\,(1 - g_1) \quad (3)$$

Statistic is used to test Ho hypothesis. Where;

$$g_1 = [(n+1)/(r-n)]\lfloor(2p^2 + 3p - 1)/6(p+1)\rfloor \quad (4)$$

$$g_3 = \left[(r-n)\log_e|S_b| - \sum_{i=1}^{n}(q-1)\log_e|S_i|\right] \quad (5)$$

Where, n is the number of BSF level, p is the number of WSF level, r is the number of total subject, q is the number of subjects in each level of BSF, S_i is variance-covariance matrix of order n x n concerning the treatment and periods and $S_b^2 = \sum_{i=1}^{n} S_i / n.q$ statistic from Eq. 3 is compared with X^2_{α,v_1} value, where, $v_1 = [p + p(p-1)/2](n-1)$. If $q > \chi^2_{\alpha,v_1}$ then H_0 is rejected.

$$\text{If } p>5 \text{ then } F = g_3\,[(1 - g_1) - (v_1/v_2)]/v_1 \quad (6)$$

statistic is used to test H_0 hypothesis, where g_1, g_3 and v1 are the same in Eq. 3, $v_2 = \lfloor(v_1 + 2)/(g_2 - g_1^2)\rfloor$, where, $g_2 = \lfloor(n^2 + n + 1)(p^2 + p - 2)\rfloor/\lfloor 6(r-n)^2\rfloor$. F statistic from Eq. 6 is compared with F_{α,v_1,v_2} value. If $F > F_{\alpha,v_1,v_2}$ then H_0 is rejected. This indicates that the variance-covariance structure is not homogenous in each level of BSF. In this case, both univariate and multivariate tests could not used for testing the factor effects. Therefore, the suitable transformation should be applied to data and then the assumptions should be tested again. If the assumptions could not be valid, the non-parametric test should be applied to data.

If the variance-covariance structure is homogenous in each level of BSF then the hypothesis which means the pooled variance-covariance matrix is ideal uniform

$$H_0 = \Sigma_b = \Sigma_0 \quad (7)$$

is tested, where, Σ_b is pooled variance-covariance matrix, Σ_0 is ideal uniform matrix. There are two cases to test the H_0 hypothesis in Eq. 7.

$$\text{If } p \leq 5 \text{ then } q = h_3\,(1 - h_1) \quad (8)$$

statistic is used to test Ho hypothesis, where;

$$h_1 = [p(p+1)^2(2p-3)]/\left[6\left\{\sum_{i=1}^{n}(q-1)\right\}(p-1)(p^2+p-4)\right] \quad (9)$$

$$h_3 = \left[-\sum_{i=1}^{n}(q-1)\log_e(|S_b|/|S_0|)\right] \quad (10)$$

q statistic from Eq. 8 is compared with X^2_{α,v_1} value, where $v_1 = (p^2 + p - 4)/2$. If $q < \chi^2_{\alpha,v_1}$ then H_0 is not rejected.

$$\text{If } p>5 \text{ then the statistic } F = h_3\,[1 - h_1 - (v_1/v_2)]/v_1 \quad (11)$$

is used to test H_0 hypothesis, where h_1, h_3 and v_1 are the same in Eq. 8 $v_2 = (v_1 + 2)/(h_2 - h_1^2)$.

Where:

$$h_2 = [(p-1)p(p+1)(p+2)]/\left[6(\sum_{i=1}^{n}(q-1)^2(p^2+p-4)\right]$$

F statistic from Eq. 11 is compared with F_{α,v_1,v_2} value. If $F < F_{\alpha,v_1,v_2}$ then H_0 is not rejected. Thus, the univariate tests for repeated measures could be used for testing the effects of the factors. If the H_0 hypothesis is rejected, the multivariate tests given below should be used for analysis of the data contains repeated measures.

If the two assumptions given above are not valid, neither univariate nor multivariate tests can be invalid. This leads the re-testing of hypothesis with applied suitable transformations in data[2].

Testing of BSF effects: The Ho hypothesis which will be tested is

$$S_h = P_2'Y'X(X'X)^{-1}T\left[T'(X'X)^{-1}T\right]^{-1}T'(X'X)^{-1}X'YP_2$$

$$H_0 : T_2'MP_1 = 0 \tag{12}$$

Where, P_1 is a matrix of order px1 and its elements are +1, T_2' is a matrix of order n-1xn and addition of the elements of each row is zero, M is parameter matrix.

The test statistic

$$U = \frac{|S_e|}{|S_e + S_h|} \tag{13}$$

is used testing the hypothesis in Eq. 12, where $S_h = P'_1Y'X(X'X)^{-1}T_2[T'_2(X'X)^{-1}T_2]^{-1}T'_2(X'X)^{-1}X'YP_1$ and $S_e = P'_1Y'[I - X(X'X)^{-1}X']YP_1$, where X is design matrix with nxr dimensional, Y is data matrix with rxp dimensional. U statistic from Eq. 13 is compared with U_{α,p,v_1,v_2} value[9].

Where, $v_1 = (n-1)(p-1)$ and $v_2 = (r-n-1)(p-1)$. If $U > U_{\alpha,p,v_1,v_2}$, then, H_0 is not rejected[6].

Testing of WSF effects: The H_0 hypothesis which will be tested is

$$H_0 : T_2'MP_2 = 0 \tag{14}$$

Where, P_2 is a matrix of order nxn-1 and addition of the elements of each row is zero. T_1 is a matrix of order nx1 and its elements are +1.

The test statistic given in Eq. 13 is used testing of Ho hypothesis in 14. Where, $S_h = P'_2Y'X(X'X)^{-1}T_1[T'_1(X'X)^{-1}T_1]^{-1}T'_1(X'X)^{-1}X'YP_2$ and $S_e = P'_2Y'[I - X(X'X)^{-1}X']YP_2$ U statistic from Eq. 13 is compared with U_{α,p,v_1,v_2} value[9], where, $v_1 = p - 1$ and $v_2 = (r - n - 1)(p - 1)$. If $U > U_{\alpha,p,v_1,v_2}$, then, H_0 is not rejected.

Testing of interaction effect: The H_0 hypothesis which will be tested is:

$$H_0 : T_2'MP_2 = 0 \tag{15}$$

Where, P_2 and T_2 is a matrix given in Eq. 12 and 14. The test statistic given in Eq. 13 is used testing of H_0 hypothesis in 15, where, $S_h = P'_2Y'X(X'X)^{-1}T_2[T'_2(X'X)^{-1}T_2]^{-1}T'_2(X'X)^{-1}X'YP_2$ and $S_e = P'_2Y'[I - X(X'X)^{-1}X']YP_2$ U statistic from Eq. 13 is compared with U_{α,p,v_1,v_2} value[9], where, $v_1 = (n-1)(p-1)$ and $v_2 = (r-n-1)(p-1)$. If $U > U_{\alpha,p,v_1,v_2}$ then H_0 is not rejected.

RESULTS AND DISCUSSION

The data regarding diastase activity was used for illustration of the assumption tests for validation and multivariate test of the Honey origins, period and Honey origins X period interaction effects on diastase activity.

Validation of the assumptions: In the first stage, $H_0 = \sum_1 = \sum_2 = ... = \sum_n$ hypothesis related with the variance-covariance structure is homogenous in each level of BSF was tested and because of being $p \le 5$ then q statistic from Eq. 3 was computed as q = 13.17. H_0 hypothesis was not rejected. Thus, the variance-covariance structure is homogenous in each honey origin for diastase data.

In the next stage $H_0 = \sum_b = \sum_0$ hypothesis related with pooled variance-covariance matrix is ideal uniform matrix was checked. Because of being $p \le 5$ then q statistic from Equation 8 was computed as q = 17.460. So $17.460 > \chi^2_{0.05,\,8} = 15.507$ that H_0 hypothesis was rejected and thus, the pooled variance-covariance structure from all treatment groups is not homogenous. In this case, because of not validation of the assumption, which is the pooled variance-covariance structure from all honey origins is homogenous, for the data regarding diastase activity, the multivariate test method of repeated measures was applied to diastase activity data instead of univariate

Table 1: The univariate and multivariate test results of the Honey Origin, Period and Honey Origin x Period interaction effects on diastase activity

	Univariate method			Multivariate method (U statistic)	
SOV	DF	MS	Sig.	U	Sig.
Between subject					
Honey origin	2	31.452	**	0.18399	*
Error-1	12	10.622			
Within subject					
Period	3	7.067	**	0.25695	**
Honey origin x period	6	1.733	-	0.12617	*
Error-2	36	1.394			

-: not significant, *: Significant at $\alpha = 0.05$, **: Significant at $\alpha = 0.01$

Table 2: The effects of honey origin and period on diastase activity

Honey origins	Period 1	Period 2	Period 3	Period 4	Total
Orange	12.9±1.37	9.58±1.28	8.10±0.80	7.22±0.44	9.45±0.69a
High plateau	19.94±1.24	15.50±0.97	16.3±0.97	10.98±0.88	15.68±0.87b
Cotton	20.16±1.37	13.94±1.07	19.94±1.24	12.18±1.13	16.55±1.02b
Total	14.02±0.466a	12.94±0.535ab	13.56±0.535ab	12.46±0.44b	

*Different letters refer to statistically differentness ($p<0.05$)

test method to test the effects of the honey origins, period and the honey origins X period interaction on diastase activity. In addition, the two-way repeated measurement design which is one of the univariate tests for repeated measures[7,8], was applied to diastase data to compare the results of each method and given in Table 1. The source of between subject variations consists of honey origins and Error 1 and within subject variation, consists of period, period x honey origins and error 2. Error 1 was originated from variation hive in the same honey origin and period. Error 2 was originated from error 1 x period interaction effect. In this study, honey origins effect was analyzed according to error 1, while period effect was found to be significant ($p<0.01$) but honey origins X periods interaction effect was not significant on diastase activity ($p>0.05$).

In addition, the variation of the honey origins means versus the periods for diastase activity is shown in Table 2 which indicates however diastase activity in orange and high plateau honeys has similar variation for different periods; variation cotton honey is different from other two honeys. Therefore results clearly showed that diastase activity variation for different honey origins (Honey origins X Period interaction effect) is significantly different in different periods (Table 2).

Effects of BSF (Honey origins): U statistic for testing of honey origins effect from Eq. 13 was computed as U = 0.25695 and critical values concerning with U statistic as $U_{0,\ 05,\ 4,\ 3,33}$ = 0.25265 and $U_{0,\ 01,4,}$ = 0.31167. Due to 0.18399<0.25265, H_0 hypothesis was rejected. That was to say there were significances differences among the honey origins. The same as, the univariate test results of this parameter also showed that there were significant differences among the honey origins on Diastase activity (Table 1).

Effects of the WSF (period): U statistic for testing of the period effect was computed as U = 0.1267 and the critical values concerning with U statistic as $U_{0,\ 05\ 4,\ 3,\ 33}$ = 0.4934 and $U_{0,\ 01,\ 4,\ 3,\ 33}$ = 0.41451. Due to 0.256953<0.41451, H_0 hypothesis was rejected. That was to say there were significant differences among the periods on diastase activity. The similar inference was obtained from univariate test of that characteristic (Table 1).

Effects of WSF (Honey origin X Period interaction): U statistic for testing of honey origins treatment X Period interaction effect was computed as U = 0.126173 and the critical values concerning with U statistic as $U_{0,\ 05,\ 4,\ 6,\ 33}$ = 0.25265 and $U_{0,\ 01,\ 4,\ 6,\ 33}$ = 0.31167. Due to 0.126173<0.25265, H_0 hypothesis was rejected. That was to say honey origins X Period interaction was significant on diastase activity ($p<0.05$). Otherwise, univariate test results showed that honey origins X Period interaction was not significant on diastase activity. That is different result compared the univariate test method. That case indicates that the wrong statistical techniques causes to faulty decisions.

CONCLUSIONS

When H_0 hypothesis is rejected, univariate analysis methods will be valid only in some extent. These analyses will be invalid in the case of time period and time period X treatment interaction. In the event of mismatching of hypothesis after statistical test, the method of Cole and Grizzle[10], Multivariate Test Statistics, can be used to assess the effect of honey origins, period and honey origins X period interaction as has been done in the current study. In addition to this method, Hotelling T^2 test statistics, which is also multivariate test, can be used.

The analysis of data in a multivariate analysis of variance of repeated measures (RMANOVA) is quite complex. The complexity of RMANOVA sources largely from the fact that RMANOVA is quite sensitive to violations of one of major assumptions of the test. If data for one or more of samples to be analyzed by repeated measures factor ANOVA come from a population whose distributions violates the assumption of normality, or outliers are present, the ANOVA on the original data may provide misleading results, or may not be the most powerful test available. In such cases, transforming the data or using nonparametric tests may provide a better analysis. Transformations are applied to correct the problems of non-normality or unequal variances. Nonparametric tests are tests that do not make the usual distributional assumptions of the normal-theory-based tests. For the one-way RMANOVA and two-way RMANOVA, the most common nonparametric alternative tests are Friedman's test[11].

According to current assessment to experimental data, it can be concluded that multivariate analysis of variance of repeated measures method will replace the standard univariate repeated measures techniques that occur in practice when univariate repeated measures model assumptions are not satisfied. Thus, the researcher will make a true statistical decision.

REFERENCES

1. Grieve, P.A., 1984. Tests of sphericity of normal distributions and the analysis of repeated measures designs. Psychometrika, 49: 257-267.
2. Bek, Y. and E. Efe, 1986. The analysis of repeated measures for t≥2. Ç.U. J. Agric. Fac., 1: 52-66.
3. Geisser, S. and S.W. Greenhouse, 1958. An extension of Box's results on the view of the F distribution in multivariate analysis. Anal. Math. Stat., 29: 885-891
4. Jaccard, J. and L. Ackerman, 1985. Repeated measures of means in clinical research. J. Consulting Clin. Psychol., 53: 426-428.
5. SPSS, 1999. SPSS for Windows. Release 10.01 Version. SPSS Inc.
6. Timm, N.H., 1980. Multivarate Analysis of Variance of Repeated Measurements. Krishnaiah, P.R. (Ed.), Handbook of Statistics. North-Holland Publishing Company. Holland.
7. Gill, J.L., 1986. Repeated measurement: Sensitive tests for experiments with few animals. J. Anim. Sci., 63: 943-954.
8. Gill, J.L., 1988. Repeated measurement: Split-plot trend analysis versus analysis of first differences. Biometrics, 44: 289-297.
9. Gill, J.L., 1981. Design Analysis of Experiments in the Animal and Medical Sciences. The Lowa State University Press. USA.
10. Cole, J.W.L. and J.E. Grizzle, 1966. Application of multivariate analysis of variance to repeated measurements experiments. Biometrics, 22: 810-822.
11. Dytham, C., 2000. Choosing and Using Statistics. Blackwell Science Ltd. Great Britain.