Vol. 4, No.6 June 2014 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

On Improving Precision of a Repeated Measures Analysis of Variance

[1] Adetutu, O. M, [2] Umar, A. E, [3] Lanlege, D. I

[1] Department of Mathematics and Statistics, Federal University of Technology, Minna, Niger State, Nigeria
[2] Department of Mathematics and Computer Science, Niger State Polytechnics, Zungaru, Nigeria
[3] Department of Mathematics and Computer Science, Ibrahim Badamasi Babangida University, Lapai, Niger State, Nigeria

[1] xtutu2@yahoo.com

ABSTRACT

This paper investigates the use of repeated measures analysis of variance (RMANOVA) in analyzing repeated measurements. We focus on how serial auto correlation of repeated measures data can be removed in order to improve the validity of decision on repeated measures since this is widely used tool of applied statistician because its users are confronted with what seems to be a myriad of decisions, even in its simple application. In this study, the higher order interactions $(a.w.g)$ and $(day * a.w.g)$ are used as $error_1$ and $error_2$ in order to improve the precision of our result. Ten (10) data sets were used to justify this paper and the computation was done with the aid of Statistical Package for Social Sciences (SPSS). The test on the overall estimates was significant at $5\%\ (\alpha = 0.05)$

Keywords: *Autocorrelation, Between-Subject, Covariance, Interaction, Measures and Within-Subject*

1. INTRODUCTION

Repeated measures analysis is a widely used tool of applied statistician. However, its users are confronted with what seems to be a myriad of decisions, even in its simple application. Repeated measurements arise in many diverse fields, and are possibly even more common than single measurements.

What distinguishes such observations from those in more traditional statistical data modeling is that the same variable is measured on the same observational unit more than once Crowder and Hand (1990).

In longitudinal studies, individuals may be monitored over a period of time to record the developing pattern of their observed values. Over such a period, the conditions may be deliberately changed, as in crossover trials, to study the effect on the individual. Even in studies which are not intentionally longitudinal, once a sample of individual units has been assembled, or identified, it is often easier and more efficient in practice to monitor and observe than repeatedly rather than to discard each one after a single observation and start afresh with another sample (Keselman et al, 1999a).

Repeated Measures Designs (RMDs) are quite versatile, and researchers used many different designs and call the designs by many different names. For example, a one way repeated measures ANOVA may be considered as a one-factor within subjects ANOVA. Two-way repeated measures ANOVA may be referred to as a two-way within subjects ANOVA. These designs are called related samples models, matched samples models, longitudinal studies and within-subjects designs (Montgomery, 1992).

The need for repeated measures analysis rather than classical approach is given below according to Hand and Crowder (1999):

a. In drug, nutrition or learning experiments where the objective is to determine the effect of

   different sequences of treatment applications on subjects of experimental units.

b. When the objective is to discover whether or not a trend can be traced among the responses obtained by successive applications of several treatments on a single experimental unit.

c. For some experiments, the experimental units are scarce and expensive and have to be used repeatedly. For example, in small clinics, or in the development of large military systems such as Aerospace vehicles, Aero planes, radar, computers, human beings or animals, etc.

d. In situations where the nature of the experiment calls for special training of experimental units over a long period of time. In order to minimize cost and time, the experimenter should take advantage of the trained experimental unit for repeated measurements.

e. In some experiments where the treatments effects do not have a serious damaging effect on the experimental units can be used for successive experiments.

The basic data format may be identified tersely as '*n* individuals $\times$ *P* measurements'. The individuals may be humans, litters of animals, pieces of equipment, geographical locations, or any other units for which the observations are properly regarded as a collection of connected measurements. The measurements on an individual are recorded values of a variable made at different times. If more than one variable is recorded at each time, the data form a three-way array: individuals' $\times$ measurements $\times$ variables (1999b).

Vol. 4, No.6 June 2014 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

Crowder and Hand (1990) proposed a natural partitioning of the variation in (α, β) into a between-cattle groups component and a residual within-groups components in the ANOVA tradition. If each measurement come from a different individual that is,independent without the complication of being repeated, the traditional approach is to perform nevertheless a standard ANOVA, but using a standard ANOVA in these case of repeated measurement is not appropriate because it fails to model the correlation between the repeated measures: the data violate the assumption of independence (Boik,1997).

Complication of being repeated, the traditional approach is to perform nevertheless a standard ANOVA (Keselman et al, 2001). But using a standard ANOVA in these case of repeated measurement is not appropriate because it fails to model the correlation between the repeated measures: the data violate the assumption of independence, (Crowder and Hand ,1990). Also, work of Winer (1971) have shown that α_i^{C} and α_{ij}^{CD} are uncorrelated.

2. AIM AND OBJECTIVES OF THIS PAPER

The aim of this paper is to x-ray the latest developments in repeated measures data analysis strategies as well as improving the precision of repeated measurements data result. The objectives are to:

a. Analysis the repeated measures involving two sources of variations (milk production from the exotic cattle which is subdivided into four groups and days which serves as the time the produced by the cattle is measured).
b. Ascertain the degree of significance of Cattle and Days as the two sources of variations by sacrificing the higher order interactions (Cattle within Group and interaction between Days $\times$Cattle within Group) as $error_1$ and $error_2$ using Repeated Measures Analysis of Variance (RMANOVA).

3. DATA USED

The data for this paper is a primary data of repeated measurements of the experiment on a test protein diet on the milk production of exotic breed of cattle called Holstein during the first three weeks of locatation investigated at Maizube Farms limited, Minna and can be found in Adetutu (2008). Four groups of cattle numbering nine, seven, ten and nine for groups 1, 2, 3 and 4 respectively were used in the experiment. Group 1 is a control diet while groups 2, 3 and 4 are with 10%, 20% and 40% protein replacement respectively.

The milk production were recorded on alterante days in decilitres and the questions concern the possible differences in milk production profiles between groups as well as knowing differences in milk production profiles between the groups.

The format of the data collected from the experiment carried out was identified tersely as '*n* individuals $\times$ *P* measurements'. Each of the cattle in their respective groups were being monitored over a period of 21 days to record the increase in pattern of their observed value.

4. MATERIALS AND METHODS

4.1 Assumptions of the Model

Some assumptions are required to formalize the specification of the model $y_{ij} = \mu_{ij} + \alpha_{ij} + e_{ij}$ (Hand and Crowder, 1990). It is revealing to divide these assumptions into three strata namely:

a. The *A*-stratum
b. The *B*-stratum
c. The *C*-stratum

The *B*-stratum concerns notation and terminology.

A₁ The α_{ij} for a given *j*, vary over the population of individuals with mean 0 and variance $\sigma^2{}_{\alpha_j}$. Thus $E(\alpha_{ij}) = 0$ and $V(\alpha_{ij}) = \sigma^2{}_{\alpha_j}$, which is really no restriction at all since any non-zero mean is absorbed in μ_{ij} .

A₂ The e_{ij} , for given *j*, vary over individuals with mean 0 and variance $\sigma^2{}_j$; Thus $E(e_{ij}) = 0$ and $V(e_{ij}) = \sigma^2{}_j$ and just as in *A₁.*, there is no restriction.

The B-stratum specifies interactions between the random components of the model in terms of their covariance, that is, the correlation structure is laid down.

B₁ For different *i*, the α -profiles are uncorrelated, i.e. $c(\alpha_{ij}, \alpha_{i'j'}) = 0$ for $i \neq i'$. For given *i*, $c(\alpha_{ij}, \alpha_{ij'}) = \alpha_{\sigma_{jj'}}$ (homogeneous over *i*, but otherwise as yet unconstrained).

B₂ The errors are all uncorrelated, i.e. $c(e_{ij}, e_{i'j'}) = 0$ if $i \neq i'$ or $j \neq j'$.

B₃ The random components, α_{ij} and e_{ij} , are uncorrelated i.e. $c(\alpha_{ij}, e_{i'j'}) = 0$ for all *i*, *j*, i' , j' .

Many of the development proceeds quite happily under *A* and *B*, one has to bring *C* for the strict distributional results which support for example, the

Vol. 4, No.6 June 2014 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

quoting of exact *P*-values for *F*-test. *E*, *V* and *C* will denote respectively Expectation, Variance and Covariance over hypothetical repetitions of the observations with individual as randomly selected from their population.

C The α_{ij} and e_{ij} are normally distributed.

The consequences of assumptions A and B for the means and covariance of the y_{ij} are as follows:

i. $E(y_{ij}) = \mu_{ij}$ Follows simply from the zeroing from the means of α_{ij} and e_{ij} in *A1.* and *A2.*

ii. $c(y_{ij}, y_{i'j'})$ = $E[(y_{ij} - \mu_{ij})(y_{i'j'} - \mu_{i'j'})]$

$= E[(\alpha_{ij} - e_{ij})(\alpha_{i'j'} - e_{i'j'})]$

$= E[(\alpha_{ij}\alpha_{i'j'} + \alpha_{ij}e_{i'j'} + \alpha_{i'j'}e_{ij} + e_{ij}e_{i'j'})]$

Now from A_1 and B_1,

$E(\alpha_{ij}\alpha_{i'j'}) = \delta_{ii'}\sigma_{\alpha_{jj'}}$ Where $\delta_{ii'}$ is the Kronecker delta (Crowder and Hand, 1990), taking values 1 when $i = i'$ and 0 when $i \neq i'$.

Focusing on A_2 and B_2,

$E(e_{ij}e_{i'j'}) = \delta_{ii'}\delta_{jj'}\sigma^2{}_j$

From B_3, $E(\alpha_{ij}e_{i'j'}) = 0$ and $E(\alpha_{i'j'}e_{ij}) = 0$

Therefore, the model specification for the first two moments of the data is

$E(y_{ij}) = \mu_{ij}$

$c(y_{ij}, y_{i'j'}) = \delta_{ii'}(\sigma_{\alpha_{jj'}} + \delta_{jj'}\sigma^2{}_j)$

The observations from different individuals ($i \neq i'$) are uncorrelated, according to the model specification for the first two moment of the data those on the same individual have correlations $\dfrac{\sigma_{\alpha_{jj'}}}{\{(\sigma_{\alpha_{jj}} + \sigma^2{}_j)(\sigma_{\alpha_{j'j'}} + \sigma^2{}_{j'})\}^{1/2}}$.

In the special but commonly applied case where $\alpha^2{}_j = \sigma^2$ and

$\sigma_{\alpha_{jj'}} = \sigma^2{}_\alpha$ (for each j, j') that is $\dfrac{\sigma^2{}_\alpha}{(\sigma^2{}_\alpha + \sigma^2)}$.

This correlation ranges from 0 to 1 as $\dfrac{\sigma^2{}_\alpha}{\sigma^2}$ ranges from 0 to ∞. It measures the strength of the 'personal touch' and as a correlation between different measurements on the same individual, is called an interclass correlation.

It is worthy of mention that starting with basically independent observation, correlation arise naturally through the random effect α_{ij}.

$Y_{(i)k}$ is the effect of k^{th} experimental unit in i^{th} cattle in group i.e. *c.w.g.* (error), $\varepsilon_{(i)jk}$ is the error term i.e. Days× *c.w.g.*(error)

4.2 Model for the Analysis

The fundamental model for this analysis is the partition: of

$y_{ij} = \mu_{ij} + \alpha_{ij} + e_{ij}$ into three components.

y_{ij} denotes the j^{th} measurement made on the i^{th} individual

μ_{ij} denotes the mean level of y_{ij} over hypothetical replication of the set-up with randomly selected individuals from the population. Hence, μ_{ij} is a fixed parameter of the set-up, taking a unique value irrespective of the individual, μ_{ij} is an immutable constant of the inverse.

α_{ij} denotes the consistence departure of y_{ij} from μ_{ij} for the particular individual actually appearing as in our sample. Hence, under hypothetical replications with the same individual, y_{ij} has mean $\mu_{ij} + \alpha_{ij}$. α_{ij} Varies randomly over the population of individuals, it has a random effect and it is a lasting characteristic of the individual.

e_{ij} is the error term representing the departure of y_{ij} from $\mu_{ij} + \alpha_{ij}$ in the particular occasion with the particular individual. e_{ij} is but a fleecing aberration of the moment, no consistency over repetition is accorded to the e_{ij}.

However, if μ_{ij} does not depend on the particular individual, the subscript *i* could be dropped and the fixed effect could be written as μ_j. This is when all the individuals in the sample are treated on an equal footing with no differences in applied treatments or recorded backgrounds to distinguish them. If otherwise, for instance, when there are different treatment groups, it is convenient to retain the subscripts I to indicate the special factor affecting the observations for that individual.

Vol. 4, No.6 June 2014 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

The vector $\mu_i = \left(\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip}\right)'$ corresponds to the *P* measurements on the i^{th} individual.

4.3 Estimate of Repeated Measures Model

The repeated measures involving two sources of variations cattle and day is given by equation (1)

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + Y_{(i)k} + \varepsilon_{(i)jk} :$$
i = 1, 2,…,35 ; j = 1, 2,…, 12; k = 1, 2, 3, 4.

(1)

Where μ is the overall mean

α_i is the effect of i^{th} cattle in a group

β_j is the effect of j^{th} day

$(\alpha\beta)_{ij}$ is the interaction effect of cattle in a Group and Days

The source of total variations is separated into two parts:

The Sum of Square of Cattle (SS_{cattle}) is given by the equation (2)

$$SS_{cattle} = \sum_{i=1}^{35}\sum_{k=1}^{4}\frac{y_{i.k}^{\ 2}}{12} - CT$$

(2)

Where $CT = \left[\sum_{i=1}^{35}\sum_{j=1}^{12}\sum_{k=1}^{4} y_{ijk}^{\ 2}\right] \Big/ 1680$

And the Sum of Square of Days (SS_{Days}) is given by equation (3)

$$SS_{Days} = \sum_{j=1}^{12}\frac{y_{.j.}^{\ 2}}{140} - CT$$

(3)

For a repeated measures of the form given in the equations (1), the Total Sum of Square(SS_{Total}) and Sum of Square of Groups (SS_{Groups}) are given in the equations (4) and (5)

Therefore $SS_{Total} = \sum_{i=1}^{35}\sum_{j=1}^{12}\sum_{k=1}^{4} y_{ijk}^{\ 2} - CT$

(4)

Also, $SS_{Groups} = \sum_{k=1}^{4}\frac{y_{..k}^{\ 2}}{420} - CT$

(5)

The interactions between Days and Cattle is given by the equation (6)

$$SS_{Days\times Cattle} = \left[\sum_{i=1}^{35}\sum_{j=1}^{12}\frac{y_{ij.}^{\ 2}}{4} - CT\right] - SS_{Days} - SS_{Cattle}$$

(6)

In order to increase the precision of the repeated measures, the higher order interactions Cattle within Group $(c.w.g)$ and $Days \times Cattle\ within\ the\ Group\ (Days \times c.w.g)$ are to be sacrificed as $Error_1$ and $Error_2$ (Montgomery, 1991).

The Sums of Squares of the interactions between Cattle within Group and $Days \times Cattle\ within\ Group$ are given by the equations (7) and (8)

$$SS_{Error1} \text{ i.e. } SS_{c.w.g.} = SS_{Cattle} - SS_{Groups}$$

(7)

$$SS_{Error2} \text{ i.e. } SS_{Days\times c.w.g.} = SS_{Days\times Cattle} - SS_{Days\times Group}$$

(8)

$Error_1$ is used to test the effect of cattle while $Error_2$ is used to test the effect of Days and effect of Days$\times$Cattle.

For this study, we have the following hypotheses:
$H_0: D_0 = D_{00} = 0$ *VS* H_1:*Not* H_0

5. RESULTS

The estimate obtained with the aid of Statistical Package for Social Sciences (SPSS) are presented in the tables 1-3 for the model (RMANOVA), Test of Within – Subjects Effects and Test of between- Subjects Effects below:

Vol. 4, No.6 June 2014 ISSN 2225-7217
ARPN Journal of Science and Technology

http://www.ejournalofscience.org

Table 1: Repeated Measures Analysis of Variance

Source of variation	Degrees of freedom	Sums of Squares	Expected means squares	F	Sig.
Days	11	1,514,589.11			
Cattle	34	251,904.66			
Groups	3	104,378.64	7.31	2.28	0.034
c.w.g. (Error1)	31	147,526.02			
Days*Cattle	374	294,851.22			
Days*Groups	33	83,916.01	4.11	1.34	0.003
Days*c.w.g. (Error2)	341	210,935.21			
Total	419	2,061,344.99			

Table 2: Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Days	Sphericity Assumed	1398439.902	11	127130.900	196.833	.000
	Greenhouse-Geisser	1398439.902	1.814	771042.489	196.833	.000
	Huynh-Feldt	1398439.902	2.106	663884.417	196.833	.000
	Lower-bound	1398439.902	1.000	1398439.902	196.833	.000
Days × Groups	Sphericity Assumed	89409.885	33	2709.390	4.195	.000
	Greenhouse-Geisser	89409.885	5.441	16432.316	4.195	.002
	Huynh-Feldt	89409.885	6.319	14148.583	4.195	.001
	Lower-bound	89409.885	3.000	29803.295	4.195	.013
Error(factor1)	Sphericity Assumed	220245.353	341	645.881		
	Greenhouse-Geisser	220245.353	56.225	3917.234		
	Huynh-Feldt	220245.353	65.300	3372.824		
	Lower-bound	220245.353	31.000	7104.689		

Table 3: Tests of Between-Subjects Effects

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	6182004.100	1	6182004.100	1200.697	.000
group	98424.485	3	32808.162	6.372	.002
Error	159609.077	31	5148.680		

6. DISCUSSION OF RESULTS

From the above results, we observed that the milk production profiles of the four groups differ in some way by the protein replacement in the cattle diet as explained by the outcome of our results in (9) and (10) at $\alpha = 0.05$ ($p \le 0.034$). We reject H_0 and conclude that Groups are significantly different. It implied that the four groups of cattle are significantly different from one another.

Similarly, interaction between Days and Groups is found to be significantly different as shown by (10) at $\alpha = 0.05$ ($p \le 0.0030$). This means that there is a change in the milk production profile of the four different groups of Cattle with respect to Days.

However, as a result of autocorrelation of repeated data, transformation (average) was carried out and the result was re presented by equation (12) which shown that the rate of change across (between) the groups is significantly different and suggest that each group have a peculiar rate of change but all the groups have a similar intercept (starting point).

7. CONCLUSION AND RECOMMENDATIONS

In view of the investigator's time, effort and aspirations, it might be thought just a trifle insensitive of statistician to ignore most of his data; thus, the analyses might simply be repeated for each time point (days) but such repeated tests do not in general provide a useful description of the difference response curves.

For the validity of the ANOVA, the data within groups need to be normally distributed, each group having the same variance which is not always true of repeated measures due to autocorrelation in most cases. To have high precision from our Repeated Measures Analysis of Variance (RMANOVA), the higher order interactions have to be sacrificed as error and the transformation of repeated measures data is also necessary in order to remove the autocorrelation from the data and make it more normally distributed.

Hence, it is recommended that effort should be made to improve their precision towards making correct decision.

REFERENCES

[1] Adetutu, O. M. (2008): "Analysis of Repeated Measures". Unpublished M. Sc. Thesis. University of Ilorin, Ilorin. Nigeria.

[2] Runyon, R. P. Coleman, K. R. Pittenger D. J.(2000), Fundamentals of Behavioural Statistics, McGraw-Hill Companies, p.634.

[3] Crowder, M. J. and Hand, D. J.(1990) Analysis of Repeated Measures. Chapman and Hall.

[4] Dytham, C. (2000) Choosing and Using Statistics, Blackwell Science Ltd., Great Britain, p. 218.

[5] Keselman, J. C., Lix, L.., & Keselman, H. J.(1996) :The analysis of repeated measurements : A quantitative research synthesis. British Journal of Mathematics and M Statistics Psychology, 49:275-298.

[6] Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999*a*). The analysis of repeated measurements: A comparison of mixed-model Satterthwaite *F* tests and a nonpooled adjusted degrees of freedom multivariate test. Communications in Statistics—Theory and Methods, 28: 2967-2999.

[7] Keselman, H. J., Algina, J., Kowalchuk, R. K., & Wolfinger, R. D. (1999*b*). A comparison of recent approaches to the analysis of repeated measurements. British Journal of Mathematics and Psychological, 52, 63-78.

[8] Keselman, H. J., Algina, J., Wilcox, R. R., & Kowalchuk, R. K.(2001) :Testing repeated measures hypotheses when covariance matrices are heterogeneous : Revisiting the robustness of the Welch- James test again. Educational and Psychological Measurement, 60, 925 – 938.

[9] Boik, R. J.(1997): Analysis of repeated measures under second-stage sphericity: An Empirical. Bayes approach. Journal of Educational and Behavioural Statistics, 22, 155-192

[10] Montgomery, D.C.(1992): Design and Analysis of Experiments. Third Edition. New York: John Wiley & Sons.

[11] Winer B. J.(1971) Statistical Principles in Experimental Design. Second Edition. New York: McGraw-Hill.

ABOUT THE AUTHORS

Mr. O. M. Adetutu is an Assistant Lecturer in the Department of Mathematics and Statistics, Federal University o Technology, Minna, Nigeria. He holds a Master of Science (M.Sc) degree in Statistics. His research interests are in the areas of Design and Analysis of Experiments and Biostatistics

Dr. A. E. is a principal lecturer in the Department of Mathematics and Computer Science, Niger State Polytechnic Zungaru, Nigeria.

Dr. D. I. Lanlege is a Lecturer II in the Department of Mathematics and Computer Science, Ibrahim Badamasi Babangida University, Lapai, Niger State. He holds a Ph.D. degree in Applied Mathematics from Federal University of Technology, Minna, Niger state, Nigeria. His research interests are in the areas of Biometrics and Applications.