International Journal of Engineering Sciences &Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

RESEARCHERID
THOMSON REUTERS

[Ekerikevwe * *et al.*, 8(5): May, 2019] ISSN: 2277-9655
IC™ Value: 3.00 Impact Factor: 5.164
CODEN: IJESS7

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A GENERALISED LINEAR MODEL FOR FIELD EXPERIMENTS: AN ANCOVA-POST EXAMPLE

K. I. Ekerikevwe[*1] & K. A. Odior[2]
[*1&2]Department of Statistics, School of Applied Sciences & Technology, Delta State Polytechnic, Otefe – Oghara, Nigeria.

DOI: 10.5281/zenodo.3228750

ABSTRACT

In this study, we present a general linear model which blends analysis of variance (ANOVA) and regression when an independent variable has a powerful correlation with the dependent variable and when the independent variables do not interact with other independent variables while predicting the value of the dependent variable. This model is generally applied to balance the effect of comparatively more powerful non interacting variables in order to avoid uncertainty among the independent variables. Data from an observational study with repeated measures (pre-post) were obtained and analysed. The efficiency of the model to determine the differences in means of four treatments before and after adjustment of the field experimental data was discussed. The study was well supported by an empirical example

KEYWORDS: Experiment, Treatments, Model, Repeated measures, Concomitant.

1. INTRODUCTION

A simple repeated measures experiment or observational study in which a subject is assessed twice –once at the beginning of the study and again at the end-analysis of covariance is one of the obvious possibilities for its analysis. In the method, initial observations are being used to measure environmental influences, the object or subject at this stages are yet to be influenced by the treatments. The situation where such measurements are influenced by the treatments, the subjects are reassessed to study the effects of these treatments. Analysis of covariance (ANCOVA) is a technique which is useful for improving the precision of an experiment (Montgomery, 1976; Brookman2017). Analysis of covariance model evaluates whether the means of a dependent variable (DV) are equal across levels of a categorical independent variable (IV) often called a treatment, while statistically controlling for the effects of other continuous variables that are not of primary interest, known as covariates (CV) or nuisance variables. Mathematically, ANCOVA decomposes the variance in the DV into variance explained by the CV(s), variance explained by the categorical IV, and residual variance. Intuitively, ANCOVA can be thought of as 'adjusting' the DV by the group means of the CV(s).In this study, we use the analysis of variance techniques to model the effect of four treatment means on the growth of rubber trees in South-South Nigeria. The field experimental data on height gain of the young rubber trees before (pre) and after (post) the application of the treatments, were obtained from the Rubber Research Institute , Nigeria; the performance of these treatments was well supported with empirical results obtained through analysis of covariance statistical model.

2. LITERATURE REVIEW

Montgomery et al (2017), Seltman (2018) and Pouratian (2002) demonstrate the severity of experimental post-treatment bias analytically and document the magnitude of the potential distortions it induces using visualizations and reanalyses of real-world data. They conclude by providing applied researchers with recommendations for best practice. Frison and Pocock (1992) discuss three methods for analyzing data from pre-post designs: ANOVA with the post measurement as the response variable (ANOVA-POST), ANOVA with the change from pre-treatment to post-treatment as the response variable (ANOVACHANGE), and ANCOVA

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

with the post measurement as the response variable (ANCOVA-POST), adjusting for the pre-treatment measurement. Brogan and Brogan (1980) compare the use of ANOVACHANGE with RANOVA. Corley et al (1971) studied the Morphological characters of oil palm seedlings such as seedling height, number of leaves, stem girth, petiole depth, petiole width, number of leaflets/leaf, number of primary roots, total root volume and biomass of seedlings were recorded at 4 stages at 3 month interval.

O'Connell (2017) worked on repeated measures data that are summarized into pre-post-treatment measurements. They used various methods that exist in the literature for estimating and testing treatment effect, including ANOVA, analysis of covariance (ANCOVA), and linear mixed modeling (LMM). They considered five methods common in the literature, and discuss them in terms of supporting simulations and theoretical derivations of variance. Consistent with existing literature, their results demonstrate that each method leads to unbiased treatment effect estimates, and based on precision of estimates, 95% coverage probability, and power, ANCOVA modeling of either change scores or post-treatment score as the outcome, prove to be the most effective. They further demonstrate each method in terms of a real data example to exemplify comparisons in real clinical context. Calabrese (2011) characterized the stimulus-response function of auditory neurons using a generalized linear model (GLM). In this model, each cell's input is described by: 1) a stimulus filter (STRF); and 2) a post-spike filter, which captures dependencies on the neuron's spiking history. The output of their model is given by a series of spike trains rather than instantaneous firing rate, allowing the prediction of spike train responses to novel stimuli. They fit the model by maximum penalized likelihood to the spiking activity of zebra finch auditory midbrain neurons in response to conspecific vocalizations (songs) and modulation limited (ml) noise. They then compare this model to normalized reverse correlation (NRC), the traditional method for STRF estimation, in terms of predictive power and the basic tuning properties of the estimated STRFs. They find that a GLM with a sparse prior predicts novel responses to both stimulus classes significantly better than NRC.
Among these methods, ANCOVA-POST is generally regarded as the preferred approach, given that it typically leads to unbiased treatment effect estimate with the lowest variance relative to ANOVAPOST or ANOVA-CHANGE [Matthews, (1990); Frison(1992); Brogan and Kutner(1980); Huck (1975); Jennings (1988); Dimitrov (2003)]. Meanwhile, ANCOVA has been criticized as being biased in the case of unequal pre-treatment mean measurements between groups (Samuels (1986); Van Breukelen, (2006)). Liang and Zeger (2000) opine that in the simple case with only two responses (i.e. pre- and post-treatment measurements); ANCOVAPOST produces an unbiased estimate only in the case of equal pretreatment measurements. This study, therefore, uses ANCOVA-POST to evaluate the performance of four treatments. We further dem
onstrate the method in terms of a real data example to exemplify a pre-post field experimental context.

3. METHODOLOGY

When we have heterogeneity in experimental units sometimes restrictions on the randomization (blocking) can improve the test for treatment effects (Grace-Martin, 2019). In some cases, we don't have the opportunity to construct blocks, but can recognize and measure a continuous variable as contributing to the heterogeneity in the experimental units. These sources of extraneous variability historically have been referred to as 'nuisance' or 'concomitant' variables. More recently, these variables are referred to as 'covariates'. When a continuous covariate is included in an ANOVA we have the analysis of covariance (ANCOVA). The continuous covariates enter the model as regression variables, and we have to be careful to go through several steps to employ the ANCOVA method. Inclusion of covariates in ANCOVA models often means the difference between concluding there are or are not significant differences among treatment means using ANOVA. To use a covariate in ANCOVA, we have to go through several steps. First, we need to establish that for at least one of the treatment groups there is a significant regression relationship with the covariate. Otherwise, including the covariate in the model won't improve the estimation of treatment means. Secondly, we have to be sure that the regression relationship of the response with the covariate has the same slope for each treatment group. This is an extremely important point. In our example, we need to be sure that the lines for Males and Females are parallel (have equal slope). Depending on the outcome of the test for equal slopes, we have two alternative ways to finish up the ANCOVA: 1) fit a common slope model and adjust the treatment SS for the presence of the covariate, or 2) evaluate the differences in means at least three levels of the covariate. According to Analysis of covariance (ANCOVA) is useful when you want to improve precision by removing extraneous sources of variation from your study by including a covariate. The analysis of covariance uses features from both analysis of variance and multiple regressions. The covariance analysis is potentially very useful in many ways. Some of which are: It

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

removes the effects which are due to an environmental source of variation. The covariance adjustment, removes biases due to regression that is biases due to

$$\beta\left(\overline{X}-\overline{X}\right)$$

It helps to analyze the nature of treatment effects properly variable is playing its part in producing the treatments effects. It can be profitably employed to eliminate missing values by setting y = 0 for each missing value and introducing a dummy covariance x in such a way that x = 0 for all others. According to Cochran (1950) opined that in covariance analysis, the treatment mean $\overline{y}$, is adjusted by the amount $b(\overline{X} - \overline{X})$, the effect of the adjustment is to change each $\overline{y_j}$, to the value that it would be expected to have if all treatments had the same x mean. It is in this way that the technique removes the effect of variations in the $\overline{X_i}$.Also, during adjustment in covariance analysis, one (1) degree of freedom (df) must be subtracted in the residual SS for the additional regression parameter. ANCOVA is used for several purposes: in experimental designs, to control for factors which cannot be randomized but which can be measured on an interval scale. In regression models, to fit regressions where there are both categorical and interval independents.

The model has the following assumptions: (At least one categorical and at least one interval independent). (Low measurement error of the covariate). (Homogeneity of covariate regression coefficients; i.e. "parallel lines model") The covariate coefficients (the slopes of the regression lines) (the more likely it is to make Type I errors - accepting a false null hypothesis). There is a statistical test of the assumption of homogeneity of regression coefficients.

(Additivity) ANCOVA is robust against violations of additivity but in severe violations the researcher may transform the data, as by using a logarithmic transformation to change a multiplicative model into an additive model. Note, however, that ANCOVA automatically handles interaction effects and thus is not an additive procedure in the sense of regression models without interaction terms. (Independence of the error term) The error term is independent of the covariates and the categorical independents. Randomization in experimental designs assures this assumption will be met. (Independent variables orthogonal to covariates) The independents are orthogonal. If the covariate is influenced by the categorical independents, then the control adjustment ANCOVA makes on the dependent variable prior to assessing the effects of the categorical independents will be biased since some indirect effects of the independents will be removed from the dependent. (Homogeneity of variances) There is homogeneity of variances in the cells formed by the independent categorical variable Heteroscedasticity is lack of homogeneity of variances, in violation of this assumption. (Multivariate normality) For purposes of significance testing, variables follow multivariate normal distributions. (Compound sphericity) The groups display sphericity (the variance of the difference between the estimated means for any two different groups is the same. Tests or adjustments for lack of sphericity are usually actually based on possible lack of compound symmetry.

4. THE MODEL

A mathematical model may be formulated that underlies each analysis of variance. This model expresses the response variable as the sum of parameters of the population. For example, a linear mathematical model for a two factor experiment is given as:

$$Y_{ij} = \mu + t_i + \beta\left(\overline{X}_{ij} - \overline{X}_{oo}\right) + e_{ij}$$

Y_{ij} = total effect

μ = true mean

t_j = treatment effect

$\overline{X}_{ij}$ = mean of the corresponding independent variable on which the Y_{ij} has a linear regression with β as a regression coefficient. $\overline{X}_{..}$ = the grand mean. e_{ij} = Error effect

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Note that this model is the sum of various constants. This type of model is called a linear model. It becomes the mathematical basis for our discussion of the analysis of covariance.

Table 1: Formula Table for ANCOVA-POST Adjustment

Source	Df	Regression SS	Adjusted SS	MS	F
Treatment	T-1	RSS_{Total} - RSS_E	$ASS_T = ASS_{Total} - ASS_E$	$\Sigma X_j EY_j/tn$	
Error	n-t-1	$RSS_E = b(SXY_E)$	$ASS_E = SSY_E – b(SXY_E)$	$MS_E = ASS_E/df_E$	
Total	N-2	$RSS_{Total} = b(SXY_{Total})$	$ASS_{Total} = SSY_{Total} – b(SXY_{Total})$		

5. DATA EXAMPLE

The data for this research were obtained from Rubber Research Institute in Nigeria. As a way of determining the growth of the young rubber trees, data on height of young rubber trees before (pre) and after treatments application were collected and analysed. The heights of ten (10) randomly selected young rubber trees were observed first before the application of the treatment (pre) and second after the application of the treatments (post). The interval between the pre-post observations was twelve (12) months. The measurements were taken in centimetres. The pre-post field experimental data are presented below.

Table 2: The table of heights gain per palm before and after fertilizers application.

S/N	PLOT I K		PLOT II Mg		PLOT III P		PLOT IV N	
	(Pre) Cm	(Post) cm	(Pre) Cm	(Post) cm	(Pre) Cm	(Post) cm	(Pre) Cm	(Post) cm
1	16.1	115.5	25.0	87.4	36.1	111.0	17.4	125.0
2	14.4	113.4	22.2	76.5	18.1	110.2	15.3	125.4
3	17.2	113.2	20.7	86.2	18.2	113.3	17.2	115.5
4	9.0	116.6	36.5	81.0	25.4	112.1	16.4	135.4
5	12.3	105.3	22.4	95.6	15.0	111.5	11.8	12.7
6	15.1	111.0	23.2	89.4	19.4	115.5	21.3	135.0
7	13.1	111.5	25.3	83.5	23.5	115.6	11.0	115.9
8	12.2	107.1	24.3	83.8	18.2	115.7	15.5	132.5
9	16.4	113.3	29.1	72.5	15.1	112.4	5.4	123.5
10	12.0	111.5	28.4	85.3	26.4	116.9	15.6	117.4

Source: Rubber Research Institute, Iyanomo, Benin City, Nigeria.

6. RESULTS OF ANALYSIS

Table 3: ANCOVA Tests Of Between-Subjects Effects: Pre- Fertilizer Effect

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	904.798[a]	4	226.199	14.631	.000	.626
Intercept	1392.426	1	1392.426	90.066	.000	.720
PRE_FERTILIZER_EFFECT	4.698	1	4.698	.304	.585	.009
GROUP	861.185	3	287.062	18.568	.000	.614
Error	541.102	35	15.460			
Total	15810.000	40				
Corrected Total	1445.900	39				
R-Squared = .626	Adjusted R^2 =.583					

Table 4: ANCOVA-POST Pairwise Comparisons of Treatment Effects

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
K	Mg	5.514	2.081	.072	-.307	11.334
	P	.543	2.015	1.000	-5.094	6.179
	N	-8.064*	1.810	.000	-13.125	-3.003
Mg	K	-5.514	2.081	.072	-11.334	.307
	P	-4.971*	1.763	.047	-9.902	-.040
	N	-13.578*	1.887	.000	-18.855	-8.300
P	K	-.543	2.015	1.000	-6.179	5.094
	Mg	4.971*	1.763	.047	.040	9.902
	N	-8.607*	1.844	.000	-13.765	-3.449
N	K	8.064*	1.810	.000	3.003	13.125
	Mg	13.578*	1.887	.000	8.300	18.855
	P	8.607*	1.844	.000	3.449	13.765

Based on estimated marginal means
*. The mean difference is significant at the .05 level.
b. Adjustment for multiple comparisons: Bonferroni.

Table 5: Adjusted ANCOVA and Adjusted Y Means for ANCOVA-POST (Height)

Source of Variation	**Sum of Squares REGRESSION**	**Degree of Freedom**	**Sum of Squares ADJUSTED**	**Mean Square**	$F_{calculated}$
Total	-1.58 x -2575.28	38	6648.74		
Error	0.007	35	1396.28 – 0.87	39.89	
Fertilizer		3	4282.42	1427.51	35.79
Fertilizer	$\overline{X}_j$	$\overline{X}_j - \overline{X}_{oo}$	$b(\overline{X}_j - \overline{X}_{oo})$	$\overline{Y}_j$	$\overline{Y}_j - b(\overline{X}_j - \overline{X}_{oo})$
K	12.40	-4.44	0.04	107.11	107.07
Mg	24.51	7.67	-0.08	78.26	78.34
P	16.42	-0.42	0.00	107.92	107.92
N	14.04	-2.80	0.03	119.71	119.68

Discussion of Results

From Table 3, we can see that the pre-treatments effect does not differ significantly from each other in PLOT I-IV. However, Table 4 compares the post-fertilizer effects which shows that there is a significant differences in the performance of the four treatments after 12 months of their application on the subjects with Sulphate of Ammonia (N) having the most significant contribution to the growth of the young trees since $P(0.00) < P(0.05)$. Also, the adjusted Y means of the height measurements using the ANCOVA-POST model $\overline{Y}_j - b(\overline{X}_j - \overline{X}_{oo})$ were computed with relative efficiency as shown in Table 5. The results show that the treatment (Sulphate of Ammonia, as denoted by N) has the overall most significant contribution on the performance of the young rubber trees with respect to their growth.

7. SUMMARY AND CONCLUSION

Using the ANCOVA-POST, the study has found that young rubber trees benefit immensely from nitrogen through the application of sulphate of ammonia; hence, Nitrogen nutrients or fertilizers should always have the

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

highest proportion whenever fertilizers are applied on young rubber trees both at nursery and field stages. Same quantity or proportion could be used for both potassium and phosphate fertilizers. Magnesium should only be provided in small quantity in the soil where palms are grown to avoid absolute deficiency in it. Sulphate of potash and rock phosphate have effects which do not differ significantly from each other and their effects are less on the growth of oil palms when compared to that of sulphate of ammonia. Finally, sulphate of magnesia has the lowest effect on the growth of young rubber trees. This paper therefore recommends that, in field experiments, where the subject was assessed twice-once at the beginning of the study and again at the end; a generalised linear model of ANCOVA-POST will produces an unbiased estimates when the pre-treatment measurements are not differ significantly.

REFERENCES

[1] Brogan, D.R., Kutner, M.H. (1980) Comparative analyses of pretest-posttest research designs. The American Statistician.
[2] Broockman, David E., Joshua L. Kalla, and Jasjeet S. Sekhon (2017) "The Design of Field Experiments With Survey Outcomes: A Framework for Selecting More Efficient, Robust, and Ethical Designs." Forthcoming, Political Analysis. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract .
[3] Calabrese, A., Schumacher JW, Schneider DM, Paninski L, Woolley SMN (2011) A Generalized Linear Model for Estimating Spectrotemporal Receptive Fields from Responses to Natural Sounds. PLoSONE6(1):e16104. https://doi.org/10.1371/journal.pone.0016104
[4] Cochran, W. G. (1950). "Experimental Designs." John Wiley & Sons, USA.
[5] Corley, R. H. V. Hardon, J. J. Tan, G. Y. (1971). Analysis of growth of the oil palm (ElaeisguineensisJacq.) I. Estimation of growth parameters and application inbreeding. Euphytica 20(2):307-315 • DOI: 10.1007/BF00056093.
[6] Dimitrov DM, Rumrill J, Phillip D (2003) Pretest-posttest designs and measurement of change. Work: A Journal of Prevention, Assessment and Rehabilitation.
[7] Frison L, Pocock SJ (1992) Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Statistics in medicine 11: 1685-1704.
[8] Grace-Martin, K. (2019) Analyzing Pre-Post Data with Repeated Measures or ANCOVA at https://www.theanalysisfactor.com/pre-post-data-repeated-measures/
[9] Huck SW, McLean RA (1975) Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. Psychological Bulletin 82: 511.
[10] Jennings E (1988) Models for Pretest-Posttest Data: Repeated Measures ANOVA Revisited. Journal of Educational and Behavioral Statistics 13: 273-280.
[11] Liang KY, Zeger SL (2000) Longitudinal data analysis of continuous and discrete responses for pre-post designs. SankhyÄ: The Indian Journal of Statistics Series B: 134-148.
[12] Matthews J, Altman DG, Campbell M, Royston P (1990) Analysis of serial measurements in medical research. BMJ: British Medical Journal.
[13] Montgomery, D.C. (1976). "Design & Analysis of Experiments". John Wiley & Sons N.Y.
[14] Montgomery, J. M., Nyhan, B. and Torres, M. (2017) How conditioning on post-treatment variables can ruin your experiment and what to do about it. American Journal of Political Science Dataverse within the Harvard Dataverse Network at http://dx.doi.org/10.7910/DVN/EZSJ1S.
[15] Pouratian, N. and Toga, A. W. (2002). Brain Mapping: The Methods (Second Edition), Science https://www.sciencedirect.com/topics/neuroscience/generalized- linear-model. Direct Publisher.
[16] Samuels ML (1986) Use of analysis of covariance in clinical trials: a clarification. Controlled clinical trials 7: 325-329.
[17] Seltman, H.J.(2018) Experimental Design and Analysis @http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf July.

[18] O'Connell NS, Dai L, Jiang Y, Speiser JL, Ward R, et al. (2017) Methods for Analysis of Pre- Post Data in Clinical Research: A Comparison of Five Common Methods. J Biom Biostat 8: 334. doi:10.4172/2155-6180.1000334

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

[19] Van Breukelen GJ (2006) ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. Journal of clinical epidemiology.

Appendix A

Main Treatment Combinations Used In the Field Experiment

K: As sulphate of potash in 1:1:1:2 at 56grams per rubber tree
Mg: As sulphate of magnesia in 1:1:1:2 at 56grams rubber tree
P: As rock phosphate in 1:1:1:2 at 56grams rubber tree
N: As sulphate of ammonia in 1:1:1:2 at 56grams rubber tree