Abstract No: PS-08

Reduction of experimental error in coconut research by choosing proper data analysis techniques

W. H. H. Fernando[1*], K. P. Waidyarathne [2] and D. D. M. Jayasundara[1]

[1]Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka
[2]Plant Physiology Division, Coconut Research Institute, Sri Lanka
himashaharshani635@gmail.com*

Accurate data analysis techniques are essential in field experiments to correctly understand the influence of independent variables on the dependent variable/s. This study compares different statistical techniques used in analyzing longitudinal data (nut yield data collected in multiple years) of a coconut field experiment. Longitudinal studies are necessary for coconut research due to its perennial nature. However, these experiments often have high variability due to the heterogeneity nature of coconut palms, where individual palms display inconsistent temporal behaviour. High variation among the individuals in similarly treated plots makes treatment mean sensitive to those fluctuations ultimately masking the true treatment effect. Even careful planning of the experiment cannot ensure the total elimination of this component. The study highlights the ways in which how this unaccountable variability should be handled to obtain a precise research output. There are many types of statistical techniques used in analyzing data from different experimental designs to achieve optimal research outcomes. This study compares different statistical techniques applied to a randomized complete block design (RCBD), the most frequently used experimental design in coconut research, using a long term coconut fertilizer study. The example illustrates the appropriate types of analyses to meet the precise analysis output by evaluating the model residuals and the Coefficient of Variations (CV). CV, the ratio of the standard deviation to the mean (total average of the design), is a measure of relative variability. In particular ANOVA, Mean Square Error of ANOVA can be used as the standard deviation of the design because Standard Error (SE) of a statistic (usually an estimate of a parameter) represents the standard deviation of its sampling distribution. In this study, Repeated Measures Analysis of Variance (RMANOVA) was used as the classical method. Improved methods used were Linear mixed model and Multivariate Analysis of Variance(MANOVA) with two principal components (representing ≥ 78% variation of the data) as dependent variables. Adequacy of all methods was accepted after checking normality with the Shapiro-Wilk test, homogeneity of variance with Levene's test, and independence of residuals with the Box-Pierce test. CV resulted from RMANOVA applied on RCBD was 39.2%, while it was 16.51% from the Liner mixed model. The lowest CV (10.04%) resulted from MANOVA with two principal components indicates that it can be more efficiently used to analyze long term experiments of coconuts. The consistency of the results should be studied further with a few more similar kinds of data sets. In addition to the above statistical analysis techniques, Bayesian inference methods will be studied for further improvements in the results.

Keywords: Coefficient of Variance, Mean Square Error, Randomized Complete Block Design, Repeated Measures Analysis of Variance