

Abstract No: PO-48

A complex sample design based analysis of determinants of secondary job holding of heads of households in Sri Lanka

P. A. H. S. Pannala^{1*}, K. A. D. Deshani¹

¹ Department of Statistics, University of Colombo, Sri Lanka
pahspannala@gmail.com*

Heads of Households (HHs') play the most prominent role in families. This study was mainly focused on HHs' who were engaged in secondary jobs (SJ) in Sri Lanka. Exploring methods and techniques of incorporating complex sample design into analysis and identifying factors associated with secondary job enrolment of HHs' in Sri Lanka by developing a statistical model using complex sample design were main objectives of the study. The data of the Annual Labour Force Survey-2019 was collected from the Department of Census & Statistics. The sampling method of the survey was two-stage stratified. A sample where sampling method was not simple random sampling is called a complex sample. The study emphasized that analysing and modelling techniques developed for simple random samples could not be used for data from a complex sample. Thus, study used descriptive, univariate, bi-variate analysis techniques developed for complex sample data. Since both sample design and weights were used analysis is design-based. Complex sample binary logistic regression modelling techniques were used to acquire the status "being secondary employed or not". High class imbalances were noticed, and over-sampling, under-sampling, and SMOTE techniques were tried, and under-sampling was selected as the best way of balancing the data. Two models were developed, for the original data and for the under-sampled data. Nagelkerke R square and classification table were considered to compare the overall fit of two models. The overall accuracy of models developed for original and under sampled data given by classification table was 90.4% and 71.2% respectively. But class-wise accuracy was high in model fitted for balanced data set with under-sampling. Model developed for original data correctly predicts only 0.8% of SJ enrolment of HHs' and 99.9% correctly non-SJ enrolment while model fitted for balanced data correctly predict 41.1% of the SJ enrolment of HHs' and 86.9% non-secondary job enrolment. Model developed with interaction terms for the balanced data set had overall accuracy 74.1% which is not a considerable improvement compared to main effects model for the same dataset with 71.2% overall accuracy and, consisted of more predictor variables again led to choose main effect model. According to the model derived for the balanced data, being in rural sector, being male, engaging in occupations managers, senior officials and legislators, professionals, services and sales workers, skilled agricultural, forestry and fishery workers, monthly income below rs.15,000.00, increased the likelihood of being a secondary employed HH. Moreover, no published work could be found on a design-based analysis on this topic. Thus, study leads the government and employers to find out determinants of SJ holding of employed HHs' and to adjust their plans and improve the quality of primary and SJs.

Keywords: Balanced data, Complex sample, Design based, Heads of households, Secondary job