

Variance Estimation in the Presence of Nearest Neighbor Imputed Data

¹H. Temesgen, B. Eskelson², and T. Barrett

Abstract: Missing data is a common problem virtually in all forest inventory and monitoring programs. Nearest neighbor imputation (NN) methods have been used to compensate for missing data and to fill-in selected attributes for non-sampled polygons. Often, the imputed values are treated as if they were observed values. Such procedure is appropriate for estimating totals, means, proportions, and other first-order population quantities if the NN method does not result in sizeable bias. However, treating the imputed data set as the complete data set and using the ordinary variance formulas can provide invalid results in estimating uncertainty (variance) and in conducting statistical inferences. When the proportion of missing data is large, as often happens with forestry applications using nearest neighbor methods, the potential for error in estimated variance increases.

In this presentation, we discuss the suitability and predictive abilities of selected variance estimation techniques under a NN imputation approach, and examine bootstrap methods to quantify imputation variance. For basal area and volume per ha, the imputation variance is not severe when the imputation rate is low. We also demonstrate how the magnitude of the imputation variance supports forest monitoring and assessment endeavors.

Key words and phrases: jackknife, bootstrap, bootstrap confidence interval, two-phase sampling

The authors are respectively, Assistant Professor, Department of Forest Resources, Oregon State University; PhD Student, Department of Forest Sciences, Oregon State University; and Research Forester, PNW Research Station, Portland, OR.

¹Corresponding author: Tel. +1 541-737-8549; fax .+1 541-737-3049; Email: hailemariam.temesgen@oregonstate.edu,