



# Regression analysis for exponential family data in a finite population setup using two-stage cluster sample

Brajendra C. Sutradhar<sup>1,2</sup>

Received: 7 January 2022 / Revised: 22 July 2022 / Accepted: 2 August 2022 /  
Published online: 14 September 2022  
© The Institute of Statistical Mathematics, Tokyo 2022

## Abstract

Over the last four decades, the cluster regression analysis in a finite population (FP) setup for an exponential family such as linear or binary data was done by using a two-stage cluster sample chosen from the FP but by treating the sample as though it is a single-stage cluster sample from a super-population (SP) which contains the FP as a hypothetical sample. Because the responses within a cluster in the FP are correlated, the aforementioned sample mis-specification makes the sample-based so-called GLS (generalized least square) estimators design biased and inconsistent. In this paper, we demonstrate for the exponential family data how to avoid the sampling mis-specification and accommodate the cluster correlations to obtain unbiased and consistent estimates for the FP parameters. The asymptotic normality of the regression estimators is also given for the construction of confidence intervals when needed.

**Keywords** Clusters under a finite population · Clusters selected in first stage · Individuals selected in second stage from a selected cluster · Invalid inferences for regression effects using GLS estimates · Doubly weighted estimation · Unbiasedness · Consistency and asymptotic normality

---

✉ Brajendra C. Sutradhar  
bsutradh@mun.ca

<sup>1</sup> Department of Mathematics and Statistics, Memorial University, St. John's, NL A1C 5S7, Canada

<sup>2</sup> Present Address: 28 Musgrave Street, St. John's, NL A1A 5E7, Canada