Assessing selection bias in regression coefficients estimated from nonprobability samples with applications to genetics and demographic surveys.

Summary: Selection bias is a serious potential problem for inference about relationships of scientific interest based on samples without well-defined probability sampling mechanisms. Motivated by the potential for selection bias in: (a) estimated relationships of polygenic scores (PGSs) with phenotypes in genetic studies of volunteers and (b) estimated differences in subgroup means in surveys of smartphone users, we derive novel measures of selection bias for estimates of the coefficients in linear and probit regression models fitted to nonprobability samples, when aggregate-level auxiliary data are available for the selected sample and the target population. The measures arise from normal pattern-mixture models that allow analysts to examine the sensitivity of their inferences to assumptions about nonignorable selection in these samples. We examine the effectiveness of the proposed measures in a simulation study and then use them to quantify the selection bias in: (a) estimated PGS-phenotype relationships in a large study of volunteers recruited via Facebook and (b) estimated subgroup differences in mean past-year employment duration in a nonprobability sample of low-educated smartphone users. We evaluate the performance of the measures in these applications using benchmark estimates from large probability samples.

MSC:

62P10 Applications of statistics to biology and medical sciences; meta analysis
62J05 Linear regression; mixed models
62F07 Statistical ranking and selection procedures
92D10 Genetics and epigenetics

Keywords:
linear regression; probit regression; nonprobability samples; selection bias; polygenic scores; National Survey of Family Growth

Full Text: DOI arXiv

References:


[18] Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In \textit{Annals of Economic and Social Measurement} 5 475-492. NBER.


[41] Petrolia, D. R. and Bhattacharjee, S. (2009). Revisiting incentive effects: Evidence from a random-sample mail survey on


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.