Using R for exploring sampling designs at Statistics Norway

Susie Jentoft\(^1\), Johan Heldal\(^1\)

1. Division for Methods, Statistics Norway
* Contact author: Susie.Jentoft@ssb.no

**Keywords:** sampling, official statistics, Rcmdr, RcmdrPlugin.sampling

The use of \( R \) at Statistics Norway is mostly restricted to the Division for Methods. It is used as an analytic, analysis and development tool. However, the benefits of \( R \) over other software packages are growing and we are looking for more ways to integrate the software into our everyday processes.

One area of interest for integration is in the planning phase of survey sampling. We often use complex sampling designs, particularly in face-to-face interview surveys, to cluster the participants in manageable and cost-effective groups. However, clustering participants generally reduces the precision of the estimates from the survey. Investigating this balance between cost and accuracy is an integral part of the planning process. We have been developing a package called \texttt{RcmdrPlugin.sampling} as a tool for exploring sampling designs and selecting samples. This package uses tools from the \texttt{sampling} package and builds them into the \texttt{Rcmdr} interface. This is to give the package a user friendly interface that can appeal to people with little experience of \( R \). It provides a platform for selecting a simple random sample, stratified samples and multi-stage samples with or without stratification. If key variable and cost data are available, estimates for variances and survey costs can be calculated for different stratification and clustering designs.

This package has been designed as a versatile and useful tool for both Statistics Norway and others who deal with planning sample surveys. We hope over time to continue to add to the functionality of the package and to add extra features for enhanced visualisation of the proposed sample.