Real World Sampling & Weighting: Methodology in Social Surveys at NatCen

Angus Mayo | BASS Politics and Data Analytics

Overview of the Data Fellowship

As a Survey Statistician, my role at the **National Centre for Social Research** (**NatCen**) consisted of ensuring the integrity and representativeness of survey data. This work takes place across two stages of the social research process: sampling (before fieldwork is conducted) and weighting (after field data are collected).

All work was conducted across multiple programming languages, including R and the syntaxes of SPSS and Stata.

Approach to Sampling and Weighting

Representative sampling is crucial in social research, as it ensures findings are generelisable and equally reliable across all segments of the population. Within the sampling stage, my work included using historical reponse rates to calculate the optimal sizes of strata from which to draw survey samples. By distributing the sample this way, every demographic stratum can be measured with similar accuracy, so estimates inferred from survey responses have comparable margins of error for each group.

Even with precise sampling calculations, raw survey samples rarely mirror the population, so weights are used to correct these imbalances and reduce bias. At the weighting stage, I implemented automated variable selection through stepwise logistic regression to estimate

A simplified representation of the survey weighting process

non-response and compute adjustment factors. I also conducted calibration weighting, applying constrained optimisation algorithms to adjust survey weights so that the sample aligned with known population benchmarks (e.g., age, sex, region), while assessing where to trim extreme adjustments.

*Examples in the news of **British Social Attitudes** data*

Projects

- Calculated sampling strata sizes for *British Social Attitudes*, NatCen's flagship survey, which will be used in the 2025 round.

- Explored the relationship between design effect and bias through alternative calibration weighting models for the *NatCen Panel*.

- Updated/translated legacy SPSS syntax into R for the *Gambling Survey for Great Britain* to be used in future rounds.

Key Skills Developed

A highly technical placement, my Data Fellowship at NatCen built on a range of statistical, programming and professional skills, including:

- Communicating with researchers and other statisticians to determine and write up sampling specifications;

- Maneuvering and creating advanced functions across Excel spreadsheets, used for computing strata sizes, calculating population totals for calibration and checking bias;

- Logistic regression to model non-response, including stepwise methods for selection of predictors;

- Producing functional, replicable and updateable code within R and SPSS;

- Subjective judgement of extreme weights/adjustments for the purpose of trimming;

- Sharing ideas with team members and contributing to discussion, including chairing a team meeting.