Using Calibration Weighting in Samples with Non-probability Components

Presenter: Jamie Ridenhour and Phillip S. Kott

Collaborators: Matthew Farrelly, Kian Kamyab & Joe McMichael

$$\sum_R d_k \left[1 + \exp(\mathbf{m}_k{}^T \mathbf{g})\right] \mathbf{c}_k = \mathbf{T}_\mathbf{c}$$

⇑ **Model variables** ⇑ **Calibration variables** ⇑ **Calibration targets**

⇑

Our idea of a pretty picture

Overview

- Two ABS samples of the US on attitudes to marijuana needed to be combined with two social-media recruited samples.
- Previously, a similar exercise was conducted in Oregon with one ABS sample and one Facebook-recruited sample.
- Lessons from the latter are applied to the former.

The US Sample Frames

Five Frames

Frame 1 – Mail respondents of first ABS survey

Frame 2 – Web respondents of first ABS survey

Frame 3 – Respondents to first social media survey

Frame 4 – Respondents of second ABS survey

Frame 5 – Respondents to second social media survey

ABS samples were stratified (by state marijuana laws) probability samples of addresses. One adult selected per household.

Frame 1 had to be discarded because age of respondent was not collected.

Survey items for the remaining frames were considered identical

The Oregon Marijuana Study

An ABS of one adult per Oregon household was given a 20-minute questionnaire on marijuana use and attitudes.

Roughly half responded via mail, half Internet

More responses were recruited via Facebook.

Poor response on race and household size questions.

How can we weight the result to draw inferences? (This question was not asked until after the data was collected)

Potential Calibration Variables

Sample Size – 1,989
(mail response – 722; mail-to-web – 640; recruit – 627)

Missing number of adults in household – over 800
(745 for ABS respondents)

Missing race = black – over 1,300

Used to calibrate the ABS sample to the population

Missing Age group (six levels) – 3
Missing Sex – 76
Missing Education (three levels) – 173

Added to calibrate recruit cohort to mail-to-web cohort

In politics TODAY, do you consider yourself ….
Republican, Democrat, Independent,
No preference,
No or invalid answer (*treated as a separate level*)

The Selection Model

The probability that an Oregon adult was sampled and then responded to the ABS survey is assumed to be a logistic function of three categorical variables: age group, sex, and education level. (Better would be to assume only a probability of response, if the probabilities of selection were known)

The probability that an Oregon adult was recruited into the sample via Facebook is assumed to be a logistic function of the above three categorical variables and party affiliation.

The population that would respond by Internet when given the chance (represented by the mail-to-web cohort) is assumed to be the same as the population that could be recruited via Facebook. *An assumption that will be tested.*

SAS/SUDAAN Code

Recruit cohort: TYPE = 1; X = 1; Z = 1; ABS = 0

Mail-to-web cohort: TYPE = 2; X = 0; Z = -1; ABS = 1

Mail cohort: TYPE = 3; X = 0; Z = 0; ABS = 1

```
PROC WTADJX DATA = D ADJUST = POST  DESIGN = WR;
WEIGHT _ONE_;  NEST _ONE_;  LOWERBD 1;  VAR [ ….];
CLASS SEX AGE EDU PARTY;   * after imputing missing values;
MODEL _ONE_ = SEX*ABS  AGE*ABS  EDU*ABS SEX*X  AGE*X
EDU*X PARTY*X/NOINT;   (NOINT = no intercept)
CALVARS SEX*ABS  AGE*ABS  EDU*ABS SEX*Z  AGE*Z  EDU*Z
PARTY*Z/NOINT;
POSTWGT [population totals for the categories,  16 zeroes];
VDIFFVAR TYPE (1,2);
(WTFINAL is the output calibrated weight)
```

Holm-Bonferroni Procedure

The conservative HB procedure is not only a overall multiple comparison test but also assesses each individual comparison.

For 20 items, sort whether there was a response and differences among respondents by their *p*-values.

For HB20_.1:

Difference with lowest *p*-value out of 20 is significant at .1 level if *p*-value is less than HB20_.1 critical value (.1/20).

Difference with second lowest *p*-value is significant at .1 level if *p*-value is less than HB20.1 critical value (.1/19).

Continue until first not-significant difference.

Smallest *p* Values vs Critical Holm-Bonferroni Values

VARIABLE	Estimated difference	*p* value	HB20_.05	HB20_.1
More DUI?	0.11	0.00247	0.00250	0.00500
Edible MJ in public?	-0.23	0.00371	0.00256	0.00526
How legal?	0.11	0.00658	0.00263	0.00556
Adult frequency?	-0.13	0.01619	0.00270	0.00588
Is edible MJ safer?	-0.17	0.02260	0.00278	0.00625
Guest use in home?	-0.18	0.04079	0.00286	0.00667
Is vaping safer?	0.10	0.05260	0.00294	0.00714
More teenage use?	0.12	0.08722	0.00303	0.00769
Response to vaping Q	0.05	0.09704	0.00313	0.00833

Jackknife Weights (from Kott 2006)

Randomly sort ABS and recruit respondent samples.

Systematically assign respondents to one of 30 jackknife groups.

Create the r^{th} set of jackknife replicate weights by setting the replicate weights of respondents in the r^{th} group to zero and multiply the calibrated weight for respondents outside the group by 30/29.

Recalibrate each replicate without a *lowerbd*.

Scale the calibrated and jackknife weights assigned to mail-to-web (by .65) and recruit (by .35) cohorts to eliminate double counting.

Returning to the US Samples

Frame 2 – Web respondents of first ABS survey

Frame 3 – Respondents to first social media survey

Frame 4 – Respondents of second ABS survey

Frame 5 – Respondents to second social media survey

Sample from Frame 4 calibrated to populations in strata, age groups, education groups, and gender.

Sample from Frame 2 calibrated to respondents with internet access in Frame 4 by strata, age groups, education groups, gender, and politics.

Samples from Frame 3 and 5 each calibrated to social media users in Frame 2 by strata, age groups, education groups, gender, and politics.

No testing was done in making these decisions (resource constraints)

Combining the Cohorts to Avoid Double Counting

Divide the respondent sample into the following cohorts:

F3 (first social-media frame),

F5 (second social-media frame),

$F2_{SM}$ (first ABS internet respondents with social media)

$F2_{R}$ (the remaining first ABS internet respondents)

$F4_{SM}$ (second ABS respondents with social media)

$F4_{INT}$ (second ABS respondents with internet but without social media)

$F4_{R}$ (the remaining F4 respondents – mail respondents without social media)

We assume that $F4_{SM}$, $F2_{SM}$, F3, and F5 all represent the same subpopulation after calibration weighting.

We likewise assume that $F4_{INT}$ and $F2_{R}$ represent the same subpopulation after calibration weighting.

Combining the Cohorts to Avoid Double Counting

Compute $n^*_{4INT}=\frac{\left(\sum_{4INT} \text{TMPWGT}_j\right)^2}{\sum_{4INT} \text{TMPWGT}_j^2}$, where TMPWGT_j is the calibrated weight; that is, $n^*_{4INT}= n$/(Unequal Weighting Effect)

Compute the other effective cohort sample sizes analogously.

Assign the respondents in $F4_R$ the final weight $\text{FNLWGT}_k = \text{TMPWGT}_k$.

Composite respondents with internet but without social media in $F4_{INT}$ and $F2_R$:

Assign the respondents in $F4_{INT}$ $\text{FNLWGT}_k = \frac{n^*_{4INT}}{n^*_{4INT}+n^*_{2R}} \text{TMPWGT}_k$.

Assign the respondents in $F2_R$ $\text{FNLWGT}_k = \frac{n^*_{2R}}{n^*_{4INT}+n^*_{2R}} \text{TMPWGT}_k$.

Composite the social-media-using respondents in $F4_{SM}$, $F2_{SM}$, F3, and F5:

Assign the respondents in $F4_{SM}$ $\text{FNLWGT}_k = \frac{n^*_{4SM}}{n^*_{4SM}+n^*_{2SM}+n^*_3+n^*_5} \text{TMPWGT}_k$

and the respondents in $F2_{SM}$, F3, and F5 analogously

Some Concluding Remarks

Think about analysis before data are collected.

Using nonprobability samples relies on assumptions, which need to be clearly stated and tested when possible.

Selection modeling is analogous to nonresponse modeling.

One can run an unweighted logistic regression on the blended sample so long as all the variables used in weighting (stratum, age group, education group, gender, and politics) are covariates in the model. One needs to assume that the model is correct ($E(y_k - p(\mathbf{x}_k)|\mathbf{x}_k) = 0$ for any $\mathbf{x}_k$) and that the "selection" of the respondents is a function of the model covariates.

Useful References

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Kott, P. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 133–142.

Kott, P. (2017). A partially successful attempt to integrate a web-recruited cohort into an address-based sample. Presented at NISS/WSS Workshop on Inference from Nonprobability Samples, Washington DC. (available online).

Kott, P. (2018) A design-sensitive approach to fitting regression models with complex survey data, *Statistics Surveys*, 12, 1-17.

RTI International (2012). *SUDAAN Language Manual, Release 11.0.* Research Triangle Park, NC: RTI International.

Singh, A., Dever, J., and Iannacchione, V. (2004). Efficient estimation for surveys with nonresponse follow-up using dual-frame calibration. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 3919–3930.

Tille, Y. and Matei, A., (2013). *Package* '*Sampling.*' A software routine available online at http://cran.r-project.org/web/packages/sampling/sampling.pdf (procedure: *gencalib*).