Poststratification When Population Counts Are Unavailable

Anja Zgodic
Jan M Eberth
Alexander C McLain
University of South Carolina

Background

- Small Area Estimation (SAE) produces estimates for small geographical areas, e.g. a county.

- Within each small area, data is sampled unequally and sociodemographic groups are represented unequally.

- Poststratification (PS) is used to adjust SAE estimates for unequal representation of sociodemographic groups.

- PS is often a weighted average of estimates of all sociodemographic groups.

Problem

- To compute the weighted average for PS, we need population counts for all sociodemographic groups. **These counts are not always available**.

Solution

- For strata of sociodemographic groups with missing counts, one can obtain small area prevalence by using simplifying assumptions and secondary data sources.

- Multiplying this prevalence with available sociodemographic group counts enables the computation of the weighted average of PS.

Dataset

- 2016 National Survey of Children's Health (NSCH)

For small area prevalence estimates, **poststratification** can be conducted **despite of missing population counts**.

With **simplifying assumptions** and **secondary data sources**, estimate the **prevalence** of missing sociodemographic group counts.

Update total population counts using **prevalence estimates of missing counts.**

Poststratification is then **possible.**

Take a picture
for more information

UNIVERSITY OF South Carolina

Working Example

- We fit a SAE model for county-level rates of child overweight/obesity:
$logit(p_{rgaek}) = \boldsymbol{X}_{rgae}\boldsymbol{\beta}_{rgae} + \boldsymbol{X}_k\boldsymbol{\beta}_k + b_k$
b_k is a random intercept for each county k conditional on neighboring counties (intrinsic conditional autoregressive).

- This model provides estimates of overweight/obesity rates for each county k, for 256 strata of child-level variables: $\hat{p}_{rgaek}$
  - 4 race/ethnicity groups, r
  - 2 gender groups, g
  - 8 age groups, a
  - 4 parental education groups, e

- Census population counts are only available for 64 race/ethnicity, gender, and age strata N_{rgak}, but we need N_{rgaek}.

- Assuming child gender and age have no influence on parental education $P_k(e|rga) = P_k(e|r)$, we need the prevalence $P_k(e|r)$, to multiply it with N_{rgak}.

- $P_k(rgae) = P_k(e|r) \times P_k(rga)$, where $P_k(rga)$ is available in secondary data sources (Census Bureau for this example).

- $P_{S(k)}(e|r)$ is available from the NSCH.

- Assuming the ratio of state-to-county prevalence of being in a parental education group is the same for all race/ethnicity groups, we can get $P_k(e|r)$ by multiplying $Ratio_{S(k)} \times P_{S(k)}(e|r)$

- $Ratio_{S(k)}$ is available from the Census

- $P_k(rgae)$ is now estimated, $P_k(rgae) = P_k(e|r) \times P_k(rga)$.

- We get the last missing piece: $N_{rgaek} = P_k(rgae) \times N_{rgak}$

- We can finally conduct PS:
$\hat{p}_k = \frac{\sum_r \sum_g \sum_a \sum_e \hat{p}_{rgaek} N_{rgaek}}{\sum_r \sum_g \sum_a \sum_e N_{rgaek}}$

- We have an estimate of obesity/overweight for each county, $\hat{p}_k$. This estimate reflects the underlying distribution of the county population.