Error Analysis

David J Dorer Dorer Community Service Foundation

An important part of statistical model building and estimation is determining the precision of the output estimates or cell values in the output tables.

One method for computing errors is called "replicate weights." Instead of a single weight in the synthetic data set a series of different weights are produced. See the R `survey` package for details about weights and replicate weights. The replicate weights allow you to repeat the analysis separately for each replicate weight. The resulting multiple estimates, one for each replicate weight can then combined to produce and estimate and margin of error or "MoE" just like downloaded ACS tables. There are 2 functions for this process:

```
PAT.repweights(<data.frame>)
PAT.reptable(<formula>,<replicate weight object>)
```

`PAT.reptable()` produces a multiway tabulation with an estimate "Est" and margin of error "MoE" component/dimension.

The R function PAT.repweights uses the `survey` package function as.svrepdesign to create the replicate weights: n

```
PAT.repweights<-function(dat,seed=1000,rep=100) {
    require(survey)
    design<-svydesign(id=~1,data=dat,weights=dat$weights);
    if(!is.na(seed)) set.seed(seed);
    as.svrepdesign(design=design,type="mrbbootstrap",replicates=rep);
};
```

The `bootstrap` method is used to create the replicate weights.

Internal and External Validity

Any statistical estimates of population sizes or associated rates should be checked for internal and external validity. For example the function `PAT.test.model()` cross checks the small area estimate of a synthetic data cross tabulation at the PUMA geography with the corresponding detail table at the same PUMA geography. This is an example of an "internal validity" check. An example of an external validity check would be to compare a statistic for a geography that uses data sources other than ACS data. For example a statistic that is published by a State or Federal agency such as a school district. Many schools publish statistics for their school age population. You can compute the same or similar statistic using the PAT package and compare the results.

References

Validation of Multilevel Regression and Poststratification Methodology for Small Area Estimation of Health Indicators From the Behavioral Risk Factor Surveillance System https://academic.oup.com/aje/article/182/2/127/93984

2017-2021 Accuracy of PUMS https://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2017_2021Accuracy

American Community Survey Design and Methodology Report https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/2022/acs_design_methodology_report_2022.pdf

v1.0 26 Dec 2023 13:23