RRFSS Analysis Training Sessions 2023 Agenda

Session #1 TUESDAY, NOVEMBER 14th, 1:00pm - 3:00pm

1. Overview of using ISR-provided weights (when each weight should be applied)

2. Overview of methods and formula used to produce the ISR-provided weights.

i) Checking weights introduction

ii) Dual-frame landline and cell phone weights

3. Replicating the ISR-provided weights.

RRFSS Analysis Training Sessions 2023 Agenda

Session #2 TUESDAY, NOVEMBER 21st, 10:00am - 12:00pm

1. Overview of methods and formula for weights if combining cycles

i) Deriving sample weights for combining cycles

2. Overview of methods and formula for weights in clustered analyses

i) Deriving sample weights for within/sub-PHU geographies

RRFSS Analysis Training Sessions 2023 Agenda

Session #3 TUESDAY, NOVEMBER 28th, 10:00am - 12:00pm

1. Age-standardization

i) Overview and discussion

2. Post-stratification

i) Overview and discussion – Deriving post-stratification weights by age-group and sex

ii) Additional considerations

RRFSS Analysis Training Sessions 2023
Agenda

Session #3 TUESDAY, NOVEMBER 28th, 10:00am - 12:00pm Continued

3. Other survey data analysis topics

i) Calculating weights with missing data – how to handle
ii) Checking weights overview
iii) Cell phone sampling frame additional considerations
iv) Finite Population Correction
v) Resampling
vi) Alternative variance estimation methods/options
vii) Regression analysis with survey data
viii) When weights are/are not needed
ix) CSAplan for SPSS users

4. Additional questions/topics

a. Please submit questions ahead of time. Please submit by e-mail to Hugh and Liza.

1. Age-standardization

i) Overview and discussion

- a procedure applied in epidemiology to adjust rates (e.g., mortality, answers to a survey question) to a given age distribution (e.g., for a population based on Census data) compared to the age distribution in the original sample or population
- this adjustment involves a weighted average, so weights to population are in effect used
- the Stata command dstdize can be employed to calculate direct “standardized rates, a weighted average of the stratum-specific rates”

2. Post-stratification

i) Overview and discussion – Deriving post-stratification weights by age-group and sex

- can help obtain more accurate and representative results

- for some surveys, age group and gender/sex quotas are set based on current Census information as part the sampling scheme – if those quotas are closely achieved, post stratification may not be needed

- age and gender/sex can also be used as control variables in statistical analyses such as regression

2. Post-stratification

An Example from the Halton PHU of Post-stratification Weights to Population by Age Group and Gender

Age Group	Male (1)	Female (5)	Total
18-34 (1)	66264	63099	129363
35-64 (2)	124197	130546	254743
65+ (3)	43978	55376	99354
Total	234439	249021	483460

RRFSS 2023 Jan. to Apr.
An Example from the Halton PHU of
Post-stratification by Age Group and Gender
Weights to Population

Tot_popn_STATSCAN_Halton_c43=66264 if age34_ISR_new==1 & RGENDER==1

Tot_popn_STATSCAN_Halton_c43=63099 if age34_ISR_new==1 & RGENDER==5

Tot_popn_STATSCAN_Halton_c43=124197 if age34_ISR_new==2 & RGENDER==1

Tot_popn_STATSCAN_Halton_c43=130546 if age34_ISR_new==2 & RGENDER==5

Tot_popn_STATSCAN_Halton_c43=43978 if age34_ISR_new==3 & RGENDER==1

Tot_popn_STATSCAN_Halton_c43=55376 if age34_ISR_new==3 & RGENDER==5

RRFSS 2023 Jan. to Apr.
An Example from the Halton PHU of
Post-stratification by Age Group and Gender
Weights to Population

WeightBYagegroup_gender_and_Popn =
Tot_popn_STATSCAN_Halton_c43 / tot_respondents_in_agegroup_gend

Subpopulation size

subsample size

WeightBYagegroup_gender_and_Popn
"Weight by each age group (age34_ISR_new) and gender in Halton for Cycle 43“

2. Post-stratification

ii) Additional considerations

- in addition to age group and gender,
income and education
are often used in large population surveys
- can be helpful, but may not entirely correct
misrepresentative samples

3. Other survey data analysis topics

i) Calculating weights with missing data – how to handle

- accept some weights are missing

or

- before computing the weights:

impute the missing values of variables needed to compute the weights

or

- after computing the weights:

impute values of the missing weights

3. Other survey data analysis topics

ii) Checking weights overview

- weight means
- weight sums
- weight range, outliers: are there extremely large or small weights to consider for “trimming”
- weight variability
- weight percentiles: “unexpected shifts in distribution”

3. Other survey data analysis topics

ii) Checking weights overview

- Record count
- Coefficient of variation (cv)
- Design effects
- Standard errors: with survey weight compared to replicate weights (e.g., bootstrap replicate weights) if the latter are in use
- Nonresponse bias
- Other related surveys to compare

3. Other survey data analysis topics

ii) Checking weights overview

- compare common estimates (e.g., totals, means, relative frequencies (%), proportions, standard errors) and analyses without and with weights applied

- do the same for other statistical analyses commonly done (e.g., regression)

3. Other survey data analysis topics

iii) Cell phone sampling frame additional considerations

- acceptable maximum values for number of cell phones for respondent and household
- same for the number of landlines for respondent and household
- incorporation of any relevant population figures on cell phone and landline use

3. Other survey data analysis topics

iv) Finite Population Correction

N: Population Size

n: Sample Size

$$\sqrt{\frac{N-n}{N-1}}$$

- sampling without replacement strictly speaking changes the probabilities of successive selections
- can multiply an estimate's standard deviation by this factor for increased accuracy of the standard deviation

3. Other survey data analysis topics

iv) Finite Population Correction

N: Population Size

n: Sample Size

$$\sqrt{\frac{N-n}{N-1}}$$

- commonly not considered needed when *n/N* is less that 0.1 (10%)
- RRFSS 2023 Jan.-Apr. Halton example:

  338/483460 = 0.0006991271 …

  √[(483460-338)/(483460-1)] = 0.9996514 …

3. Other survey data analysis topics

v) Resampling

- common examples are the jackknife and the bootstrap
- jackknife: take one case out successively to create n replicate sample – there are variations
- bootstrap: resample the sample with replacement to create m replicate samples of size n; m is commonly 500 or 1,000 – there are variations
- standard errors and confidence intervals (e.g., 95%) of estimates (e.g., totals, means, proportions, relative frequencies (%)) can be calculated from the range of values of that estimate over all m replicate sample

3. Other survey data analysis topics

vi) Alternative variance estimation methods/options

- the jackknife and bootstrap procedures can be applied widely to various data and statistical operations
- the calculation of variance and standard errors is one example
- in Stata the svyset statement can include various variance estimation specifications such as:

vce(linearized) – the default option
vce(jackknife)
vce(bootstrap)

- for RRFSS, generally vce(linearized) can be used

3. Other survey data analysis topics

vii) Regression analysis with survey data

- there are statistical diagnostics tests for whether or not weights need to be applied when performing a regression
- Bollen et al. article (2016) is a helpful overview
- the Valliant and Dever book (2018) gives some worked examples

3. Other survey data analysis topics

viii) When weights are/are not needed

- generally helpful for the increased accuracy (reduced bias) of population estimates of one variable
- there are various exceptions such as when the weights are not correlated with the variable of interest leading to little difference in the estimate without and with weights (the latter may lead to a larger standard error, reduced precision, “more noise”)
- statistical analyses with multiple variables may or may not require the use of weights

3. Other survey data analysis topics

ix) CSAplan for SPSS users

```
CSPLAN ANALYSIS
  /PLAN FILE="H:\Projects1\RRFSS 2022\RRFS Sep Data\Hugh's
work\RRFSS_HHLD_and_Popn_weights.csaplan"
  /PLANVARS ANALYSISWEIGHT=WeightBYNadults_and_PopnPerYear
  /SRSESTIMATOR TYPE=WOR
  /PRINT PLAN
  /DESIGN STRATA=CS_H_Unit CLUSTER=CS_HHLD_ID
  /ESTIMATOR TYPE=WR.
```

4. Additional questions/topics