



NDWS Weighting Tutorial

with SAS, Stata, and R code demonstrations

Why use weights?

Without weights

- Unweighted counts describe respondents
- Every respondent counts the same
- Good for sample-size checks
- Not a population estimate

With weights + survey design

- Weighted estimates describe the target population
- Weights adjust for unequal selection and response patterns
- Accounts for stratification and clustering
- Produces design-based SEs and CIs

Specify the survey design to get valid standard errors, confidence intervals, & inferences.

Design declaration: three ingredients

Weight

FinalWeight

How much each respondent represents

Strata

Stratum

Groups used in sample design

Cluster / PSU

**pSampleCluster /
NDWS_FACID /
1**

Primary sampling unit; variable may differ by file

- Use the variable names provided in your NDWS files and user guides.
- The weight is not a “quality score” -- each case contributes a different amount to the estimate

SAS demonstration: PROC SURVEYMEANS

Example: AL_Staff_RUF

```
PROC SURVEYMEANS DATA=AL_Staff_RUF;  
  STRATA stratum;  
  CLUSTER NDWS_FACID;  
  /* or pSampleCluster if you use AL or NH PUFs */  
  WEIGHT FinalWeight;  
  VAR JobOverall;  
RUN;
```

- What this does:
 - Estimates the weighted mean of JobOverall
 - Uses strata and cluster information for SEs
 - Reports N, mean, SE, and CI
- Same pattern for weighted means and many other survey procedures.

Data Summary	
Number of Strata	13
Number of Clusters	120
Number of Observations	1559
Sum of Weights	710161.386

Statistics						
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
JobOverall	Level of satisfaction: Overall job	1559	3.542414	0.030043	3.48285851	3.60197013

Stata demonstration: *svyset*, then *svy*:

Example: CC_RUF

```
svyset [pweight = FinalWeight],  
strata(stratum)
```

```
svy: mean PatientPanel
```

Survey: Mean estimation

```
Number of strata =    18      Number of obs   =    6,554  
Number of PSUs   = 6,554      Population size = 480,537.93  
Design df        =          6,536
```

	Linearized			
	Mean	std. err.	[95% conf. interval]	
PatientPanel	947.66	16.11493	916.0695	979.2505

- What this does:
 - Estimates the weighted mean of PatientPanel
 - Uses strata and cluster information for SEs
 - Reports N, mean, SE, and CI
- The sample frame of CC is a list of all eligible clinicians, so no cluster/PSU variable is needed
- After *svyset*, use *svy*: for population estimates and design-based standard errors

Without strata, the point estimates are the same, but the standard errors differ.

R demonstration: survey package

Example: NH_Staff_PUF

```
library(survey)

nh_des <- svydesign(
  ids      = ~pSampleCluster,
  # or ~NDWS_FACID for NH / AL RUFs
  # or ~1 for CC surveys
  strata   = ~Stratum,
  weights  = ~FinalWeight,
  data     = nh_staff_puf,
  nest     = TRUE)

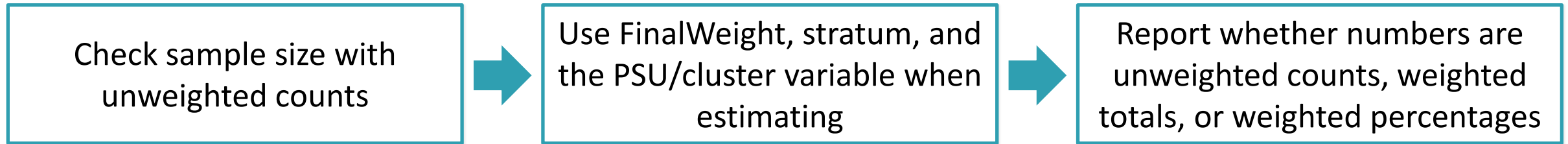
joboverall_mean <- svymean(
  ~JobOverall,
  nh_des,
  na.rm = TRUE)

joboverall_mean
confint(joboverall_mean)
```

- R translation:
 - svydesign() = design declaration
 - svymean() = weighted mean + SE
 - confint() = 95% confidence interval
- svymean() output displays only the weighted mean and its standard error by default; confint() is requested to show the CI

	mean	SE
JobOverall	3.4557	0.0227
	2.5 %	97.5 %
JobOverall	3.411195	3.500125

Takeaway



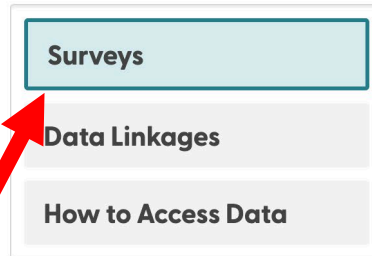
- Unweighted count shows “How many **respondents** are in the data?”
- Weighted estimate helps answer “What does this represent in the **population**?”
- Declare the design first; then use survey-aware commands
- Check out the variable names; Stata and R are case-sensitive

Differences in each language

	SAS	Stata	R
Survey weight	weight FinalWeight;	[pweight = FinalWeight]	weights = ~FinalWeight
Stratum	strata stratum;	strata(stratum)	strata = ~stratum
Cluster / PSU	cluster NDWS_FACID;	svyset NDWS_FACID	ids = ~NDWS_FACID
No cluster / PSU	Omit it	Omit it	ids = ~1
Cluster IDs nested within strata	Handled through the combined design specification	Handled through svyset	nest = TRUE
95% confidence interval	Reported by PROC SURVEYMEANS	Reported by svy: mean	confint() must be requested

- If no strata (e.g., NDWS CC Public Use Files (PUFs)) -> skip it
- In NH & AL PUFs, use `pSampleCluster` for cluster/PSU

Learn more on the NDWS website ndws.org



NDWS library



Community Clinician

- CC Wave 1, Version 2 (PDF, 422 KB) ↗
- CC Wave 1, Short Form (PDF, 378 KB) ↗

Wave 1 Final User Guide

- CC User Guide (PDF, 601 KB) ↗

While new organizations were recruited during Wave 2 for the Nursing Home, Assisted Living, and Home Care Surveys, the Community Clinician Survey continued recruitment from the original Wave 1 sample. The version 2 instrument posted above has three additional questions (items 70, 72, 73) added during the recruitment phase; otherwise it is identical to the original Wave 1 instrument. In addition, select respondents received a short-form version of the instrument, which is available here.



Nursing Home

Wave 2 Questionnaires

- Nursing Home Staff (PDF, 393 KB) ↗
- NH Staff Follow-Up (PDF, 441 KB) ↗
- NH Administrator (PDF, 360 KB) ↗

Wave 2 Sample Frame

- Overview (PDF, 173 KB) ↗
- SAS Files (ZIP, 21 KB) ↗

Wave 2 User Guide

- NH/AL User Guide (PDF, 483 KB) ↗



Assisted Living

Wave 2 Questionnaires

- Assisted Living Staff (PDF, 390 KB) ↗
- AL Staff Follow-Up (PDF, 392 KB) ↗
- AL Administrator (PDF, 363 KB) ↗

Acknowledgement

Funding provided by the National
Institute on Aging (U54AG084520)



*JAGS Special
Collection*

Thank you!

www.ndws.org | info@ndws.org