MAP ACADEMY

Methodology, Analytics & Psychometrics

Analyzing Data from Complex Sampling Designs: An Overview and Illustration

Natalie Koziol, MA, MS

Methodologist, MAP Academy

Presented on 12/12/14

Outline

- Probability (random) sampling
- Sampling strategies
- Inferential frameworks
- Data analysis considerations
- Data analysis example

PROBABILITY SAMPLING

Requirements

1) The set of all possible samples, given the sampling strategy, can be defined
2) Each possible sample has a known probability of being selected, $P(S = s)$
3) Each population unit has a nonzero probability of being selected, $\pi_i > 0$
   - π_i is the "inclusion probability" of unit i
   - $\pi_i = \sum_{i \in s} P(S = s)$
4) A random mechanism is used to select a sample with probability $P(S = s)$

Example

- Target population
  - Family pets
- Sampling frame
  - List of all units in the target population

    - ❑ Hugo
    - ❑ Nala
    - ❑ Smokey
    - ❑ Pepper
- Sampling strategy
  - Obtain simple random sample of size $n = 2$

Example, cont'd

- Define set of all possible samples and determine sample selection probabilities

s	Sample Units	$P(S = s)$
s_1	Hugo, Nala	$1/6$
s_2	Hugo, Smokey	$1/6$
s_3	Hugo, Pepper	$1/6$
s_4	Nala, Smokey	$1/6$
s_5	Nala, Pepper	$1/6$
s_6	Smokey, Pepper	$1/6$

Example, cont'd

- Calculate inclusion probabilities

Population Unit	π_i
Hugo	$\mathrm{P}(S = s_1) + \mathrm{P}(S = s_2) + \mathrm{P}(S = s_3) = 1/2$
Nala	$\mathrm{P}(S = s_1) + \mathrm{P}(S = s_4) + \mathrm{P}(S = s_5) = 1/2$
Smokey	$\mathrm{P}(S = s_2) + \mathrm{P}(S = s_4) + \mathrm{P}(S = s_6) = 1/2$
Pepper	$\mathrm{P}(S = s_3) + \mathrm{P}(S = s_5) + \mathrm{P}(S = s_6) = 1/2$

Example, cont'd

- Use random mechanism to select sample
  - e.g., Use the SURVEYSELECT procedure in SAS

```
DATA PopPets;
    INPUT name $;
    DATALINES;
Hugo
Nala
Smokey
Pepper
;
PROC SURVEYSELECT DATA=PopPets METHOD=SRS n=2 SEED=21701
                  OUT=SamplePets; RUN;
PROC PRINT DATA=SamplePets; RUN;
```

Obs	name
1	Hugo
2	Nala

Non-Probability Sampling

- Convenience sampling, purposive sampling
- Generally cheaper and less complex than probability sampling
- May be the only option
  - e.g., When studying hidden or hard-to reach populations
- More susceptible to selection bias than probability sampling!
  - Selection bias results from the sampled population not matching the target population
  - Threatens the external validity (generalizability) of inferences

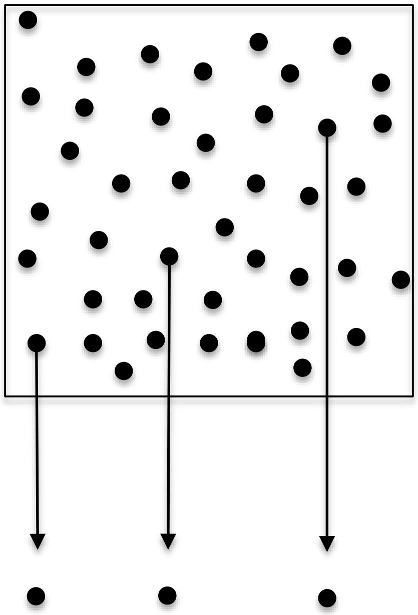
SAMPLING STRATEGIES

Random Sampling Strategies

- Element sampling
- Stratified sampling
- Cluster sampling

Element Sampling

- Basis for all other sampling strategies
- Sampling unit = observation unit
- Types
  - Simple random sampling (SRS)
  - Bernoulli sampling
  - Poisson sampling
  - Systematic sampling



Element Sampling: SRS

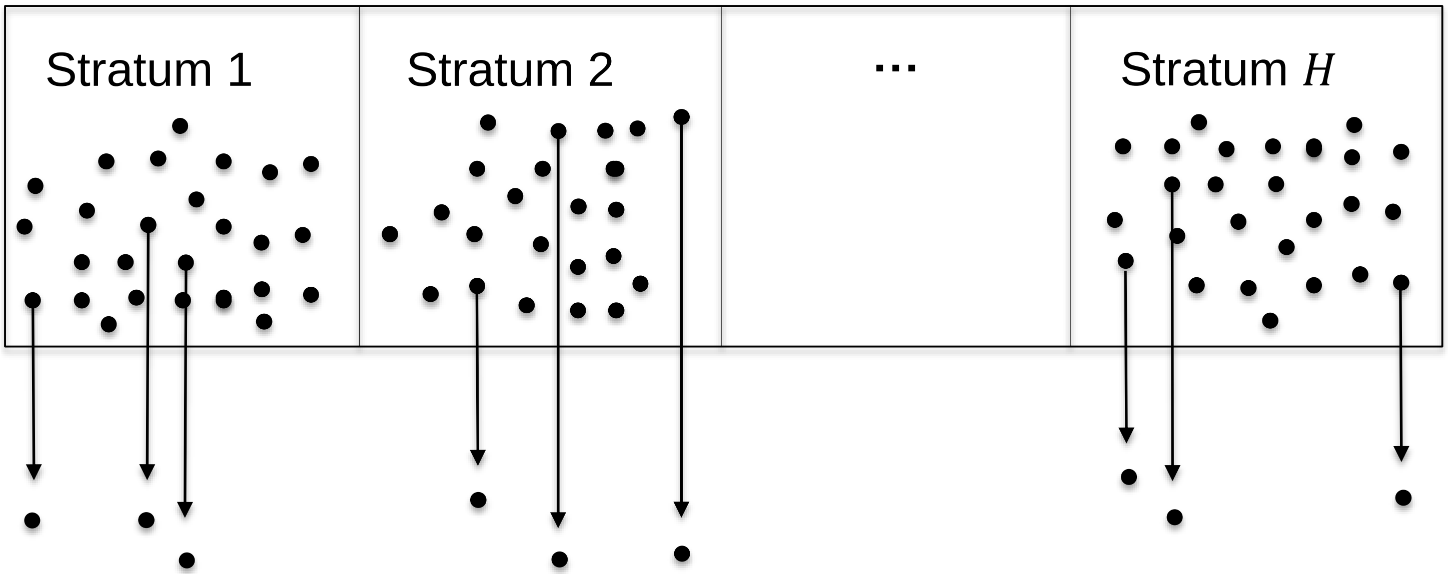
- Randomly select n units from a population of N units
- With replacement (SRSWR)
  - Sampled unit placed back in population after each draw
  - Units can be sampled more than once
  - Also referred to as unrestricted sampling (URS)
- Without replacement (SRSWOR)
  - Sampled unit NOT placed back in population after draw
  - Units CANNOT be sampled more than once
  - Also referred to simply as simple random sampling
- $\pi_i = n/N$

Element Sampling: Other Types

- Bernoulli sampling
  - Similar to SRSWOR but n is a random variable
  - Specify constant inclusion probability ($\pi_i = \pi$)
  - Select each unit with probability π
- Poisson sampling
  - Similar to Bernoulli sampling but unequal inclusion probabilities
- Systematic random sampling
  - Randomly select starting point from sampling frame and then sample at fixed interval of N/n
  - Special type of clustering but often acts like SRS

Stratified Sampling

- Divide the population into H strata
- Perform element sampling independently within each stratum



Stratified Sampling

- Equal allocation
  - n_h is constant across all h
- Proportional allocation
  - n_h is proportional to N_h
- Optimal allocation
  - Greater proportion of units selected from strata that are large, heterogeneous, and inexpensive to sample
  - Neyman allocation is a special case

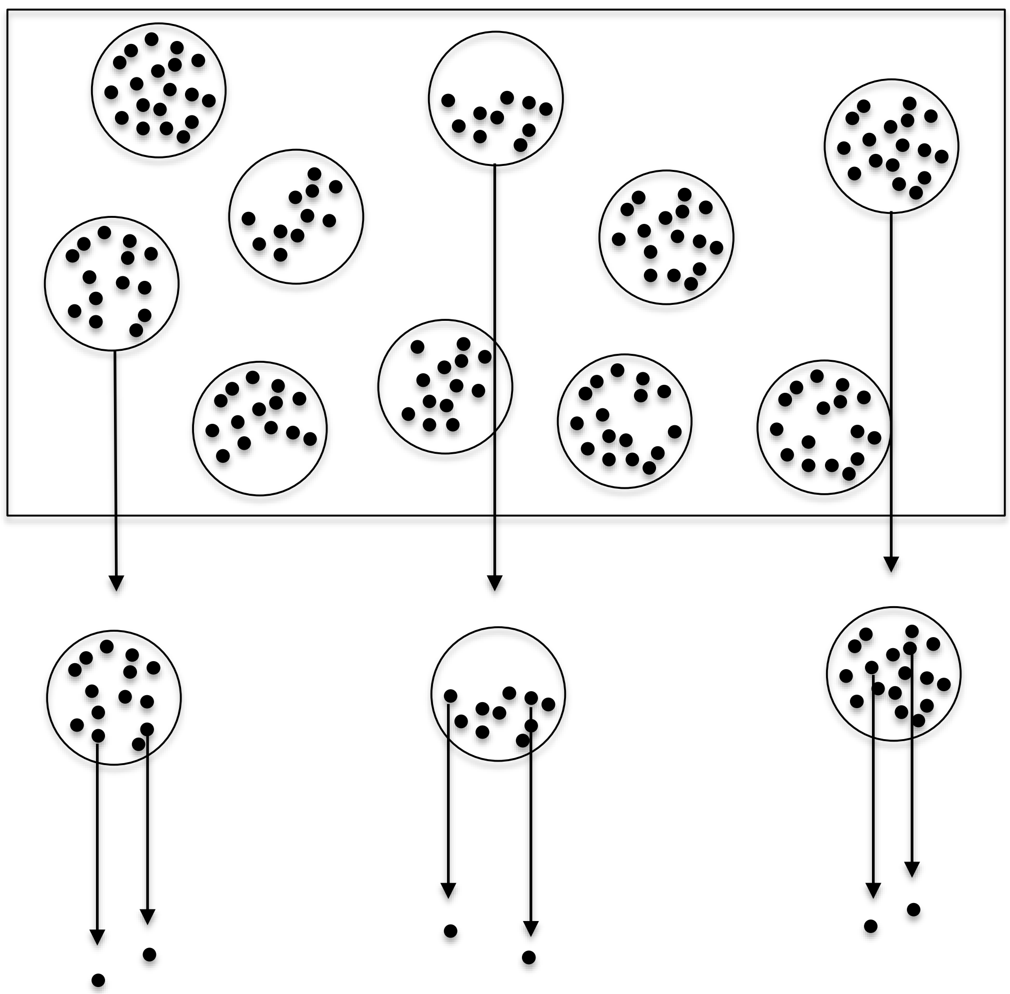
Stratified Sampling

- More control over sample representativeness
  - Less chance of obtaining a “bad” sample
- Potentially more efficient method of sampling
  - Allows variation in sampling frame, design, and field procedures across strata
- Enables domain (subpopulation) analysis
- Greater precision (smaller standard errors)

Cluster Sampling

- Primary sampling unit ≠ observation unit
- One-stage clustering
  - Use element sampling strategy to sample clusters of units
    - Clusters = primary sampling units (PSUs)
  - Observe all units within each sampled PSU
- Two-stage clustering
  - Stage 1: Use element sampling strategy to sample PSUs
  - Stage 2: Use element sampling to sample individual units within the sampled PSUs
    - Individual units = second-stage units (SSUs)

Cluster Sampling



Cluster Sampling

- Methods for sampling PSUs
  - Equal probability sampling methods
  - Probability proportional to size (PPS) methods
    - Inclusion probability of PSU is proportional to a measure of the PSU's size
    - Several different PPS methods (e.g., WR, WOR, systematic, Brewer, Murthy, Sampford)

Cluster Sampling

- Disadvantages
  - Less precision (larger standard errors)
- Advantages
  - May be the only option
  - May be the more time and cost efficient option
  - Permits multilevel inferences

INFERENTIAL FRAMEWORKS

Inferential Frameworks

- Goal of sampling is to make inferences about the population
- Need formal statistical framework to link sample to population
  - Design-based framework (randomization theory)
  - Model-based framework
  - Hybrid framework

Design-Based Framework

- Requires probability sampling
  - Inclusion indicators (Z_i's) are random variables

$$Z_i = \begin{cases} 1 \text{ if unit } i \text{ is in the sample} \\ 0 \text{ otherwise} \end{cases}$$

  - Measured outcomes (Y_i's) are assumed to be fixed quantities
- Design-based estimators
  - Use of design weights
  - Standard errors derived from the design
- Permits descriptive inferences about finite population parameters
  - Parameters are generally simple functions (e.g., mean, total) of the Y_i's

Model-Based Framework

- Does not require probability sampling
  - Y_i's are random variables
    - Specify a hypothetical probability model for Y_i
  - If probability sampling is used, then Z_i's are also random variables
- Model-based estimators
  - Design features specified as part of the model (e.g., use multilevel modeling, truncated regression)
  - Standard errors derived from the model
- Permits predictive inferences about infinite (super-) population parameters
  - Parameters are the parameters of the model (e.g., regression coefficients)

Contrasting Weaknesses

- Weaknesses of design-based framework
  - Doesn’t lend itself to answering the types of questions relevant to social science research
    - Limited to simple univariate/bivariate investigations
    - Limited to description
- Weaknesses of model-based framework
  - Inferences susceptible to model misspecification
  - Cumbersome reliance on model specification to account for sample design features
    - Results in highly parameterized models (blurs interpretation, reduces statistical power)
    - Complete and appropriate specification is difficult

Hybrid Framework

- Combines the traditional frameworks
  - Relies on model specification and design-adjusted estimation
  - Assuming probability sampling, provides descriptive inferences about finite population parameters
  - Assuming correct model specification, provides predictive inferences about infinite population parameters
- Continuum of modeling options
  - Aggregated approaches
    - Rely more heavily on adjusted estimation
    - The focus of this presentation
  - Disaggregated approaches
    - Rely more heavily on model specification

DATA ANALYSIS CONSIDERATIONS

Accounting for the Design

- Need to account for design features in order to obtain valid inferences
- Adjustments
  - Weighting
  - Alternative variance estimators
  - Finite population correction (FPC)
  - Domain analysis
- Requires statistical software that can handle complex sampling designs

Design Weights

- Need to account for unequal inclusion probabilities
  - Weight each sample observation by the inverse of its inclusion probability
    - $w_i = 1/\pi_i$
- Generally do not need to account for equal inclusion probabilities
  - Self-weighting sample
  - Weighting may still be necessary if computing totals or performing multilevel modeling

Example

Stratum	Unit	Height	π_{ih}	w_{ih}
Male	1	72	$1/2$	2
Male	2	70	$1/2$	2
Male	3	74	$1/2$	2
Male	4	72	$1/2$	2
Female	5	64	$1/3$	3
Female	6	66	$1/3$	3
Female	7	62	$1/3$	3
Female	8	63	$1/3$	3
Female	9	64	$1/3$	3
Female	10	65	$1/3$	3

Average height in the population

$= 67.2$ inches

Unweighted sample estimate

$$= \frac{70 + 74 + 63 + 65}{4}$$

$= 68$ inches

Weighted sample estimate

$$= \frac{70 \times 2 + 74 \times 2 + 63 \times 3 + 65 \times 3}{2 + 2 + 3 + 3}$$

$= 67.2$ inches

Complexities of Weighting

- Weight adjustments
  - Complex adjustments may be made to design weights to account for nonresponse
  - $\widetilde{w}_i = 1/(\pi_i \hat{\varphi}_i)$ where $\hat{\varphi}_i$ is the estimated probability that unit i responds
- Multiple weight options
  - Secondary datasets often include multiple weight options
  - Appropriate weight depends on several factors
    - Type of analysis (e.g., longitudinal vs. cross-sectional)
    - Unit of analysis (e.g., child vs. school)
    - Respondent (e.g., parent-report, direct observation of child)

Alternative Variance Estimators

- Need to adjust standard errors (SEs) to account for the design
- Assumption of independent and identically distributed random variables is untenable outside of SRS
- SEs will tend to be overestimated in the presence of stratification
- SEs will tend to be underestimated in the presence of clustering

Alternative Variance Estimators

- Closed-form (theoretical) solutions for SEs only available for very simple analyses
- Use an approximation method
  - Taylor series (linearization) methods
  - Random group methods
  - Resampling and replication methods
    - Balanced repeated replication (BRR)
    - Jackknife
    - Bootstrap
  - Generalized variance functions

Finite Population Correction

- Downward adjustment made to SEs when sampling without replacement
  - Increase in sampling fraction results in decrease in sampling variability
- fpc $= (1 - f)$
  - f is the sampling fraction of the PSUs
  - For SRSWOR, $f = n/N$
- Only available when using Taylor series variance estimation method
- Typically ignored in practice when $f < .05$

Domain Analysis

- Researchers are often interested in particular subgroups of the population
- SEs and inferential tests will generally be incorrect if analyses are performed separately by subgroups
- A more appropriate approach is to conduct a domain (subpopulation) analysis
  - Zero-weight approach
  - Multiple-group approach

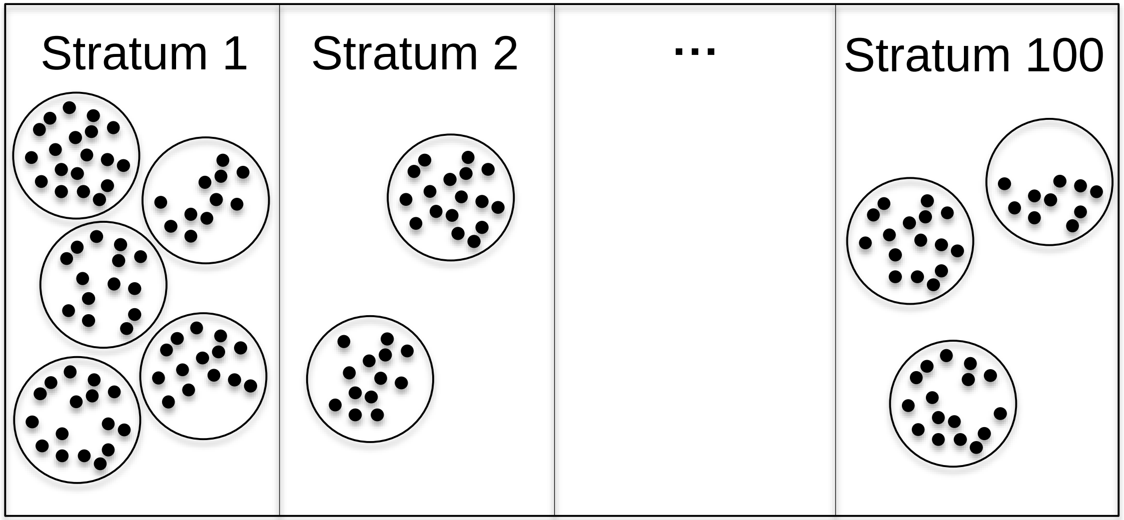
Statistical Software Options

- AM statistical software
- Data Analysis System (DAS) (for NCES data)
- M*plus*
- PowerStats (for NCES data)
- R package "survey"
- SAS survey procedures
- SPSS complex samples module
- Stata
- SUDAAN

DATA ANALYSIS EXAMPLE

Simulated Population

- 1,000 PSUs nested within 100 strata
  - 2 to 18 PSUs nested within each stratum
- 24,587 total SSUs nested within the PSUs
  - 10 to 40 SSUs nested within each PSU



Sampling Design

- First stage
  - Sampled 2 PSUs without replacement from each stratum with probability proportional to size (PPS)
  - 200 total PSUs sampled
- Second stage
  - Sampled 5 SSUs from each PSU using SRSWOR
  - 1,000 total SSUs sampled

Sample Data File

- First 10 cases*

Obs	stratum	psu	ssu	x1	x2	x3	x4	y1	y2	wt
1	1	367	8881	1	0.62625	1	1.09675	0	-0.47781	40.3066
2	1	367	8887	1	0.62625	1	1.35400	1	1.30236	40.3066
3	1	367	8894	1	0.62625	0	-0.66075	0	-1.59707	40.3066
4	1	367	8900	1	0.62625	1	1.54148	1	2.37954	40.3066
5	1	367	8903	1	0.62625	1	2.27028	1	1.68341	40.3066
6	1	479	11627	1	-0.00300	0	0.60575	1	1.20603	41.0316
7	1	479	11649	1	-0.00300	1	1.50519	1	2.05838	41.0316
8	1	479	11652	1	-0.00300	1	-0.29731	1	0.02299	41.0316
9	1	479	11654	1	-0.00300	1	-0.08718	0	-0.73679	41.0316
10	1	479	11658	1	-0.00300	0	0.03111	1	0.84361	41.0316

*BRR & Jackknife replicate weights (BRRrep1-BRRrep104, JKrep1-JKrep200) not shown

Analysis 1

- Examine descriptive statistics for y_1 and y_2
- Use Jackknife method for variance estimation

Analysis 1: M*plus*

```
TITLE: Analysis 1;
DATA: FILE = sample.csv;
VARIABLE: NAMES ARE stratum psu ssu x1 x2 x3 x4 y1 y2 wt
                    BRRrep1-BRRrep104
                    JKrep1-JKrep200;
          USEVARIABLES ARE y1 y2;
          CATEGORICAL ARE y1;
          WEIGHT = wt;
          REPWEIGHTS = JKrep1-JKrep200;
ANALYSIS: TYPE = COMPLEX;
          ESTIMATOR = ML;
          REPSE=JACKKNIFE2;
MODEL:
OUTPUT: SAMPSTAT;
```

```
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES


  Y1
    Category 1      0.505       505.042
    Category 2      0.495       494.958
```

```
                                                    Two-Tailed
                 Estimate       S.E.  Est./S.E.    P-Value


Means
   Y2              -0.011      0.088     -0.127      0.899
```

Analysis 1: R

```
library(survey)
JKrepwts <- sample[,115:314]
design.1 <- svrepdesign(data = sample, type = "JKn", repweights = JKrepwts,
                        weights = ~wt, scale = 1,
                        rscales = rep(1, ncol(repweights)))
svymean(~as.factor(y1), design = design.1)
svytotal(~as.factor(y1), design = design.1)
svymean(~y2, design = design.1)
```

```
                    mean     SE
as.factor(y1)0 0.50504 0.0274
as.factor(y1)1 0.49496 0.0274
```

```
                 total     SE
as.factor(y1)0 12426 671.29
as.factor(y1)1 12178 677.90
```

```
         mean     SE
y2 -0.011101 0.0876
```

Analysis 1: SAS

```
PROC SURVEYFREQ DATA=sample VARMETHOD=JK;
TABLE y1;
WEIGHT wt;
REPWEIGHTS JKrep1-JKrep200 / JKCOEF = 1; RUN;

PROC SURVEYMEANS DATA=sample VARMETHOD=JK;
VAR y2;
WEIGHT wt;
REPWEIGHTS JKrep1-JKrep200 / JKCOEF = 1; RUN;
```

Table of y1

y1	Frequency	Weighted Frequency	Std Dev of Wgt Freq	Percent	Std Err of Percent
0	498	12426	671.28751	50.5042	2.7390
1	502	12178	677.90276	49.4958	2.7390
Total	1000	24604	62.22160	100.000	

Statistics

Variable	N	Mean	Std Error of Mean	95% CL for Mean	
y2	1000	-0.011101	0.087576	-0.1837921	0.16159069

Analysis 2

- Estimate a logistic regression model to determine the effect of x_4 on y_1
- Use Taylor series method for variance estimation
- Perform domain analysis for subpopulation $x_1 = 1$

Analysis 2: M*plus*

```
TITLE: Analysis 2;
DATA: FILE = sample.csv;
VARIABLE: NAMES ARE stratum psu ssu x1 x2 x3 x4 y1 y2 wt
                    BRRrep1-BRRrep104
                    JKrep1-JKrep200;
          USEVARIABLES ARE y1 X4;
          CATEGORICAL ARE y1;
          WEIGHT = wt;
          CLUSTER = psu;
          STRATIFICATION = stratum;
          SUBPOPULATION = x1 EQ 1;
ANALYSIS: TYPE = COMPLEX;
          ESTIMATOR = MLR;
MODEL: y1 ON x4;
```

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

Y1          ON
   X4                 1.213      0.115     10.520      0.000

Thresholds
   Y1$1              -0.101      0.115     -0.881      0.378
```

Analysis 2: R

```
library(survey)
design.2 <- svydesign(data = sample, id = ~psu, strata = ~stratum,
                      weights = ~wt)
results.2 <- svyglm(y1 ~ x4, design = subset(design.2, x1 == 1),
                    family = quasibinomial)
summary(results.2)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1012     0.1149   0.881    0.383
x4            1.2134     0.1153  10.520 1.37e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis 2: SAS

```
PROC SURVEYLOGISTIC DATA=sample VARMETHOD=TAYLOR;
CLUSTER psu;
STRATA stratum;
DOMAIN x1;
MODEL y1 (DESCENDING) = x4;
WEIGHT wt; RUN;
```

Analysis of Maximum Likelihood Estimates

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.1012	0.1149	0.88	0.3805
x4	1.2134	0.1154	10.52	<.0001

NOTE: The degrees of freedom for the t tests is 100.

Analysis 3

- Estimate a multiple linear regression model to determine the effects of x_2 and x_3 on y_2
- Use BRR method for variance estimation

Analysis 3: M*plus*

```
TITLE: Analysis 3;
DATA: FILE = sample.csv;
VARIABLE: NAMES ARE stratum psu ssu x1 x2 x3 x4 y1 y2 wt
                    BRRrep1-BRRrep104
                    JKrep1-JKrep200;
          USEVARIABLES ARE y2 x2 x3;
          WEIGHT = wt;
          REPWEIGHTS = BRRrep1-BRRrep104;
ANALYSIS: TYPE = COMPLEX;
          ESTIMATOR = ML;
          REPSE = BRR;
MODEL: y2 ON x2 x3;
```

```
                                                   Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Y2         ON
   X2                 -0.513      0.053     -9.597      0.000
   X3                 -0.313      0.108     -2.904      0.004

Intercepts
   Y2                  0.096      0.068      1.410      0.159

Residual Variances
   Y2                  2.028      0.099     20.588      0.000
```

Analysis 3: R

```
library(survey)
BRRwts <- sample[,11:114]
design.3 <- svrepdesign(data = sample, type = "BRR", repweights = BRRwts,
                        weights = ~wt)
results.3 <- svyglm(y2 ~ x2 + x3, design = design.3)
summary(results.3)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09632    0.06832   1.410  0.16178
x2          -0.51324    0.05348  -9.596  9.1e-16 ***
x3          -0.31362    0.10807  -2.902  0.00458 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis 3: SAS

```
PROC SURVEYREG DATA=sample VARMETHOD=BRR;
MODEL y2 = x2 x3 / SOLUTION;
WEIGHT wt;
REPWEIGHTS BRRrep1-BRRrep104; RUN;
```

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.0963208	0.06832472	1.41	0.1616
x2	-0.5132398	0.05348241	-9.60	<.0001
x3	-0.3136247	0.10806653	-2.90	0.0045

Note: The degrees of freedom for the t tests is 104.

REFERENCES

References

General sampling references

- Kish, L. (1965). *Survey sampling*. New York, NY: Wiley.
- Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). Boston, MA: Brooks/Cole.
- Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. New York, NY: Springer-Verlag.
- Skinner, C. J., Holt, D., & Smith, T. M. F. (Eds.). (1989). *Analysis of complex surveys*. New York, NY: John Wiley & Sons.
- Wolter, K. M. (2007). *Introduction to variance estimation* (2nd ed.). New York, NY: Springer.

References

References for structural equation modeling of data from complex sampling designs

- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 12*, 411-434.
- Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. Mplus Web Note. Muthén & Muthén.
- Stapleton, L. M. (2006). An assessment of practical solutions for structural equation modeling with complex sample data. *Structural Equation Modeling: A Multidisciplinary Journal, 13*, 28-58.

References

References for weighted multilevel modeling

- Asparouhov, T. (2004). Weighting for unequal probability of selection in multilevel modeling. Mplus Web Notes: No. 8. Muthén & Muthén.
- *Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics – Theory and Methods, 35*, 439-460.
- Asparouhov, T., & Muthén, B. (2006). Multilevel modeling of complex survey data. In *Proceedings of the Joint Statistical Meeting: ASA Section on Survey Research Methods* (pp. 2718-2726).
- Cai, T. (2013). Investigation of ways to handle sampling weights for multilevel model analyses. *Sociological Methodology, 43*, 178-219.
- Carle, A. C. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology, 9*(49).
- Grilli, L., & Pratesi, M. (2004). Weighted estimation in multilevel ordinal and binary models in the presence of informative sampling designs. *Survey Methodology, 30*, 93-103.
- Kovačević, M. S., & Rai, S. N. (2003). A pseudo maximum likelihood approach to multilevel modeling of survey data. *Communications in Statistics – Theory and Methods, 32*, 103-121.
- Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. *Journal of the Royal Statistical Society, Series B, 60*, 23-40.
- *Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society, Series B, 60*, 23-56.
- Stapleton, L. M. (2002). The incorporation of sample weights into multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 475-502.
- Stapleton, L. M. (2012). Evaluation of conditional weight approximations for two-level models. *Communications in Statistics – Simulation and Computation, 41*, 182-204.
- Stapleton, L. M. (2014). Incorporating sampling weights into single- and multilevel analyses. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 363-388). Boca Raton, FL: CRC Press.

References

References for inferential frameworks

- Hansen, M., Madow, W., & Tepping, B. (1983). An evaluation of model-dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association, 78*, 776-793.
- Kalton, G. (2002). Models in the practice of survey sampling (revisited). *Journal of Official Statistics, 18*, 129-154.
- Little, R. J. A. (2014). Survey sampling: Past controversies, current orthodoxy, and future paradigms. In X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott, & J.-L. Wang (Eds.), *Past, present, and future of statistical science* (pp. 413-428). Boca Raton, FL: CRC Press.
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316.
- *Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. *Multivariate Behavioral Research, 44*, 711-740.
- Wu, J.-Y., & Kwok, O.-M. (2012). Using SEM to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 16-35.

References

References for software

- AM statistical software
  - http://am.air.org/ (homepage)
- Data Analysis System (DAS)
  - http://nces.ed.gov/das/ (homepage)
- M*plus*
  - http://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r6_web.pdf (user's guide)
- PowerStats
  - http://nces.ed.gov/datalab/ (homepage)
- R package "survey"
  - http://cran.r-project.org/web/packages/survey/index.html (links to user's guide and vignettes)
  - http://r-survey.r-forge.r-project.org/survey/ (package homepage)
- SAS
  - http://support.sas.com/documentation/cdl/en/statug/67523/PDF/default/statug.pdf (user's guide)
  - http://support.sas.com/documentation/cdl/en/statug/67523/HTML/default/viewer.htm#statug_introsamp_sect001.htm (overview)
- SPSS Complex Samples module
  - http://library.uvm.edu/services/statistics/SPSS22Manuals/IBM%20SPSS%20Complex%20Samples.pdf (user's guide)
- Stata
  - http://www.stata.com/manuals13/u.pdf (user's guide)
- SUDAAN
  - http://www.rti.org/sudaan/ (homepage)
- Comparisons among programs
  - http://www.hcp.med.harvard.edu/statistics/survey-soft/

QUESTIONS? COMMENTS?

For more information, please contact Natalie Koziol at nak371@gmail.com