Survey Weighting
Population Studies Center Short Course

Richard Valliant

University of Michigan

Nov. 12, 2019

1. Basic Steps in Weighting
2. Weight Calibration
3. Nonprobability Sampling

Software & Files You Should Have

- Base R, R packages: PracTools, survey, sampling, doBy
- Access to Stata (optional)
- Files for this course–slides, examples
- Examples.zip—has both R and Stata examples
- Examples are taken from two books:
  Valliant, R., Dever, J.A., & Kreuter, F. (2018). *Practical Tools for Designing and Weighting Survey Samples*, 2nd edition. New York: Springer.

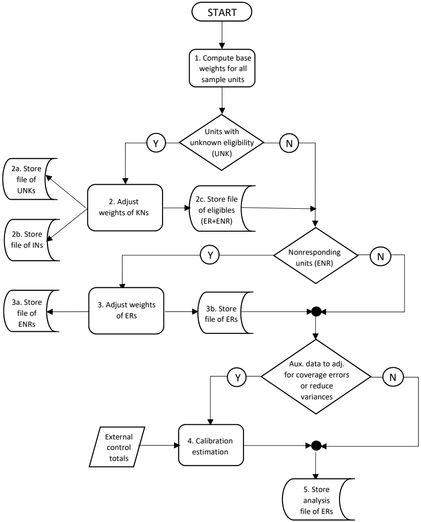
  Valliant, R. & Dever, J.A. (2018). *Survey Weights: A Step-by-Step Guide to Calculation*. College Station TX: Stata Press.

Course Module

General steps in a probability sample

1. Base weights
2. Unknown eligibility
3. Nonresponse adjustment
4. Calibration to population controls

General steps–schematic picture



Base Weights

Base Weights

- Base weight (aka sampling, design, or selection weight) = inverse of selection probability
  $d_{0i} = 1/\pi_i$
- *srs*: $d_{0i} = N/n$
- *stsrs*: $d_{0i} = N_h/n_h$
- Probability proportional to x: $d_{0i} = N\bar{x}_U/nx_i$
- Multistage sample: a weight is computed for each stage of selection and multiplied together to construct the (element-level) base weight

Stratified simple random sample without replacement *stsrswor*–R example

File `stsrs.R`

```
require(PracTools)
require(sampling)
data(nhis)
table(nhis$educ_r)
#   1    2    3    4
#1964  719  933  295
# sort data to put pop in education level order
pop <- nhis[order(nhis$educ_r),]
# select stsrswor
nh <- c(3,4,5,6)
stsrs <- strata(pop, stratanames = c("educ_r"),
            size=nh, method="srswor")
# extract the sample data
samdat <- getdata(pop, stsrs)
samdat$wts <- 1 / samdat$Prob
```

Stratified simple random sample without replacement *stsrswor*–Stata example

File `Ex.2.5_stsrs.do`

```
use "J:\UMich weighting 2019\nhis.dta", clear
sort educ_r
tabulate educ_r, matcell(Nh)
matrix nh = (3, 4, 5, 6)
mat list Nh
mat list nh

* method using foreach loop
foreach i of local edlev{
local n = nh[1,`i']
sample `n' if educ_r == `i', count
}
sort educ_r
gen stsrswt = Nh[1,1]/nh[1,1] if educ_r == 1
replace stsrswt = Nh[2,1]/nh[1,2] if educ_r == 2
replace stsrswt = Nh[3,1]/nh[1,3] if educ_r == 3
replace stsrswt = Nh[4,1]/nh[1,4] if educ_r == 4
list, clean
```

Exercise 11 - Calculate and verify base weights for single-stage stratified design

Stratum	Business Sector	Population Size
1	Manufacturing	6,000
2	Retail	15,000
3	Wholesale	4,000
4	Service	20,000
5	Finance	5,000
Total		50,000

a) Calculate a proportional allocation for the five-stratum design using an overall sampling rate of 5 percent. What is your overall sample size?

b) Calculate the base weights. How might you verify that the weights are correct?

Nonresponse Adjustments

Methods of analyzing nonresponse

- Deterministic

  Every unit is an R or an NR, no random choice

- Stochastic response

  Every unit has a probability of being an R or an NR

Stochastic is the point-of-view that underlies methods used in practice

Types of Stochastic Missingness

- Missing Completely at Random (MCAR)–every unit has same probability of response. Respondents are just a random subsample of initial sample.
- Missing at Random (MAR)–probability of response does not depend on y but does depend on some or all of the auxiliaries x. Response model can be formed that depends on x if auxiliaries known for both respondents and nonrespondents.
- Nonignorable nonresponse (NINR) aka Not missing at random (NMAR)–chances of responding depend on one or more analysis variables (y's). Dependence cannot be eliminated by modeling response based on x 's.

Issues for Nonresponse Adjustment

- How to form cells guided by analysis of response patterns, or y's, or both
- Bias under stochastic response model (Kalton & Maligalig 1991)

$$B_R\left(\hat{\bar{y}}_\pi\right) \doteq \frac{1}{N\bar{\phi}} \sum_{i \in U} (y_i - \bar{Y}_U)\left(\phi_i - \bar{\phi}\right)$$

  $\Longrightarrow$ form cells to have common mean of Y or common response propensity ϕ within each cell
- Bias under superpopulation model. Little & Vartivarian (2005) results say give primacy to forming cells where units all have a common mean.

Issues for Nonresponse Adjustment (continued)

- General approach is form cells to either
    1. Contain units that all have about same response probability, or
    2. All have a common mean of y
- #2 is hard because there are usually many y's and only means for R's are known.
- #1 is usually more feasible
    - But covariates available for forming cells may be related to both y's and response probabilities
    - Also, a limited number of covariates may be available

Propensity score adjustments

1. Fit model to predict response based on available covariates
2. Sort file (R's and NR's both) from low to high based on estimated response propensities
3. Divide into cells (5 to 10 usually enough)
4. Compute NR adjustment in each cell as sum of weights for full sample divided by sum of weights for respondents. Input weights can be base weights or UNK-eligibility adjusted weights for eligible cases. Unweighted adjustment might also be used.
5. Multiply weight of each R in a cell by NR adjustment ratio
6. Only respondents have a non-zero weight after this step.

Example 13.8 in VDK: Form classes from response propensities

2003 NHIS (`nhis.RData`) dataset has 3,911 cases. NRs are persons who answered question on personal income as: Refused, Not Ascertained, Don't Know, or only reported income as above or below $20K (`resp`. About 31% are nonrespondents by this criterion.

Covariates:

`age`	Age (continuous)
`educ_r`	Education recode (1 = high school, general education development degree (GED), or less, 2 = some college, 3 = Bachelor's or associate's degree, 4 = Master's & higher)
`hisp`	Hispanic ethnicity (1 = Hispanic, 2 = non-Hispanic)
`parents_r`	Parent(s) of sample person present in the family (1 = Yes, 2 = No)
`race`	Race (1 = White, 2 = Black, 3 = Other)

Example 13.8 continued

Use logistic model to form 5 nonresponse adjustment classes

```
p.class <- pclass(formula = resp ~ age +
           as.factor(hisp) +
           as.factor(race) +
           as.factor(parents_r) +
           as.factor(educ_r),
           type = "unwtd", data = nhis, link="logit",
           numcl=5)
 table(p.class)
```

```
 (0.453,0.631] (0.631,0.677] (0.677,0.714] (0.714,0.752] (0.752,0.818]
           778           773           788           786           786
```

Nonresponse-adjusted Weights

There are several options for computing a single adjustment in each class *c* (similar results in many applications):

1. $\hat{\phi}_c = \sum_{i \in s_c} \hat{\phi}(x_i) \Big/ n_c$, unweighted average estimated propensity where n_c is the unweighted number of cases in class *c*;
2. $\hat{\phi}_c = \sum_{i \in s_c} d_i \hat{\phi}(x_i) \Big/ \sum_{i \in s_c} d_i$, weighted average estimated propensity, where d_i is the input weight and $\sum_{i \in s_c} d_i = \hat{N}_c$ is the estimated number of pop units in class *c*;
3. $\hat{\phi}_c = n_{cR}/n_c$, unweighted response rate where n_{cR} is the unweighted number of respondents in class *c*;
4. $\hat{\phi}_c = \sum_{i \in s_{cR}} d_i / \sum_{i \in s_c} d_i$, weighted estimate of response rate; and
5. $\hat{\phi}_c = median\left[\hat{\phi}(x_i)\right]_{i \in s_c}$, unweighted median estimated propensity.

Alternative methods of computing NR adjustments within cells

- The function `NRadjClass` in `PracTools` will compute these five class adjustments given output from `pclass`.
- In NHIS data, all methods give similar results. See VDK Table 13.2.

Alternative methods of computing NR adjustments within cells

- Table 13.2. Five methods of estimating response propensities within classes based on fitting a logistic model to the NHIS data

			(1)	(2)	(3)	(4)	(5)
Class	Boundaries	Count of persons	Unweighted avg. propensity	Weighted avg. propensity	Unweighted RR	Weighted RR	Median
1	(0.453,0.631]	778	0.588	0.591	0.589	0.591	0.595
2	(0.631,0.677]	773	0.655	0.655	0.662	0.679	0.657
3	(0.677,0.714]	788	0.696	0.696	0.694	0.702	0.696
4	(0.714,0.752]	786	0.732	0.732	0.707	0.717	0.733
5	(0.752,0.818]	786	0.777	0.778	0.796	0.804	0.775

- In this case, the 5 methods differ very little on the value of the cell NR adjustment.

Checking balance on covariates

- Balance: within-class means of covariates should be about the same for Rs and NRs $\Longrightarrow$ average propensity of response will be about same for Rs and NRs
- D'Agostino (1998) gives simple method for checking covariate balance within the classes formed in propensity stratification.
- After classes formed, idea is to check on the extent of differences in the covariate means.
- The covariate means should be different between classes. Within a class, means of covariates should be the same for respondents and nonrespondents.

Checking balance (continued)

- Use the class output from Example 13.8:
  `p.class <- p.class$p.class`
- Define indicator: `resp=1` if a unit is a respondent and 0 if an NR.
- Fit models for mean of each covariate using `p.class` and `resp` as predictors.
  - For quantitative x's, fit ANOVA model, `x = p.class resp p.class*resp`
  - For dichotomous x's, fit logistic model, `logit(x) = p.class resp p.class*resp`.

Checking balance (continued)

- Coefficients on `resp` and the interaction term `p.class*resp` should be non-significant if covariate means do not differ for R's and NR's within quintile class.
- Coefficients on `p.class` should be nonzero and different from each other since units with different values of propensities, and, consequently covariates, go into the different classes.

VDK Example 13.8 ANOVA table

```
chk1 <- glm(age ~ p.class + resp + p.class*resp,
            data = nhis)
summary(chk1)
Coefficients:
                               Estimate  t value  Pr(>t)
 (Intercept)                      55.91   63.31   < 2e-16 ***
p.class(0.631,0.677]              -7.80   -5.90  4.01E-09 ***
p.class(0.677,0.714]             -10.84   -7.96   < 2e-16 ***
p.class(0.714,0.752]             -13.05   -9.48   < 2e-16 ***
p.class(0.752,0.818]             -22.86  -14.82   < 2e-16 ***
resp                              -0.02   -0.02     0.987
p.class(0.631,0.677]:resp         -0.05   -0.03     0.975
p.class(0.677,0.714]:resp         -1.35   -0.80     0.425
p.class(0.714,0.752]:resp          0.18    0.10     0.917
p.class(0.752,0.818]:resp          1.52    0.83     0.404
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Stata code for propensity classes

File `Ex.3.2_logisticNR.do` has code

```
use "nhis.dta", clear
logit resp age i.hisp i.race i.parents_r i.educ_r
predict predp, pr
* put sample units into propensity classes
pctile qpreds=predp, nq(5) genp(percent)
sort predp
egen pclass=cut(predp), at(0, 0.6306909, 0.6769733,
        0.7137994, 0.7516164, 1.0)
* code below generates 3911 values
egen pavg = mean(predp), by(pclass)
egen pmed = median(predp), by(pclass)
egen RR = mean(resp), by(pclass)
* install wtdmean from ssc
ssc inst _gwtmean
egen pavgwtd = wtmean(predp), weight(svywt) by(pclass)
egen RRwtd = wtmean(resp), weight(svywt) by(pclass)
```

Classification algorithms

- Goal is to classify units as R or NR based on covariates available for all sample cases.
- Input data are the same as for propensity modeling.
- R package `rpart` uses CART algorithm to form classes.
- Advantages of CART compared to propensity modeling are:
  - Interactions of covariates handled automatically.
  - Way in which covariates enter the model does not have to be made explicit.
  - Selection of which covariates and associated interactions should be included done automatically.
  - Variable values, whether categorical or continuous, are combined (grouped) automatically.
- We want to form classes so that we have MAR, i.e., given the x's that define classes, all units have same response probability.

Other classification algorithms

- The CART algorithm in `rpart` tends to select categorical variables with many categories too often, leading to bias.
- `cforest` is better.
- Another choice is GUIDE (Loh, 2014) but it has no R package

Weighting for multistage designs

- Depending on analytic needs, base weights may be calculated for more than one stage of selection
- Adjustments for unknown eligibility and nonresponse may be needed at each stage
- Health & Retirement Study (HRS) and National Survey of Family Growth (NSFG) are examples that UMich does

QC in weight creation

- When checking weights, remember that the numbers of records within an input file entering each step must equal the number of records on the output file exiting the step plus any records discarded.
- The sums of the incoming weights and outgoing weights in each step must balance.
- In addition:
  - Review the distribution of weights at each stage to identify any missing or extreme values.
  - Compute weighted frequencies of important survey characteristics after each weighting adjustment.
  - Compare weighted frequencies to reliable external totals.
  - Produce statistics (e.g. mean, median, minimum, maximum, unequal weighting effect) for each replicate weight after each weight adjustment (if replication variance estimation used).

Course Module

1. Basic Steps in Weighting
2. Weight Calibration
3. Nonprobability Sampling

Idea behind Use of Auxiliaries in Calibration

- Use relationship between analysis variables (Y's) and covariates (x's) to improve estimators
- Reduce variances
- Correct coverage errors
- Need
  - population totals of x's (or good estimates of them) and
  - x's for individual responding sample units
- No need for x's for individual nonsample units and non-responding sample units

Poststratification

Examples of calibration: Poststratification (PS)

Method

- Put units into groups (age groups, regions, types of business)
- Adjust weights so that estimated counts of units equal control counts

$$\hat{t}_{yPS} = \sum_{\gamma=1}^{G} N_\gamma \left(\hat{t}_{y\gamma}/\hat{N}_\gamma\right)$$

- $\hat{t}_{y\gamma} = \sum_{s_\gamma} d_i y_i$ is est'd total of y in poststratum γ based on the input weights d_i (usually base or NR-adjusted weights)
- s_γ is set of sample units in poststratum γ
- $\hat{N}_\gamma = \sum_{s_\gamma} d_k$ is est'd pop size of poststratum γ based on input weights
- N_γ is pop count or control total for poststratum γ, and *G* is the total number of poststrata.

Poststratification (continued)

- Implied final weight for unit i in poststratum γ is

$$w_i = d_i \frac{N_\gamma}{\hat{N}_\gamma}$$

  where $g_i = N_\gamma / \hat{N}_\gamma$ is PS adjustment factor.
- This is also called a *g*-weight if we write the final weight as $w_i = d_i g_i$.
  With that definition of the weight, $\hat{t}_{yPS} = \sum_{i \in s} w_i y_i$, i.e., a weighted sum of the data values.
- Note: $\sum_\gamma w_i = \frac{N_\gamma}{\hat{N}_\gamma} \sum_\gamma d_i = N_\gamma$
- Implied model for poststratification is $y_i = \mu_\gamma + \epsilon_i, \;\; i \in U_\gamma$

Example 14.2 Poststratification

File: `Example 14.2 poststrat.R`; Stata example in `Ex.4.1_poststrat.do`

Table 14.1 Percentages of persons in the large NHIS population who reported receiving Medicaid

	Age group (years)				
Hispanicity	under 18	18-24	25-44	45-64	65+
Hispanic	32.2	10.7	7.6	11.0	27.2
non-Hispanic white	12.6	6.6	3.8	3.1	3.7
non-Hispanic Black and other race/ethnicity	31.3	12.7	8.8	6.4	16.5

- Percentages depend on both age and race-ethnicity

Example 14.2 (continued)

```
    # create single variable to identify age.grp x
    # hisp.r poststrata
    # ... some code omitted; see book ...
N.PS <- xtabs(~age.grp + hisp.r, data = nhis.large1)
         # select  srswor of size n
set.seed(-1570723087)
n <- 250
N <- nrow(nhis.large1)
sam <- sample(1:N, n)
samdat <- nhis.large1[sam, ]
```

Example 14.2 (continued)

```
        # compute srs weights and sampling fraction
d <- rep(N/n, n)
f1 <- rep(n/N, n)

        # srswor design object
nhis.dsgn <- svydesign(ids = ~0,          # no clusters
                       strata = NULL,     # no strata
                       fpc = ~f1,
                       data = data.frame(samdat),
                       weights = ~d)

        # poststratified design object
ps.dsgn <- postStratify(design = nhis.dsgn,
                        strata = ~age.grp + hisp.r,
                        population = N.PS)
```

Checking PS results

- Verify that PS weights sum to pop counts using the `svytotal` function. Only first four of 15 poststrata are shown.
- SEs of estimates are zero since there is no variation from sample to sample in the estimates—they will always equal the pop counts.

```
          # Check that weights are calibrated for x's
svytotal(~interaction(age.grp, hisp.r), ps.dsgn)
                                total SE
interaction(age.grp, hisp.r)1.1 1952 0
interaction(age.grp, hisp.r)2.1  581 0
interaction(age.grp, hisp.r)3.1 1574 0
interaction(age.grp, hisp.r)4.1  704 0
```

- Weights of individual sample cases can be examined with the command, `weights(ps.dsgn)`.

Checking PS results

- The estimated proportion of persons receiving Medicaid, their SEs, and coefficients of variation (CV) are produced by:

```
        # PS standard errors and cv's
svytotal(~ as.factor(medicaid), ps.dsgn, na.rm=TRUE)
cv(svytotal(~ as.factor(medicaid), ps.dsgn, na.rm=TRUE))
        # srs standard error and cv's
svytotal(~ as.factor(medicaid), nhis.dsgn, na.rm=TRUE)
cv(svytotal(~ as.factor(medicaid), nhis.dsgn, na.rm=TRUE))
```

- `na.rm=TRUE`, is used because some cases have missing values for Medicaid; without it, results will all be `NA` (i.e., missing).
- `as.factor` forces Medicaid to be treated as a class (factor) variable.

	total	SE	CV
Poststratified	1870.8	344.5	0.184
srswor	1899.7	385.3	0.203

Raking

Raking: controls and model

- If several factor controls (age group, sex, race-ethnicity) available, we might use only marginal controls, not crosses
- A margin can be the combination of several variables, e.g., age*sex
- Implied model for raking is $y_i = \mu + \alpha_j + \beta_k + \epsilon_i,\ i \in U_{jk}$
- `Example 14.4 raking.undercoverage.R`: Raking by age group and Hispanicity

Raking: Example 14.4

Same *srswor* sample of 500 and survey design object, `nhis.dsgn`, used as in Example 14.2. `calibrate` function does the raking. Alternative is function `rake`, which will give same answer.

```
        # create marginal pop totals
N.age <- table(nhis.large$age.grp)
N.hisp <- table(nhis.large$hisp.r)
pop.totals <- c('(Intercept)' = N, N.age[-1], N.hisp[-1])

        # create raked weights
rake.dsgn <- calibrate(design = nhis.dsgn,
          formula = ~as.factor(age.grp) + as.factor(hisp.r),
          calfun = "raking",
          population = pop.totals)
```

General Regression Estimation (GREG)

General regression estimation

- Categorical and continuous variables can be used
- Estimator of total is

$$\begin{aligned}\hat{t}_{yGREG} &= \hat{t}_y + \left(\mathbf{t}_x - \hat{\mathbf{t}}_x\right)^T \hat{\mathbf{B}} \\ &= \sum_{k\in s}\left[1 + \left(\mathbf{t}_x - \hat{\mathbf{t}}_x\right)^T \left(\sum_{l\in s_A} \pi_l^{-1}\mathbf{x}_l\mathbf{x}_l^T\right)^{-1} \mathbf{x}_k\right] \pi_k^{-1} y_k\end{aligned}$$

- $\hat{t}_y$ is est'd total using input weights (base or NR-adjusted)
- $\mathbf{t}_\mathbf{x}$ is vector of pop totals of x's
- $\hat{\mathbf{t}}_x$ is vector of estimated pop totals of x's using input weights
- $\hat{\mathbf{B}}$ is (input weighted) slope of y on $\mathbf{x}$
- Underlying model for GREG is $y_i = \mathbf{x}^T\beta + \varepsilon_i$, $\varepsilon_i \sim (0, v_i)$

`calibrate` function

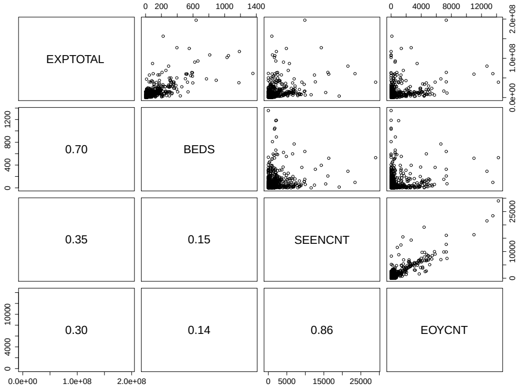
- `calibrate` function accepts a number of parameters:

`design`	survey design object
`formula`	model formula for calibration model
`population`	Vectors of population column totals for the model matrix in the calibration model, or list of such vectors for each cluster.
`calfun`	Calibration function. Allowable values are `calfun=c("linear","raking","logit","rrz")`. The function is flexible enough to accept a user-defined distance function also.

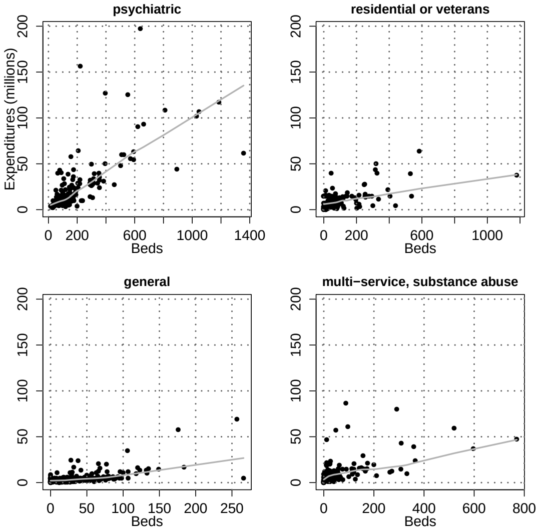
- Output
  Design object with information on strata, clusters, weights, data values for each unit in input design object

GREG example

- Survey of Mental Health Organizations population. `smho.N874` contains 874 hospitals
- Figure 14.2: Scatterplot matrix



Plots of expenditures vs beds for 4 hospital types



GREG example

File: `greg.smho98.R`

- Select *pp*(beds) sample & compute base weights

```
require(sampling)
x <- smho[,"BEDS"]
    # recode small hospitals to have a minimum MOS
x[x <= 5] <- 5
x <- sqrt(x); n <- 80
set.seed(428274453)
pk <- n*x/sum(x)
sam <- UPrandomsystematic(pk)
sam <- sam==1
sam.dat <- smho[sam, ]
d <- 1/pk[sam]
```

GREG example

- Create design object that is used in `calibrate` function to compute GREG weights.
- Calculate population (control) totals for calibration.

```
smho.dsgn <- svydesign(ids = ~0,            # no clusters
                       strata = NULL,        # no strata
                       data = data.frame(sam.dat),
                       weights = ~d)

         # Compute pop totals of auxiliaries
         # Note these are the original not the recoded x's
x.beds <- by(smho$BEDS, smho$hosp.type, sum)
x.seen <- sum(smho$SEENCNT)
x.eoy <- sum(smho$EOYCNT)
N <- nrow(smho)
pop.tots <- c(`(Intercept)` = N,
              SEENCNT = x.seen,
              EOYCNT = x.eoy,
              x.beds = x.beds)
```

GREG example

- Create design object using `calibrate`

```
sam.lin <- calibrate(design = smho.dsgn,
               formula = ~SEENCNT + EOYCNT +
                    as.factor(hosp.type):BEDS,
               population = pop.tots,
               calfun="linear")
```

- `calfun="linear"` gives GREG weights
- check whether calibration constraints are satisfied:

```
svytotal(~SEENCNT, sam.lin)
          total SE
SEENCNT 1349241  0
svytotal(~EOYCNT, sam.lin)
        total  SE
EOYCNT 505345   0
```

GREG example

- GREG can produce negative weights

```
summary(weights(smho.dsgn))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.714   5.693   8.150   8.763  10.090  33.680

summary(weights(sam.lin))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.3983  5.7470  8.8320  9.0630 10.9300 33.8300
```

GREG example

- `bounds` parameter can restrict weight changes

```
sam.linBD <- calibrate(design = smho.dsgn,
                  formula = ~SEENCNT + EOYCNT +
                      as.factor(hosp.type):BEDS,
                  population = pop.tots,
                  bounds = c(0.4, 3),
                  calfun = "linear")
```

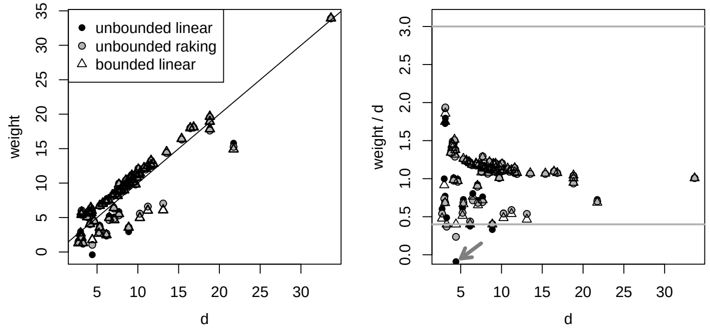
- `bounds` restricts relative change from initial weights to final weights

$$L \le \frac{w_i}{d_i} \le U \text{ for all } i \in s.$$

- $L = 0.4$ and $U = 3$ in code above

Figure 14.5

Plots of weights for the different methods of calibration in a *pps sample.* A 45^o line is drawn in the left-hand panel. Reference lines are drawn at the weight bounds, 0.4 and 3, in the right-hand panel.



GREG example (continued)

Table 14.6. Estimated totals of expenditures and proportion of hospitals with direct state financing, SEs, and CVs for π-estimate, and GREG without and with hospital controls in a *pps* sample from a subset of the Survey of Mental Health Organizations population

	Estimate or population value	SE	*CV* (%)
Total expenditures (000s)			
Population	8,774,651		
π -estimate	9,322,854	915,126	9.82
GREG 1 (no hosp controls)	9,563,683	748,596	7.83
GREG 2 (hosp controls)	9,161,491	711,633	7.77
Proportion with financing from state mental health agency			
Population	0.336		
π-estimate	0.323	0.059	18.16
GREG 1	0.303	0.051	16.92
GREG 2	0.340	0.034	9.91

Stata code

In Stata the procedure to use is `svycal`

Stata files to do the same GREG and bounded weight GREG:

`Ex.4.6_4.7_smho.greg.do`

`Ex.4.8_4.9_smho.greg.bdd.do`

Variation in Weights

Reasons for weights that vary in size

1. Varying selection probabilities as would occur in *pps* sampling or stratified sampling with different sampling rates in strata;
2. Over- or under-sampling groups of units in two-phase sampling based on domain membership;
3. Unequal response rates (and/or rates of unknown eligibility) in different subgroups leading to unequal weight adjustments;
4. Calibration to auxiliaries to reduce variances or correct for frame coverage errors.

- Varying weights may be designed into the sample, as in (1) and (2).
- Varying weights may be needed to correct for potential nonresponse bias or differential under-coverage as in (3) and (4).
- However, highly differential weights can increase variances of estimates even if they decrease bias.

Kish & other measures

- Kish (1965, 1992) design effect due to weighting, equal to one plus the relvariance of the sample weights:

$$\begin{aligned} \mathit{deff}_w &= 1 + relvar\,(w) \\ &= 1 + n^{-1}\textstyle\sum_s (w_i - \bar{w})^2 \Big/ \bar{w}^2 \end{aligned}$$

- Widely used, and possibly over-used, measure. Interpreted as the increase in variance of an estimator due to having weights that are not all the same.
- `PracTools` function: `deff` computes Kish, Henry, Spencer, and Chen-Rust design effects

Course Module

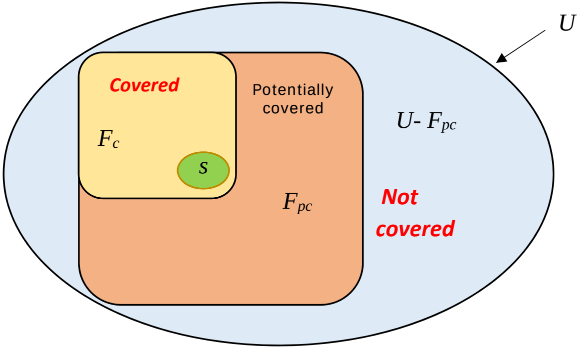
Motivation for Nonprobability Sampling

- Low response rates for many probability samples (Kohut et al. 2012)
- Ever increasing costs with ever decreasing funds
- Nonsampling errors
- The need for speed (e.g., time-sensitive measures)
- Desire to supplement subgroups underrepresented in a probability survey (e.g., because of high nonresponse)
- No defined (definable) sampling frame (e.g., IV drug users)
- Data are everywhere just waiting to be analyzed!!!

Examples of “New-ish” Sources of Data

- Twitter
- Facebook
- Snapchat
- Mechanical Turk
- SurveyMonkey
- Web-scraping
- Pop-up Surveys
- Data warehouses
- Probabilistic matching of multiple sources

Fig 18.1: Universe and sample with coverage errors



For example ...

- U = adult population
- $U - F_{pc}$ = adults without internet access
- F_{pc} = adults with internet access
- F_c = adults with internet access who visit some webpage(s)
- s = adults who volunteer for a panel

Probability vs. Nonprobability Samples

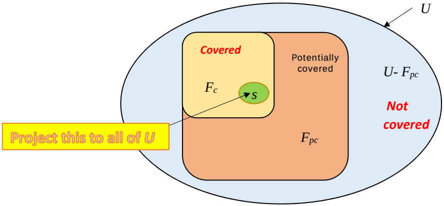
- Many applications of big data analysis use nonprobability samples. Population may not be well defined.
- Many probability surveys have such low RRs they basically are nonprobability samples
  • Pew Research response rates in typical telephone surveys dropped from 36% in 1997 to 9% in 2012 (Kohut et al. 2012)
- Recommendations for using nonprobability samples:
  - AAPOR task force reports on nonprobability samples (2013) & online samples (2010)
  - Perils and potentials of self-selected entry (Keiding & Louis 2016)

Example Studies with Nonprobability Estimates

- Analysis of medical records including text to predict heart disease (Giles & Wilcox 2011)
- Correlates of local climate & temperature with spread of infectious disease (Global Pandemic Initiative)
- MIT's Billion Prices Project–Price indexes for 22 countries from web-scraped data (Cavallo & Rigobon 2016)
- Marketing of e-cigarettes (Kim et al. 2015)
- Political polls and political issues (e.g., Clement 2016; Conway et al. 2015; Dropp & Nyhan 2016)

What is Required to Estimate a Population Total?

- Pop total $t_y = \sum_s y_i + \sum_{F_c - s} y_i + \sum_{F_{pc} - F_c} y_i + \sum_{U - F_{pc}} y_i$
- To estimate t, predict 2nd, 3rd, and 4th sums



- What if non-covered units are much different from covered?
- Difference from a bad probability sample with a good frame but low response rate:
  - No unit in $U - F_{pc}$ or $F_{pc} - F_c$ had any chance of appearing in the sample

Three Approaches to Weighting Nonprob Sample

(1) Quasi-randomization weighting

- Combine nonprob and prob samples
- Estimate Pr(inclusion in nonprob) based on covariates
- Use inverse of estimated probabilities as weight

(2) Superpopulation modeling of y's

- Fit model for y based on covariates
- Weights are based on model
- Use model weights for estimation

(3) Doubly robust

- Fit models for Pr(inclusion in nonprob) and y's
- Weights account for both

All three involve modeling (Elliott & Valliant 2017)

Quasi-randomization

- Estimate pseudo-inclusion probabilities, typically via logistic regression
- Combine nonprobability sample with a reference (probability) sample that covers whole target pop
- Two ways to estimate
  1. Estimating equation method (Chen, Li, Wu 2020)
  2. Adjusted logistic propensities
     - Combine nonprob and reference samples
     - Assign weights of 1 to nonprob cases, survey wts to reference sample
     - Run weighted logistic regression to estimate $Pr(i \in s_{nonprob})$
     - Make small bias-correction adjustment to estimated probs (Wang, Li, Valliant 2020)
     - Use inverse probs as weights for nonprob sample

Superpopulation models

- Estimated totals have form:

$$\hat{t}_y = \sum_{i \in s_{nonprob}} y_i + \sum_{i \in \bar{s}_{nonprob}} \hat{y}_i$$
$$= \sum_{i \in s_{nonprob}} w_i y_i$$

  $\bar{s}_{nonprob} = U - s_{nonprob}$
  $\hat{y}_i$ is a regression prediction of y_i

- Weight for unit i is same one as R `calibrate` or Stata `svycal` produce if input weights are all 1

See VDK (2018, ch. 18)

Doubly robust

- Approximately unbiased estimates if either (i) quasi-randomization model is correct or (ii) superpopulation model is correct
- Estimate pseudo-inclusion probabilities
- Use inverses as weights for input to R `calibrate` or Stata `svycal`
- Calibrate to pop totals of x's in superpopulation model

See VDK (2018, ch. 18)

Combining probability and nonprobability samples

- Issues
  - Does probability sample represent the target population?
  - Does nonprobability sample cover part or all of the target population?
  - Should nonprob sample be weighted to represent part or all of target pop?
- Assign weights to prob sample that inflates it to target pop
- Weight nonprob sample to the part of the target pop that it represents
- Compute a composite estimate as a weighted average of estimates from the prob sample and nonprob sample

References

- Valliant, R. & Dever, J.A. (2018). *Survey Weights: A Step-by-Step Guide to Calculation*. College Station TX: Stata Press
- Valliant, R., Dever, J.A., & Kreuter, F. (2018). *Practical Tools for Designing and Weighting Survey Samples*, 2nd edition. New York: Springer. **
- AAPOR Guidance on Reporting Precision for Nonprobability Samples (2016). AAPOR Standards Committee report chaired by C Kennedy C and T Buskirk. https://www.aapor.org/getattachment/Education-Resources/For-Researchers/AAPOR_Guidance_Nonprob_Precision_042216.pdf.aspx.
- Baker R., Brick J. M., Bates N. A., Couper M. P., Courtright M., Dennis J. M., Dillman D. A., Frankel M. R., Garland P., Groves R. M., Kennedy C., Krosnick J., Lavrakas P. J., Lee S., Link M. W., Piekarski L., Rao K., Thomas R. K., Zahs D. (2010). AAPOR report on online panels. *Public Opinion Quarterly* 74:711Ű781.
- Baker R., Brick J. M., Bates N. A., Battaglia M. P., Couper M. P., Dever J.A., Gile K., Tourangeau R. (2013). *Report of the AAPOR task force on non-probability sampling*. Tech. rep., The American Association for Public Opinion Research, Deerfield, IL.
- Cavallo A & Rigobon R (2016). The billion prices project: Using online prices for measurement and research. *The Journal of Economic Perspectives*, 30(2):151–178.
- Clement S (2016). How The Washington Post-SurveyMonkey 50-state poll was conducted. The Washington Post, 6 Sep 2016. https://www.washingtonpost.com/news/post-politics/wp/2016/09/06/how-the-washington-post-surveymonkey-50-state-poll-was-conducted/?utm_term=.5be59d7ebf32
- Chen Y, Li P, and Wu C. (2020). Doubly Robust Inference with Non-probability Survey Samples. *Journal of the American Statistical Association*. accepted, https://arxiv.org/abs/1805.06432.

- Chen S., Rust K. F. (2017). An extension of KishŠs formula for design effects to two- and three-stage designs with stratification. *Journal of Survey Statistics and Methodology* 5(2):111Ű130.
- Conway BA, Kenski K & Wang D (2015). The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. *Journal of Computer-Mediated Communication*, 20(4):363–380.
- D'Agostino, R.B. (1998). Propensity score methods for bias reduction for the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 17:2265-2281.
- Dropp K & Nyhan B (2016). The 2016 Race: It Lives. Birtherism Is Diminished but Far From Dead. The New York Times, 23 Sep 2016. `https://www.nytimes.com/2016/09/24/upshot/it-lives-birtherism-is-diminished-but-far-from-dead.html`.
- Elliott MR & Valliant R (2017). Inference from nonprobability samples. *Statistical Science*, 32(2):249–264.
- Giles T & Wilcox R (2011). IBM Watson and Medical Records Text Analytics. Presentation at the *Healthcare Information and Management Systems Society Meeting*. Accessed 25 Sep 2016: `https://www-01.ibm.com/software/ebusiness/jstart/downloads/MRTAWatsonHIMSS.pdf`.
- Henry K. A., Valliant R. (2015). A design effect measure for calibration weighting in singleŰstage samples. *Survey Methodology* 41:315Ű331.
- Kalton, G. & Maligalig, D.S. (1991). A comparison of methods of weighting adjustment for nonresponse. Proceedings of the US Bureau of the Census Annual Research Conference, pp 409-428, `http://babel.hathitrust.org/cgi/pt?id=umn.31951d00357147b;view=1up;seq=417`.
- Keiding N & Louis TA (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *Journal of the Royal Statistical Society*, Series A, 179(2):319–376.
- Kim AE, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, Guillory J, & Porter L (2015). Using Twitter Data to Gain Insights into E-cigarette Marketing and Locations of Use: An Infoveillance Study. *J Med Internet Res*, 17(11):e251.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.
- Kish, L. (1992). Weighting for unequal pi. *Journal of Official Statistics*. 8(2):183-200.

- Kohut A, Keeter S, Doherty C, Dimock M, & Christian L (2012). Assessing the representativeness of public opinion surveys, 15 May 2012. Accessed 9 Sep 2016: http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys.
- Little, R.J.A. & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology* 31:161-168.
- Loh, W. Y. (2014). Fifty years of classification and regression trees (with discussion). *International Statistical Review*, 34, 329–370.
- Spencer B. D. (2000). An approximate design effect for unequal weighting when measurements may correlate with selection probabilities. *Survey Methodology* 26(2):137Ű138.
- Wang L, Li Y, Valliant R. (2020). Adjusted Logistic Propensity Weighting Method for Non-probability Survey samples. In preparation.

1. Basic Steps in Weighting

2. Weight Calibration

3. Nonprobability Sampling