

# Subject index

## A

- AAPOR . . . see American Association of Public Opinion Research
- American Association of Public Opinion Research
  - margins of error in nonprobability samples . . . . . 132
  - reports on nonprobability samples . . . . . 6
- attrition in opt-in panels . . . . . 114

## B

- bagging (bootstrap aggregation) . . . . 42
- balanced repeated replication . . . . . 89
  - example using `survwgt` . . . . . 92
  - example using `svygen` . . . . . 91
  - Fay method . . . . . 90
  - grouping strata and PSUs . . . . . 93
  - Hadamard matrix . . . . . 90
  - replicate weights . . . . . 90
- BART . . . . . see Bayesian additive regression trees
- base weights . . . . . 14, 19
- baseball all-star voting . . . . . 109
- Bayesian additive regression trees . . . 43
- Bayesian approach
  - multilevel regression and poststratification . . . . . 131
  - nonprobability samples . . . . . 131
- book materials . . . . . xv
- boost, Stata plugin . . . . . 43
  - example using . . . . . 44
  - parameters for . . . . . 44
- boosting . . . . . 43
  - explanation of . . . . . 43

- bootstrap . . . . . 94
  - National Maternal and Infant Health Study (NMIHS) example . . . . . 96
  - number of replications . . . . . 95
  - use by Statistics Canada . . . . . 95
  - weights . . . . . 95
- BRR . . . . . see balanced repeated replication
- `bsweights` . . . . . 97

## C

- calibration . . . . . 14, 51
  - choice of covariates . . . . . 157
  - collect consistent with external sources . . . . . 156
  - contribution to variance of using sample estimates . . . . . 70
  - population totals needed . . . . . 51
  - poststratification . . . . . 52
  - superpopulation models underlying . . . . . 52
  - to sample estimates . . . . . 69
- California Health Interview Survey . . 87
- CDC . . . . . see Centers for Disease Control
- Centers for Disease Control . . . . . 109
- chain referral sampling . . . . . 108
- CHIS . . . . . see California Health Interview Survey
- classification and regression trees . . . 42
  - advantages of . . . . . 42
- clustering, uses of . . . . . 10
- collapsing adjustment cells . . . . . 102
- control totals, recovering from
  - public-use files . . . . . 55
- convenience samples . . . . . 106
- convergence of iterative procedures
  - problems in replicates . . . . . 103

coverage ..... 3  
     and selection bias ..... 110  
     overcoverage ..... 3  
     undercoverage ..... 3  
 Cramér's  $V$  ..... 41  
 current population survey ..... 4

**D**

degrees of freedom ..... 80  
 descriptive statistics ..... 4  
 design and planning stage  
     quality control at ..... 152  
 design effect because of weighting... 71  
 design unbiased ..... 7  
 deterministic model for nonresponse...  
     ..... 31  
 diagram of universe and frame ..... 3  
 directory, working ..... xv  
 disposition codes ..... 12  
     categories used in weighting... 12  
     examples of in military personnel  
     survey ..... 13  
 documentation of survey procedures...  
     ..... 159

**E**

election polls, failures of ..... 107  
 eligible units ..... 2, 3  
 estimated controls ..... 69  
 estimating equations ..... 139

**F**

Fay method of balanced repeated repli-  
 cation ..... 90  
 filenaming conventions ..... 17  
 finite population ..... 2  
     correction factors ..... 75  
 flowchart of weighting steps ..... 16  
 fpc  
     ad hoc in ultimate cluster variance  
     estimator ..... 78  
     specifying in `svyset` ..... 77  
 frame sampling ..... 2

**G**

general regression (GREG) estimator...  
     ..... 63  
     bounded weight changes ..... 67  
     example of ..... 64  
     example of bounded weight changes  
     ..... 68  
     incorrect specification in `svyset`..  
     ..... 67  
     using `svycal` ..... 64  
     weights ..... 64  
 grouping strata and PSUs ..... 85, 98  
     balanced repeated replication... 93  
     in replication variance estimates..  
     ..... 85  
     jackknife example ..... 86  
     specifying in `svyset` ..... 99  
     weight adjustments in jackknife...  
     ..... 85

**H**

Hadamard matrix ..... 90, 91  
 hierarchical regression model  
     in nonprobability samples ..... 131

**I**

ignorability of sample design in model  
     fitting ..... 140  
 imputation ..... 11  
 inference  
     Bayesian ..... 9  
     design-based ..... 7  
     model-assisted ..... 8  
     model-based ..... 7  
 Internet coverage  
     in European Union ..... 112  
     in United States ..... 113  
 interpretation of model parameter esti-  
     mates ..... 4  
`ipfraking` command ..... 60

**J**

- jackknife
  - deleting PSUs..... 83
  - example of grouping strata and PSUs..... 86
  - grouping strata and PSUs..... 85
  - JK2 version..... 88
  - replicate weights..... 83
  - stratified..... 83
- JK1 jackknife..... 84
- JK2 jackknife..... 88
- JKn jackknife..... 84

**L**

- linear substitute in variance estimation..... 81
- linearization variance estimators.... 80

**M**

- MART.. see multiple additive regression trees
- measurement error in opt-in panels.... 115
- methods of sampling
  - Bernoulli..... 10
  - clustering..... 10
  - equal probability..... 9
  - multistage..... 11
  - Poisson..... 10
  - probability proportionate to size.. 10
  - single-stage..... 11
  - stratification..... 10
- model fit
  - boxplots of predictions..... 48
  - Cramér's  $V$ ..... 41, 46
  - logistic versus boost..... 47
  - majority vote method..... 41, 45
  - skewness of predictions..... 48
- model fitting
  - census model..... 140
  - estimating equations..... 139
  - ignorability of sample design.. 140
  - setting weights to 1, accounting for strata and clusters..... 149

model fitting, *continued*

- testing whether to use weights.... 143
  - linear regression example... 144
  - logistic regression example.. 147
- using weights in..... 138
- weighted versus unweighted models..... 140
  - example of..... 141
- model parameters, estimate of..... 4
- model unbiased..... 7
- model-based weighting..... 124
  - example using `svycal`..... 126
  - formulas for weights..... 125
  - variance estimation..... 128
    - leverage adjustments..... 129
- models for estimators
  - poststratification..... 53
  - raking..... 59
- MRP..... see multilevel regression and poststratification
- multilevel regression and poststratification
  - in nonprobability samples..... 131
- multiple additive regression trees.... 43
- multiple weighting steps
  - effects on variances..... 99
  - example using `survwgt`..... 100
- multiple weights..... 134
- multistage sampling..... 25
  - uses of..... 11

**N**

- nearest-neighbor method for sample matching..... 122
- network sampling
  - chain referral..... 108
  - respondent driven..... 108
  - snowball..... 108
- nonlinear estimator..... 80
- nonprobability samples..... 6, 105
  - Bayesian approach..... 131
  - convenience samples..... 106
  - covariate requirements..... 121
  - hierarchical regression model.. 131

- nonprobability samples, *continued*
- incorrect election polls ..... 107
  - model-based weighting ..... 124
    - variance estimation ..... 128
  - network sampling ..... 108
  - observational studies ..... 107
  - opt-in web samples ..... 106
  - prediction approach ..... 125
  - presidential election polls ..... 105
  - pseudo-inclusion probabilities .. 118
  - quasirandomization approach to
    - weighting ..... 115, 116
  - reference samples ..... 117
  - river samples ..... 106
  - sample matching ..... 108, 122
  - superpopulation modeling
    - approach to weighting .... 115
  - types of ..... 106
  - variance estimation ..... 127
    - quasirandomization example ...
      - ..... 127
    - superpopulation model example
      - ..... 130
  - volunteer samples ..... 106
  - weighting ..... 115
  - Xbox sample ..... 106
- nonresponse
- MCAR, MAR, NMAR, NINR ..... 33
  - bias in deterministic model ..... 31
  - bias in stochastic model ..... 32
  - in opt-in panels ..... 114
  - models for ..... 31
- nonresponse adjustment ..... 14
- boosting ..... 43
  - classification and regression trees ..
    - ..... 42
  - example of ..... 35
  - example using propensity scores ..
    - ..... 37
  - multistage designs ..... 48
  - propensity scores ..... 36
  - random forests ..... 42
  - ratio of weight sums ..... 34
  - `survwgt`, use for ..... 100
  - tree-based algorithms ..... 41
- nonresponse adjustment, *continued*
- weighting class adjustments .... 34
  - weighting classes ..... 34
- nonresponse follow-up ..... 136
- normalized weights ..... 133
- objections to ..... 134
  - use in hierarchical linear models ..
    - ..... 134
- O**
- observational studies ..... 107
- opt-in panels
- attrition ..... 114
  - coverage and selection bias .... 110
  - measurement error ..... 115
  - nonresponse ..... 114
  - problems with ..... 110
- opt-in web samples ..... 106
- overcoverage ..... 3, 51
- P**
- population
- finite ..... 2
  - target ..... 2
- population inference, theories of ..... 7
- poststratification ..... 52
- misspecifying `svyset` ..... 55
  - model for ..... 53
  - saving weights with `svycal` .... 54
  - specifying in `svyset` ..... 55
  - superpopulation model underlying
    - ..... 53
  - using `svycal` ..... 54
  - using `svyset` ..... 53
  - weights ..... 52
- prediction approach ..... 125
- weights ..... 126
- prediction approach to weighting .. 124
- presidential election polls ..... 105
- probability proportional to size
- sampling ..... 24
  - example of ..... 25
  - measures of size ..... 25

- probability sampling ..... 5  
     base weights ..... 5  
     definition of ..... 5  
 propensity scores ..... 36  
     complementary log-log model... 37  
     forming classes based on quintiles  
         ..... 39  
     logistic model ..... 37  
     probit model ..... 37  
 public-use files  
     National Health and Nutrition Ex-  
         amination Survey (NHANES)..  
             ..... 79  
     National Maternal and Infant  
         Health Study (NMIHS) ..... 95  
     National Maternal and Infant  
         Health Survey (NMIHS) ... 147  
     replication variance estimation in..  
         ..... 82  
     Residential Energy Consumption  
         Survey (RECS) ..... 93, 144  
     Statistics Canada ..... 95  
 pwr estimate ..... 19
- Q**
- quality control ..... 151  
     analytic checks ..... 158  
     base weights ..... 153  
     data editing and file preparation..  
         ..... 154  
     design and planning stage ..... 152  
     disposition codes ..... 154  
     documentation ..... 159  
     master database ..... 153  
 quasirandomization approach to  
     weighting in nonprobability  
         samples ..... 115  
 quasirandomization to weighting... 116  
 quintiles for forming propensity classes  
     ..... 39
- R**
- raking ..... 55  
     bounds on weight adjustments.. 60  
     example of ..... 59
- raking, *continued*  
     **ipfraking** ..... 60  
     jackknife example using **survwgt**..  
         ..... 86  
     model for ..... 59  
     **sreweight** ..... 62  
 random forests ..... 42  
 random-number seed ..... 21  
 recovering control totals from public-use  
     files ..... 55  
     example of ..... 56  
 reference samples ..... 117  
     combining with nonprobability  
         sample ..... 117  
 replicate sample releases ..... 28  
 replication variance estimation ..... 81  
     balanced repeated replication... 89  
     bootstrap ..... 94  
         replicate weights ..... 95  
     convergence problems in replicates  
         ..... 103  
     grouping PSUs ..... 98  
     in public-use files ..... 82  
     JK1 jackknife ..... 84  
     JK2 jackknife ..... 88  
     JKn jackknife ..... 84, 100  
     number of bootstrap samples... 95  
     using **survwgt** ..... 92  
 respondent driven sampling ..... 108  
 response rates, decline of ..... 105  
 river samples ..... 106
- S**
- sample estimates used for calibration..  
     ..... 69  
 sample matching ..... 108  
     imputing missing values ..... 124  
     in nonprobability samples ..... 122  
     nearest-neighbor method ..... 122  
 sample releases ..... 28  
 sampling  
     frame ..... 2  
     methods of ..... 9  
 sandwich variance estimator ..... 81

- score variable in variance estimation ..  
     ..... 81
- seed, random-number ..... 21
- simple random sampling ..... 20
- skewness of predicted response propen-  
     sities ..... 48
- snowball sampling ..... 108
- srweight** command ..... 62
- srswor example ..... 20  
     fixed rate ..... 21  
     fixed size ..... 20
- srswr example ..... 21
- Stata commands  
     **ipfraking** ..... 60  
     **sample** ..... 27  
     **samplepps** ..... 27  
     **srweight** ..... 62  
     **survwgt** ..... 63, 82, 100
- Stata package, **svr** ..... 63, 82, 92, 100,  
     102
- Statistics Canada ..... 95
- stochastic model for nonresponse... 31
- stratification, uses of ..... 10
- stratified simple random sampling .. 22  
     example ..... 23
- superpopulation modeling approach to  
     weighting in nonprobability  
     samples ..... 115
- survey weights ..... 4
- survwgt**  
     command ..... 63  
     nonresponse adjustment ..... 100  
     raking example ..... 100
- svyset** command ..... 95
- systematic sampling ..... 22
- T**
- target population ..... 2
- totals, estimator of ..... 11
- tree-based algorithms ..... 41
- U**
- UCLA Institute for Digital Research and  
     Education ..... 19
- ultimate cluster variance estimator  
     multistage sampling ..... 79  
     single-stage sampling ..... 78
- undercoverage ..... 3, 51
- unequal weighting effect ..... 71  
     use in quality checks ..... 158
- universe, diagram of ..... 3
- unknown eligibility adjustment .. 14, 29
- UWE ..... see unequal weighting effect
- V**
- variance estimation ..... 75  
     ad hoc fpc ..... 79  
     balanced repeated replication... 89  
     bootstrap ..... 94  
         Rao–Wu method ..... 97  
         weights ..... 95  
     degrees of freedom ..... 80  
     effects of multiple weighting steps  
         ..... 82, 99  
     example using **survwgt** ..... 100  
     exact formulas ..... 75  
     finite population correction fac-  
         tors ..... 75  
     simple random sampling without  
         replacement ..... 75  
     grouping PSUs ..... 98  
         to form replicates ..... 85  
     jackknife ..... 83  
         example using **survwgt** ..... 86  
         grouping PSUs to form replicates  
             ..... 85  
         replicate weights ..... 83  
         stratified ..... 83  
         weight adjustments ..... 83  
     JK1 jackknife ..... 84  
     JK2 jackknife ..... 88  
     JKn jackknife ..... 84, 100  
     linearization estimators ..... 80  
         linear substitute ..... 81  
         score variable ..... 81  
     multistage example ..... 79  
     number of bootstrap samples... 95  
     one PSU per stratum design .... 99  
     replication estimators ..... 81

variance estimation, *continued*  
 sandwich estimator ..... 81  
 simple random sampling without  
 replacement  
 proportions ..... 76  
 stratified simple random sampling  
 example ..... 78  
 stratified simple random sampling  
 without replacement ..... 76  
`survwgt` procedure ..... 82  
`svr` package ..... 82  
`svyset` syntax ..... 77  
 ultimate cluster estimator ..... 77  
 with replacement workaround .. 77  
 volunteer samples ..... 106  
 pseudo-inclusion probabilities .. 118  
 volunteer web surveys ..... 109

**W**

web panel ..... 6  
 webographics ..... 118  
 weight variability ..... 71  
 weighting  
 flowchart of steps ..... 16  
 in nonprobability samples ..... 115  
 weighting class adjustments ..... 34  
 weighting versus imputation ..... 11  
 weights  
 analytic checks ..... 158  
 descriptive statistics ..... 4  
 GREG  
 using `svycal` ..... 64  
 in model fitting ..... 138  
 outlier checks ..... 158  
 quality checks ..... 157  
 raked weights with `ipfraking` .. 60  
 raked weights with `sreweight` .. 62  
 raked weights with `svycal` ..... 59  
 reasons for using  
 nonresponse ..... 1  
 probabilities of selection ..... 1  
 reduce standard errors ..... 1  
 unknown eligibility ..... 1

weights, *continued*  
 unequal weighting effect ..... 158  
 use of ..... 11  
 with-replacement sampling ..... 19

**X**

Xbox nonprobability sample ..... 106

**Z**

zip archive ..... xv