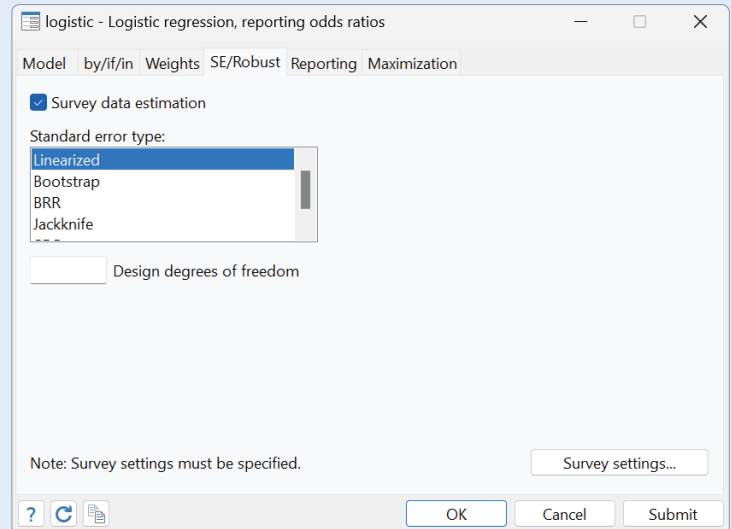


Survey data

- Account for survey design in tabulations, summary statistics, and most regression models
- Sampling design
 - Sampling weights
 - Stratification
 - Clustering
 - Multistage
 - Finite population corrections
- Variance estimates
 - Taylor-series linearization
 - Balanced and repeated replications (BRR)
 - Jackknife
 - Bootstrap
 - Successive difference replication (SDR)
- Subpopulation estimation
- Poststratification
- Raking
- Calibration
- DEFF
- MEFF



Stata analyzes data from any sampling design, whether simple or complex.

Just **svyset** it and forget it.

Simple random sample

```
. svyset _n
```

One-stage cluster design, specifying sampling weights

```
. svyset psu [pweight=pw]
```

One-stage cluster design with weights and stratification

```
. svyset psu [pweight=pw], strata(strata)
```

Two-stage design

```
. svyset psu [pweight=pw], fpc(fpc1) ||  
  _n, fpc(fpc2)
```

Two-stage design with stage-level sampling weights

```
. svyset psu, fpc(fpc1)  
  weight(pweight1) ||  
  _n, weight(pweight2)
```

BRR replicate weights

```
. svyset [pweight=pw], brrweight(brr1-brr32)
```

Specify the design just once. Then add the **svy** prefix to your command, and results are automatically adjusted to account for the sampling design.

You can account for the design when you are estimating means,
. svy: mean x

when you are estimating totals,
. svy: total x

You can also adjust for the sampling design when fitting the following:

- Logistic regression
 - Poisson regression
 - Ordered probit regression
 - Multinomial logistic regression
- Generalized linear models (GLMs)
 - Cox proportional hazards model
 - Parametric survival models
 - Instrumental-variables regression
- Selection models
 - Multilevel models
 - Structural equation models (SEMs)
 - And much more

Linear regression for the subpopulation of females

view svy1.smcl

view svy1.smcl

+

Dialog

Also see

Jump to

```
. svy, subpop(female): regress systolic_bp i.region age weight
(running regress on estimation sample)
```

Survey: Linear regression

Number of strata = 31

Number of PSUs = 62

Number of obs = 10,351

Population size = 117,157,513

Subpop. no. obs = 5,436

Subpop. size = 60,998,033

Design df = 31

F(5, 27) = 266.93

Prob > F = 0.0000

R-squared = 0.3803

systolic_bp	Coefficient	Linearized std. err.	t	P> t	[95% conf. interval]	
region						
Midwest	-.3623935	2.014345	-0.18	0.858	-4.470677	3.74589
South	-.7813662	2.123326	-0.37	0.715	-5.111919	3.549187
West	-.0837169	1.892213	-0.04	0.965	-3.942911	3.775478
age	.7584049	.0232024	32.69	0.000	.7110833	.8057265
weight	.425346	.0215081	19.78	0.000	.38148	.469212
_cons	64.29741	2.368021	27.15	0.000	59.4678	69.12702

CAP NUM INS

when you are fitting a linear regression model,
. svy: regress y x

and when you are constructing contingency tables,
. svy: tabulate x1 x2

Multistage sample, multilevel logit model

Viewer - view svy2.smcl

view svy2.smcl

Dialog

Also see

Jump to

+

```
. svy: melogit pass_read ses i.sex i.hs_grad || id_school:
(running melogit on estimation sample)
```

Survey: Mixed-effects logistic regression

Number of strata = 1

Number of PSUs = 148

Number of obs = 2,069

Population size = 346,373.74

Design df = 147

F(3, 145) = 26.60

Prob > F = 0.0000

pass_read	Coefficient	Linearized std. err.	t	P> t	[95% conf. interval]	
ses	.7580967	.0962879	7.87	0.000	.5678093	.9483841
sex Female	.6433437	.1593681	4.04	0.000	.3283952	.9582922
hs_grad Yes	-.5842494	.1751927	-3.33	0.001	-.930471	-.2380279
_cons	-1.313443	.2838087	-4.63	0.000	-1.874316	-.7525712
id_school var(_cons)	.8873707	.3117113			.4432177	1.776614

CAP NUM INS

Type or point and click

svyset - Declare survey design for dataset

Main

Weights

SE

Poststratification

Calibration

Number of stages: 2

Clear settings

Primary sampling units:

Stage 1: id_school

Stage 2: _n

Strata:

Finite pop. correction:

Sampling weight: wt1 wt2

OK Cancel Submit

regress - Linear regression

Model

by/fit/in

Weights

SE/Robust

Reporting

Survey data estimation

Standard error type: Linearized

Design degrees of freedom

OK Cancel Submit

melogit - Multilevel mixed-effects logistic regression

Model

by/fit/in

Weights

SE/Robust

Reporting

Integration

Maximization

Fixed-effects model

Dependent variable: pass_read

Independent variables: ses i.sex i.hs_grad

Offset variable:

Suppress constant term

Random-effects model

Random-effects equations: Equation 1

Create...

Edit

Disable

Enable

Options

Binomial trials per observation

Variable

Fixed number: 1

Constraints:

Manage...

Retain perfect predictor variables

OK Cancel Submit