stata

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) 	Iogistic - Logistic regression, reporting odds ratios Model by/if/in Weights SE/Robust Reporting Maximization Survey data estimation Standard error type: Inearized Bootstrap BRR Jackknife Design degrees of freedom
 Subpopulation estimation 	
 Poststratification 	Note: Survey settings must be specified.
Raking	? C 🖻 OK Cancel Submit
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

when you are fitting a linear regression model,

. svy: regress y x

and when you are constructing contingency tables,

. svy: tabulate x1 x2

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

Linear regression for the subpopulation of females

Viewer - view svy1.smcl					-		×
view svy1.smcl 🗙							
+				Dialog 👻	Also	see 👻 🛛 Ju	mp to 🕶
. svy, subpop(female) (running regress on e	: regress systoli stimation sample)	.c_bp i.re	egion age	weight			
Survey: Linear regres	sion						
Number of strata = 31 Number of PSUs = 62			Number Popula Subpop Subpop Design F(5, 2 Prob > R-squa	of obs tion size . no. obs . size df .7) . F .red	= 11 = 6 = = =	10,3 7,157,5 5,4 60,998,0 266. 0.00 0.38	951 13 36 933 31 93 90 93 90 93
systolic_bp Coeffi	Linearized cient std. err.	t	P> t	[95% co	nf.	interva	1]
region Midwest362 South781 West083	3935 2.014345 3662 2.123326 7169 1.892213	-0.18 -0.37 -0.04	0.858 0.715 0.965	-4.47067 -5.11191 -3.94291	7 9 1	3.745 3.5491 3.7754	89 .87 .78
age .758 weight .42 _cons 64.2	4049 .0232024 5346 .0215081 9741 2.368021	32.69 19.78 27.15	0.000 0.000 0.000	.711083 .3814 59.467	3 8 8	.80572 .4692 69.127	265 212 702

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 obs = 2,069 Population size = 346,373.74 Number of PSUs = 148 Design df 147 = F(3, 145) Prob > F 26.60 = 0.0000 Linearized Coefficient std. err. P>|t| [95% conf. interval] t pass read .0962879 0.000 .5678093 .9483841 ses .7580967 7.87

sex Female	.6433437	.1593681	4.04	0.000	.3283952	.9582922
hs_grad Yes _cons	5842494 -1.313443	.1751927 .2838087	-3.33 -4.63	0.001 0.000	930471 -1.874316	2380279 7525712
id_school var(_cons)	.8873707	.3117113			.4432177	1.776614

melogit - Multilevel mixed-effects logistic regression Model by/if/in Weights SE/Robust Reporting Integration Maxie 📃 svyset - Declare survey design for dataset Fixed-effects model Weights SE Poststratification Calibrati Dependent variable: Independent variables pass_read ses i.sex i.hs_grad Clear setting: Number of stages: 2 Suppress constant term Primary sampling units: Strata: Finite pop. correction: Sampling weight: Offset variable: Stage 1: id_school wt1 Stage 2: _n wt2 Random-effects model 🚍 regress - Linear regression X Random-effects equat Create. Model by/if/in Weights SE/Robust Reporting Edit Survey data estimation Disable Standard error type || id_school Bootstrap RRR Jackknife Binomial trials per observation O Var Design degrees of freedom 1 🗘 Constraints Manage.. ts" above indicates sampling of observations Retain perfect predictor variables OK Cancel Sub C ОК Cancel Survey settings. ? C ч OK Cancel Submit

Type or point and click

stata.com/survey

CAP NUM INS

And much more

Selection models

Multilevel models

Structural equation models (SEMs)

Multistage sample, multilevel logit model