

## Title

**svy estimation** — Estimation commands for survey data

## Description

Survey data analysis in Stata is essentially the same as standard data analysis. The standard syntax applies, you just need to also remember the following.

1. Use `svyset` to identify the survey design characteristics.
2. Prefix the estimation commands with “`svy:`”.

For example,

```
. use http://www.stata-press.com/data/r10/nhanes2f
. svyset psuid [pweight=finalwgt], strata(stratid)
. svy: regress zinc age age2 weight female black orace rural
```

See [SVY] `svyset` and [SVY] `svy`.

The following estimation commands support the `svy` prefix:

### Descriptive statistics

<code>mean</code>	[R] <b>mean</b> — Estimate means
<code>proportion</code>	[R] <b>proportion</b> — Estimate proportions
<code>ratio</code>	[R] <b>ratio</b> — Estimate ratios
<code>total</code>	[R] <b>total</b> — Estimate totals

### Linear regression models

<code>cnreg</code>	[R] <b>cnreg</b> — Censored-normal regression
<code>cnsreg</code>	[R] <b>cnsreg</b> — Constrained linear regression
<code>glm</code>	[R] <b>glm</b> — Generalized linear models
<code>intreg</code>	[R] <b>intreg</b> — Interval regression
<code>nl</code>	[R] <b>nl</b> — Nonlinear least-squares estimation
<code>regress</code>	[R] <b>regress</b> — Linear regression
<code>tobit</code>	[R] <b>tobit</b> — Tobit regression
<code>treatreg</code>	[R] <b>treatreg</b> — Treatment-effects model
<code>truncreg</code>	[R] <b>truncreg</b> — Truncated regression

### Survival-data regression models

<code>stcox</code>	[ST] <b>stcox</b> — Fit Cox proportional hazards model
<code>streg</code>	[ST] <b>streg</b> — Fit parametric survival models

**Binary-response regression models**

biprobit	[R] <b>biprobit</b> — Bivariate probit regression
cloglog	[R] <b>cloglog</b> — Complementary log-log regression
hetprob	[R] <b>hetprob</b> — Heteroskedastic probit model
logistic	[R] <b>logistic</b> — Logistic regression, reporting odds ratios
logit	[R] <b>logit</b> — Logistic regression, reporting coefficients
probit	[R] <b>probit</b> — Probit regression
scobit	[R] <b>scobit</b> — Skewed logistic regression

**Discrete-response regression models**

clogit	[R] <b>clogit</b> — Conditional (fixed-effects) logistic regression
mlogit	[R] <b>mlogit</b> — Multinomial (polytomous) logistic regression
mprobit	[R] <b>mprobit</b> — Multinomial probit regression
ologit	[R] <b>ologit</b> — Ordered logistic regression
oprobit	[R] <b>oprobit</b> — Ordered probit regression
slogit	[R] <b>slogit</b> — Stereotype logistic regression

**Poisson regression models**

gnbreg	Generalized negative binomial regression in [R] <b>nbreg</b>
nbreg	[R] <b>nbreg</b> — Negative binomial regression
poisson	[R] <b>poisson</b> — Poisson regression
zinb	[R] <b>zinb</b> — Zero-inflated negative binomial regression
zip	[R] <b>zip</b> — Zero-inflated Poisson regression
ztnb	[R] <b>ztnb</b> — Zero-truncated negative binomial regression
ztp	[R] <b>ztp</b> — Zero-truncated Poisson regression

**Instrumental-variables regression models**

ivprobit	[R] <b>ivprobit</b> — Probit model with endogenous regressors
ivregress	[R] <b>ivregress</b> — Single-equation instrumental-variables regression
ivtobit	[R] <b>ivtobit</b> — Tobit model with endogenous regressors

**Regression models with selection**

heckman	[R] <b>heckman</b> — Heckman selection model
heckprob	[R] <b>heckprob</b> — Probit model with sample selection

**Remarks**

Remarks are presented under the following headings:

*Overview of survey analysis in Stata*  
*Descriptive statistics*  
*Regression models*  
*Health surveys*

## Overview of survey analysis in Stata

Many Stata commands estimate the parameters of a process or population by using sample data. For example, `mean` estimates means, `ratio` estimates ratios, `regress` fits linear regression models, `poisson` fits Poisson regression models, `logistic` fits logistic regression models, etc. Some of these *estimation commands* support the `svy` prefix, that is, they may be prefixed by `svy:` to produce results appropriate for complex survey data. Whereas `poisson` is used with standard, nonsurvey data, `svy: poisson` is used with survey data. In what follows we refer to any estimation command not prefixed by `svy:` as the standard command. A standard command prefixed by `svy:` is referred to as a `svy` command.

Most standard commands (and all standard commands supported by `svy`) allow `pweights` and the `vce(cluster clustvar)` option, where *varname* corresponds to the PSU variable that you `svyset`. If your survey data exhibit only sampling weights and/or first-stage clusters, you can get by with using the standard command with `pweights` and/or `vce(cluster clustvar)`. Your parameter estimates will always be identical to those you would have obtained from the `svy` command, and the standard command uses the same robust (linearization) variance estimator as the `svy` command with a similarly `svyset` design.

Most standard commands are also fitted using maximum likelihood. When used with independently distributed, nonweighted data, the likelihood to be maximized reflects the joint probability distribution of the data given the chosen model. With complex survey data, however, this interpretation of the likelihood is no longer valid, as survey data are weighted, not independently distributed, or both. Yet for survey data, (valid) parameter estimates for a given model can be obtained using the associated likelihood function with appropriate weighting. Since the probabilistic interpretation no longer holds, the likelihood here is instead called a *pseudolikelihood*, but likelihood-ratio tests are no longer valid. See Skinner (1989, sec. 3.4.4) for a discussion of maximum pseudolikelihood estimators.

Here we highlight the other features of `svy` commands.

1. `svy` commands handle stratified sampling, but none of the standard commands do. Since stratification usually makes standard errors smaller, ignoring stratification is usually conservative. So not using `svy` with stratified sample data is not a terrible thing to do. However, to get the smallest possible “honest” standard-error estimates for stratified sampling, use `svy`.
2. `svy` commands use  $t$  statistics with  $n - L$  degrees of freedom to test the significance of coefficients, where  $n$  is the total number of sampled PSUs (clusters) and  $L$  is the number of strata in the first stage. Some of the standard commands use  $t$  statistics, but most use  $z$  statistics. If the standard command uses  $z$  statistics for its standard variance estimator, then it also uses  $z$  statistics with the robust (linearization) variance estimator. Strictly speaking,  $t$  statistics are appropriate with the robust (linearization) variance estimator; see [P] **\_robust** for the theoretical rationale. But, using  $z$  rather than  $t$  statistics yields a nontrivial difference only when there is a small number of clusters ( $< 50$ ). If a regression model command uses  $t$  statistics and the `vce(cluster clustvar)` option is specified, then the degrees of freedom used are the same as that of the `svy` command (in the absence of stratification).
3. `svy` commands produce an adjusted Wald test for the model test, and `test` can be used to produce adjusted Wald tests for other hypotheses after `svy` commands. Only unadjusted Wald tests are available if the `svy` prefix is not used. The adjustment can be important when the degrees of freedom  $n - L$  are small relative to the dimension of the test. (If the dimension is one, then the adjusted and unadjusted Wald tests are identical.) This fact along with point 2 make using the `svy` command important if the number of sampled PSUs (clusters) is small ( $< 50$ ).

4. `svy: regress` differs slightly from `regress` and `svy: ivregress` differs slightly from `ivregress` in that they use different multipliers for the variance estimator. `regress` and `ivregress` (when the `small` option is specified) use a multiplier of  $\{(N-1)/(N-k)\}\{n/(n-1)\}$ , where  $N$  is the number of observations,  $n$  is the number of clusters (PSUs), and  $k$  is the number of regressors including the constant. `svy: regress` and `svy: ivregress` use  $n/(n-1)$  instead. Thus they produce slightly different standard errors. The  $(N-1)/(N-k)$  is ad hoc and has no rigorous theoretical justification; hence, the purist `svy` commands do not use it. The `svy` commands tacitly assume that  $N \gg k$ . If  $(N-1)/(N-k)$  is not close to 1, you may be well advised to use `regress` or `ivregress` so that some punishment is inflicted on your variance estimates. Maximum likelihood estimators in Stata (e.g., `logit`) do no such adjustment but rely on the sensibilities of the analyst to ensure that  $N$  is reasonably larger than  $k$ . Thus the maximum pseudolikelihood estimators (e.g., `svy: logit`) produce the same standard errors as the corresponding maximum likelihood commands (e.g., `logit`), but  $p$ -values are slightly different because of point 2.
5. `svy` commands can produce proper estimates for subpopulations by using the `subpop()` option. Using an *if* restriction with `svy` or standard commands can yield incorrect standard error estimates for subpopulations. Often an *if* restriction will yield the same standard error as `subpop()`; most other times, the two standard errors will be slightly different; but sometimes—usually for thinly sampled subpopulations—the standard errors can be appreciably different. Hence, the `svy` command with the `subpop()` option should be used to obtain estimates for thinly sampled subpopulations. See [SVY] **subpopulation estimation** for more information.
6. `svy` commands handle zero sampling weights properly. Standard commands ignore any observation with a weight of zero. Usually, this will yield the same standard errors, but sometimes they will differ. Sampling weights of zero can arise from various postsampling adjustment procedures. If the sum of weights for one or more PSUs is zero, `svy` and standard commands will produce different standard errors, but usually this difference is very small.
7. You can `svyset iweights` and let these weights be negative. Negative sampling weights can arise from various postsampling adjustment procedures. If you want to use negative sampling weights, then you must `svyset iweights` instead of `pweights`; no standard command will allow negative sampling weights.
8. The `svy` commands compute finite population corrections (FPCs).
9. After a `svy` command, `estat effects` will compute the design effects `DEFF` and `DEFT` and the misspecification effects `MEFF` and `MEFT`.
10. `svy` commands can perform variance estimation that accounts for multiple stages of clustered sampling.
11. `svy` commands can perform variance estimation that accounts for poststratification adjustments to the sampling weights.
12. Some standard command options are not allowed with the `svy` prefix. For example, `cluster()`, `vce()`, and `weights` cannot be specified when using the `svy` prefix since `svy` is already using the clusters, variance estimation, and sampling weights identified by `svyset`. Some options are not valid with survey data, such as `beta` for standardized coefficients and `noskip` for producing optional likelihood-ratio tests. Other options are not allowed because they change how estimation results are reported (e.g., `nodisplay`, `first`, `plus`) or are not compatible with `svy`'s variance estimation methods (e.g., `irls`, `mse1`, `hc2`, `hc3`).
13. Estimation results are presented in the standard way, except that `svy` has its own table header: In addition to the sample size, model test, and  $R^2$  (if present in the output from the standard command), `svy` will also report the following information in the header:

- a. number of strata and PSUs
- b. number of poststrata, if specified to `svyset`
- c. population size estimate
- d. subpopulation sizes if the `subpop()` option was specified
- e. design degrees of freedom

## Descriptive statistics

Use `svy: mean`, `svy: ratio`, `svy: proportion`, and `svy: total` to estimate finite population and subpopulation means, ratios, proportions, and totals, respectively. You can also estimate standardized means, ratios, and proportions for survey data; see [SVY] **direct standardization**. Estimates for multiple subpopulations can be obtained using the `over()` option; see [SVY] **subpopulation estimation**.

### ► Example 1

Suppose that we need to estimate the average birthweight for the population represented by the National Maternal and Infant Health Survey (NMIHS) (Gonzalez, Krauss, and Scott 1992).

First, we gather the survey design information.

1. Primary sampling units are mothers; i.e., PSUs are individual observations—there is no separate PSU variable.
2. The `finalwgt` variable contains the sampling weights.
3. The `stratan` variable identifies strata.
4. There is no variable for the finite population correction.

Then we use `svyset` to identify the variables for sampling weights and stratification.

```
. use http://www.stata-press.com/data/r10/nmihs
. svyset [pweight=finwgt], strata(stratan)
      pweight: finwgt
           VCE: linearized
Single unit: missing
Strata 1: stratan
      SU 1: <observations>
      FPC 1: <zero>
```

Now we can use `svy: mean` to estimate the average birthweight for our population.

```
. svy: mean birthwgt
(running mean on estimation sample)

Survey: Mean estimation

Number of strata =      6           Number of obs   =   9946
Number of PSUs  =   9946           Population size = 3.9e+06
                                           Design df     =   9940
```

	Linearized		
	Mean	Std. Err.	[95% Conf. Interval]
birthwgt	3355.452	6.402741	3342.902 3368.003

From these results, we are 95% confident that the mean birthweight for our population is between 3,343 and 3,368 grams.

◀

## Regression models

As exhibited in the table at the beginning of this manual entry, many of Stata's regression model commands support the `svy` prefix. If you know how to use one of these commands with standard data, then you can also use the corresponding `svy` command with your survey data.

### ▷ Example 2

Let's model the incidence of high blood pressure with a dataset from the Second National Health and Nutrition Examination Survey (NHANES II) (McDowell et al. 1981). The survey design characteristics are already `svyset`, so we will just replay them.

```
. use http://www.stata-press.com/data/r10/nhanes2d
. svyset
      pweight: finalwgt
          VCE: linearized
Single unit: missing
Strata 1: strata
      SU 1: psu
      FPC 1: <zero>
```

Now we can use `svy: logistic` to model the incidence of high blood pressure as a function of height, weight, age, and sex (using the `female` indicator variable).

```
. svy: logistic highbp height weight age female
(running logistic on estimation sample)
```

Survey: Logistic regression

Number of strata	=	31	Number of obs	=	10351
Number of PSUs	=	62	Population size	=	1.172e+08
			Design df	=	31
			F( 4, 28)	=	178.69
			Prob > F	=	0.0000

highbp	Linearized		t	P> t	[95% Conf. Interval]	
	Odds Ratio	Std. Err.				
height	.9688567	.0056821	-5.39	0.000	.9573369	.9805151
weight	1.052489	.0032829	16.40	0.000	1.045814	1.059205
age	1.050473	.0024816	20.84	0.000	1.045424	1.055547
female	.7250086	.0641185	-3.64	0.001	.6053533	.8683151

The odds ratio for the `female` predictor is 0.73 (rounded to two decimal places), and is significantly less than 1. This finding implies that females have a lower incidence of high blood pressure than males.

Here we use the `subpop()` option to model the incidence of high blood pressure in the subpopulation identified by the `female` variable.

```
. svy, subpop(female): logistic highbp height weight age
(running logistic on estimation sample)
```

```
Survey: Logistic regression
```

```
Number of strata =      31      Number of obs   =      10351
Number of PSUs   =      62      Population size = 1.172e+08
                                          Subpop. no. of obs =      5436
                                          Subpop. size     = 60998033
                                          Design df       =      31
F( 3, 29)        =     137.05
Prob > F         =      0.0000
```

highbp	Linearized		t	P> t	[95% Conf. Interval]	
	Odds Ratio	Std. Err.				
height	.9765379	.0092443	-2.51	0.018	.957865	.9955749
weight	1.047845	.0044668	10.96	0.000	1.038774	1.056994
age	1.058105	.003541	16.88	0.000	1.050907	1.065352

Since the odds ratio for the age predictor is significantly greater than 1, we can conclude that older females are more likely to have high blood pressure than younger females.

◀

## Health surveys

There are many sources of bias when modeling the association between a disease and its risk factors (Korn, Graubard, and Midthune 1997; Korn and Graubard 1999, sec. 3.7). In cross-sectional health surveys, inference is typically restricted to the target population as it stood when the data were collected. This type of survey cannot capture the fact that participants may change their habits over time. Some health surveys collect data retrospectively, relying on the participants to recall the status of risk factors as they stood in the past. This type of survey is vulnerable to recall bias.

Longitudinal surveys collect data over time, monitoring the survey participants over several years. Although the above biases are minimized, analysts are still faced with some tough choices/situations when modeling time-to-event data. For example:

1. Time scale. When studying cancer, should we measure the time scale using the participants age or the initial date from which data were collected?
2. Time-varying covariates. Were all relevant risk factors sampled over time, or do we have only the baseline measurement?
3. Competing risks. When studying mortality, do we have the data specific to cause of death?

Binder (1983) provides the foundation for fitting most of the common parametric models using survey data. Similarly, Lin and Wei (1989) provide the foundational theory for robust inference using the proportional hazards model. Binder (1992) describes how to estimate standard errors for the proportional hazards model from survey data, and Lin (2000) provides a rigorous justification for Binder's method. Korn and Graubard (1999) discuss many aspects of model fitting using data from health surveys.

### ▶ Example 3: Cox's proportional hazards model

Suppose that we want to model the incidence of lung cancer by using three risk factors: smoking status, sex, and place of residence. Our dataset comes from a longitudinal health survey: the First

National Health and Nutrition Examination Survey (NHANES I) (Miller 1973; Engel et al. 1978) and its 1992 Epidemiologic Follow-up Study (NHEFS) (Cox et al. 1997); see the National Center for Health Statistics web site at <http://www.cdc.gov/nchs/>. We will be using data from the samples identified by NHANES I examination locations 1–65 and 66–100; thus, we will `svyset` the revised pseudo-PSU and strata variables associated with these locations. Similarly, our `pweight` variable was generated using the sampling weights for the nutrition and detailed samples for locations 1–65 and the weights for the detailed sample for locations 66–100.

```
. use http://www.stata-press.com/data/r10/nhefs
. svyset psu2 [pw=swgt2], strata(strata2)
      pweight: swgt2
          VCE: linearized
Single unit: missing
  Strata 1: strata2
      SU 1: psu2
   FPC 1: <zero>
```

The lung cancer information was taken from the 1992 NHEFS interview data. We use the participants' age for the time scale. Participants who never had lung cancer and were alive for the 1992 interview were considered censored. Participants who never had lung cancer and died before the 1992 interview were also considered censored at their age of death.

```
. stset age_lung_cancer [pw=swgt2], fail(lung_cancer)
      failure event: lung_cancer != 0 & lung_cancer < .
obs. time interval: (0, age_lung_cancer]
exit on or before: failure
      weight: [pweight=swgt2]
```

---

14407	total obs.	
5126	event time missing (age_lung_cancer>=.)	PROBABLE ERROR

---

9281	obs. remaining, representing	
83	failures in single record/single failure data	
599691	total analysis time at risk, at risk from t =	0
	earliest observed entry t =	0
	last observed exit t =	97

Although `stset` warns us that it is a “probable error” to have 5,126 observations with missing event times, we can verify from the 1992 NHEFS documentation that there were indeed 9,281 participants with complete information.

For our proportional hazards model, we pulled the risk factor information from the NHANES I and 1992 NHEFS datasets. Smoking status was taken from the 1992 NHEFS interview data, but we filled in all but 132 missing values by using the general medical history supplement data in NHANES I. Smoking status is represented by separate indicator variables for former smokers and current smokers; the base comparison group is nonsmokers. Sex was determined using the 1992 NHEFS vitality data and is represented by an indicator variable for males. Place-of-residence information was taken from the medical history questionnaire in NHANES I and is represented by separate indicator variables for rural and heavily populated (more than 1 million people) urban residences; the base comparison group is urban residences with populations of fewer than 1 million people.

```
. svy: stcox former_smoker smoker male urban1 rural
(running stcox on estimation sample)

Survey: Cox regression
Number of strata =      35      Number of obs =      9149
Number of PSUs  =     105      Population size = 1.513e+08
                                     Design df =      70
                                     F( 5, 66) =     14.07
                                     Prob > F =     0.0000
```

_t	Haz. Ratio	Linearized	t	P> t	[95% Conf. Interval]	
		Std. Err.				
former_smo~r	2.788113	.6205102	4.61	0.000	1.788705	4.345923
smoker	7.849483	2.593249	6.24	0.000	4.061457	15.17051
male	1.187611	.3445315	0.59	0.555	.6658757	2.118142
urban1	.8035074	.3285144	-0.54	0.594	.3555123	1.816039
rural	1.581674	.5281859	1.37	0.174	.8125799	3.078702

From the above results, we can see that both former and current smokers have a significantly higher risk for developing lung cancer than that of nonsmokers.

◀

#### □ Technical Note

In the previous example, we specified a sampling weight variable in the calls to both `svyset` and `stset`. When the `svy` prefix is used with `stcox` and `streg`, it identifies the sampling weight variable by using the data characteristics from both `svyset` and `stset`. `svy` will report an error if the `svyset` `pweight` variable is different from the `stset` `pweight` variable. The `svy` prefix will use the specified `pweight` variable, even if it is `svyset` but not `stset`. If a `pweight` variable is `stset` but not `svyset`, `svy` will note that it will be using the `stset` `pweight` variable and then `svyset` it.

The standard `st` commands will not use the `svyset` `pweight` variable if it is not also `stset`.

□

#### ▷ Example 4: Multiple baseline hazards

We can assess the proportional-hazards assumption across the observed race categories for the model fitted in the previous example. The race information in our 1992 NHEFS dataset is contained in the `revised_race` variable. We will use `stphplot` to produce a log-log plot for each category of `revised_race`. As described in [ST] **stcox diagnostics**, if the plotted lines are reasonably parallel, the proportional-hazards assumption has not been violated. We will use the `zero` option to reset the risk factors to their base comparison group.

(Continued on next page)

```
. sthplot, strata(revised_race) adjust(former_smoker smoker male urban1 rural) zero
> legend(col(1))
      failure_d: lung_cancer
analysis time _t: age_lung_cancer
      weight: [pweight=swgt2]
```

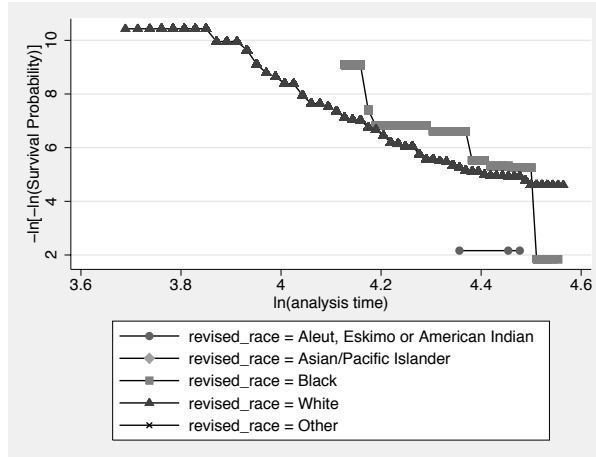


Figure 1.

As we can see from figure 1, the lines for the black and white race categories intersect. This indicates a violation of the proportional-hazards assumption, so we should consider using separate baseline hazard functions for each race category in our model fit. We do this in the following by specifying `strata(revised_race)` in our call to `svy: stcox`.

```
. svy: stcox former_smoker smoker male urban1 rural, strata(revised_race)
(running stcox on estimation sample)
```

Survey: Cox regression

Number of strata	=	35	Number of obs	=	9149
Number of PSUs	=	105	Population size	=	1.513e+08
			Design df	=	70
			F( 5, 66)	=	13.95
			Prob > F	=	0.0000

_t	Haz. Ratio	Linearized		t	P> t	[95% Conf. Interval]	
		Std. Err.					
former_smo~r	2.801797	.6280352	4.60	0.000	1.791761	4.381201	
smoker	7.954921	2.640022	6.25	0.000	4.103709	15.42038	
male	1.165724	.3390339	0.53	0.600	.6526527	2.082139	
urban1	.784031	.3120525	-0.61	0.543	.3544764	1.73412	
rural	1.490269	.5048569	1.18	0.243	.7582848	2.928851	

Stratified by revised\_race

## References

- Binder, D. A. 1983. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review* 51: 279–292.
- . 1992. Fitting Cox’s proportional hazards models for survey data. *Biometrika* 79: 139–147.
- Cox, C. S., M. E. Mussolino, S. T. Rothwell, M. A. Lane, C. D. Golden, J. H. Madans, and J. J. Feldman. 1997. Plan and operation of the NHANES I Epidemiologic Followup Study, 1992. In *Vital and Health Statistics*, vol. 1. Hyattsville, MD: National Center for Health Statistics.
- Engel, A., R. S. Murphy, K. Maurer, and E. Collins. 1978. Plan and operation of the HANES I augmentation survey of adults 25-74 years. In *Vital and Health Statistics*, vol. 1. Hyattsville, MD: National Center for Health Statistics.
- Gonzalez, J. F., Jr., N. Krauss, and C. Scott. 1992. Estimation in the 1988 National Maternal and Infant Health Survey. *Proceedings of the Section on Statistics Education, American Statistical Association* 343–348.
- Korn, E. L., and B. I. Graubard. 1999. *Analysis of Health Surveys*. New York: Wiley.
- Korn, E. L., B. I. Graubard, and D. Midthune. 1997. Time-to-event analysis of longitudinal follow-up of a survey: Choice of time-scale. *American Journal of Epidemiology* 145: 72–80.
- Lin, D. Y. 2000. On fitting Cox’s proportional hazards models to survey data. *Biometrika* 87: 37–47.
- Lin, D. Y., and L. J. Wei. 1989. The robust inference of the Cox proportional hazards model. *Journal of the American Statistical Association* 84: 1074–1078.
- McDowell, A., A. Engel, J. T. Massey, and K. Maurer. 1981. Plan and operation of the Second National Health and Nutrition Examination Survey, 1976–1980. In *Vital and Health Statistics*, vol. 1. Hyattsville, MD: National Center for Health Statistics.
- Miller, H. W. 1973. Plan and operation of the Health and Nutrition Examination Survey: United States 1971–1973. In *Vital and Health Statistics*, vol. 1. Hyattsville, MD: National Center for Health Statistics.
- Skinner, C. J. 1989. Introduction to part A. In *Analysis of Complex Surveys*, ed. C. J. Skinner, D. Holt, and T. M. F. Smith, 23–58. New York: Wiley.

## Also See

- [SVY] **svy postestimation** — Postestimation tools for svy
- [SVY] **estat** — Postestimation statistics for survey data
- [SVY] **direct standardization** — Direct standardization of means, proportions, and ratios
- [SVY] **poststratification** — Poststratification for survey data
- [SVY] **subpopulation estimation** — Subpopulation estimation for survey data
- [SVY] **variance estimation** — Variance estimation for survey data
- [U] **20 Estimation and postestimation commands**
- [SVY] **svyset** — Declare survey design for dataset
- [SVY] **svy** — The survey prefix command