

intro 10 — Fitting models with survey data

[Description](#)[Remarks and examples](#)[Also see](#)

Description

Sometimes the data are not a simple random sample from the underlying population but instead are based on a complex survey design that can include stages of clustered sampling and stratification. Estimates produced by `sem` can be adjusted for these issues.

Adjustments for survey data are provided by `sem` and `gsem`.

Remarks and examples

[stata.com](#)

Data obtained from surveys, properly treated, produce different point estimates because some observations represent a greater proportion of the underlying population than others. They also produce different standard errors because the observation-to-observation (sample-to-sample) variation is a function of the survey's design.

To obtain survey-corrected results, you first describe the characteristics of the survey with `svyset`:

```
. svyset county [pw=samplewt], fpc(n_counties) strata(states) || ///
      school, fpc(n_schools)      || ///
      student, fpc(n_students)
```

In the above, we are telling Stata that our data are from a three-stage sampling design. The first stage samples without replacement counties within state; the second, schools within each sampled county; and the third, students within schools.

Once we have done that, we can tell Stata to make the survey adjustment by prefixing statistical commands with the `svy:` prefix:

```
. svy: regress test_result teachers_per_student sex ...
```

Point estimates and standard errors will be adjusted.

You can use the `svy:` prefix with `sem` and `gsem`:

```
. svy: sem (test_result<-...) ... (teachers_per_student->...) ...
```

See the *Stata Survey Data Reference Manual* for more information on this. From a survey perspective, `sem` is not different from any other statistical command of Stata. When `gsem` is used to fit a multilevel model, stage-level sampling weights specified in the `svyset` command are applied to the corresponding hierarchical group level in the model.

Once results are estimated, you do not include the `svy:` prefix in front of the postestimation commands. You type, for instance,

```
. estat eqtest ...
```

You do not type `svy: estat eqtest ...`.

Some postestimation procedures you might ordinarily perform can be inappropriate with survey estimation results. This is because you no longer have a sample likelihood value. The postestimation command `lrtest` is an example. If you attempt to use an inappropriate postestimation command, you will be warned.

```
. lrtest ...  
lrtest is not appropriate with survey estimation results  
r(322);
```

Also see

[SEM] [intro 9](#) — Standard errors, the full story

[SEM] [intro 11](#) — Fitting models with summary statistics data (sem only)

[SVY] [Stata Survey Data Reference Manual](#)