

## How do I run analyses on subpopulations in STATA?

There are times when you may want to run analyses on a subset of the population (e.g. females; Hispanic males; teen mothers with non-missing values on analysis variables).

### If you are **NOT** using survey commands in STATA:

1. Define your subpopulation, e.g.:
  - *female*: [1 = female] [0 = male]
  - *hisp*: [1 = Hispanic males] [0 = all others]
  - *teenm*: [1 = teen mothers with non-missing values on analysis variables] [0 = all others]
    - Note: You do **not** have to use a dummy variable if you are not using survey commands. For example, you could also choose to run analyses for people who have a value of 2 on a variable that ranges from 1 to 5. (See means example below.)
2. Run analyses:
  - means
    - mean *var1* if *female*==1
    - mean *var1* if *var2*==2
  - OLS regression
    - regress *dv iv1 iv2 iv3* if *hisp*==1
  - Logistic regression
    - logistic *dv iv1 iv2 iv3* if *teenm*==1
    - logit *dv iv1 iv2 iv3* if *teenm*==1

### If you are using survey commands in STATA 9:

1. Define your subpopulation, e.g.:
  - *female*: [1 = female] [0 = male]
  - *hisp*: [1 = Hispanic males] [0 = all others]
  - *teenm*: [1 = teen mothers with non-missing values on analysis variables] [0 = all others]
2. Set your survey weights
  - Stata 9: svyset *psu\_variable* [pweight=*weightvar*], strata (*stratumvar*)
3. Run analyses:
  - means
    - svy, subpop(*female*): mean *var1*
  - OLS regression
    - svy, subpop(*hisp*): regress *dv iv1 iv2 iv3*

- Logistic regression
  - svy, subpop (teenm): `logistic dv iv1 iv2 iv3`
  - svy, subpop (teenm): `logit dv iv1 iv2 iv3`