

When working with data, it is often needed to use some of the observations in a data set and exclude the rest. The most common way to do this is using a **subsetting IF statement** in a DATA step. The basic form of a subsetting IF is

IF expression; e.g. **IF Gender = 'f';**

It is a special case of the standard IF-THEN statement. If the expression is true, then SAS continues with the DATA step. If the expression is false, then no further statements are processed for that observation, and that observation is not added to the data set being created; and SAS moves on to the next observation.

You can think of the subsetting IF as a kind of on-off switch. If the condition is true, then the switch is on and the observation is processed. If the condition is false, then that observation is turned off.



There is another alternative to subset your data, the **DELETE statement**. DELETE statements do the opposite of subsetting IFs. While the subsetting IF statement tells SAS which observations to include, the DELETE statement tells SAS which observations to exclude. The basic form is:

IF *expression* THEN DELETE;

The following two statements are equivalent (assuming there are only two values for the variable Gender, and no missing data):

IF Gender = 'f'; IF Gender = 'm' THEN DELETE;

Generally, you use the subsetting IF when it is easier to specify a condition for including observations, and use the DELETE statement when it is easier to specify a condition for excluding observations.

