

IF vs. WHERE

So far, you've learned that you can use IF conditional logic and the WHERE statement for sub-setting. But what's the difference? This might have been obvious for some of you who have extremely good attention to detail but it is not so obvious.

I have found that online resources tend to explain the differences in a fairly convoluted/confusing manner. So I'll give you a list:

When to use ONLY IF:

1. Whenever you are accessing raw data using **INPUT** statement.
2. Whenever you are using automatic variables like **FIRST.BY** or **LAST.BY**.
3. Whenever you are using **newly created variables** in a data-set.
4. Whenever you use it in combination with data set options such as **OBS** or **FIRSTOBS**.
5. Whenever you want to **conditionally** execute a statement.
6. In general, use IF statement when merging data sets to apply subset condition ***after merging data set.***

When to use ONLY WHERE:

1. Whenever you are directly using a SAS Procedure (ie Proc print)

2. Whenever using PROC SQL.

3. When sub-setting as a data set option.

4. Whenever using special operators (ie LIKE OR CONTAINS)
5. If you want to utilize an index.
6. Interested in more efficiency (assuming other points are also followed)

NO DIFFERENCE

1. When using variables in data-set using SET, MERGE, OR UPDATE.
2. When using variables in data-set using SET, MERGE, or UPDATE in data-step.

