

Kirk's Korner



Quick & Simple Tips

Kirk Paul Lafler, Software Intelligence Corporation

Methods of *Creating New Data Sets with a Subset*

A new SAS data set (or table) can be created from one or more external files or existing SAS data sets. When a data set is created this way, it contains a subset of the original external file or data set. Subsets generally contain fewer observations than the original external file or SAS data set they are derived from. The following examples illustrate popular techniques for subsetting SAS data sets.

Subsetting records from an external file with a subsetting IF statement

Records from an external file can be subset by specifying a subsetting IF statement. The subsetting IF first reads a record from an external file into memory and then determines whether the condition is true. It cannot be used in a PROC step. In the following example, a subset of patients 65 years or older are written to the SENIORS data set.

```
DATA SENIORS;
  INFILE "Patients.dat" MISSOVER;
  INPUT @22 AGE 3. @;
  IF AGE > 64;
  INPUT @1 NAME $20.
        @26 GENDER $1.;
RUN;
```

Subsetting observations in a DATA step with a WHERE statement

Observations from an existing SAS data set can be subset with a WHERE statement in a DATA step. This approach specifies a WHERE statement instead of a subsetting IF statement to select observations. It works by first evaluating whether the specified WHERE condition is true before an observation is read into the Program Data Vector (PDV) for processing. Any observations not meeting the WHERE condition are automatically eliminated from being read into the PDV. In the next example a subset of patients 65 years or older are written to the SENIORS data set.

```
DATA SENIORS;
  SET PATIENTS;
  WHERE AGE > 64;
RUN;
```

Subsetting observations in a PROC step with a WHERE statement

Observations can also be subset with a WHERE statement in a PROC step. Essentially a WHERE statement works the same way as when specified in a DATA step, except without the creation of a new data set subset. In the next example a subset of patients 65 years or older are selected for output in a PROC PRINT.

```
PROC PRINT
  DATA=PATIENTS
  NOOBS
  N;
  WHERE AGE > 64;
RUN;
```

Subsetting observations in PROC SQL

Observations can also be subset in PROC SQL with a WHERE clause. Essentially a WHERE clause can be specified with or without a CREATE TABLE statement. When used with a CREATE TABLE statement a new table subset is created which can be used for future use. As presented earlier, a WHERE clause first evaluates whether the specified condition is true before an observation is read into the Program Data Vector (PDV) for processing. In the next example PROC SQL is used to subset patients 65 years or older with a WHERE clause and are written to the SENIORS table.

```
PROC SQL;  
  CREATE TABLE SENIORS AS  
  SELECT *  
  FROM PATIENTS  
  WHERE AGE > 64;  
QUIT;
```

If you would like more information or have any questions about these tips, please contact: Kirk Lafler, Software Intelligence Corporation at KirkLafler@cs.com. Kirk's new book, Power SAS: A Survival Guide is published by Apress and provides SAS users with the largest and most comprehensive collection of SAS tips and techniques ever offered. Power SAS can be purchased on Amazon.com and BarnesandNoble.com.

SAS is the registered trademark of SAS Institute Inc., Cary, NC and other countries.