

A Gentle Introduction to Using SAS (Part VI)¹
Statistics 101
Spring 1999

In this sixth installment of our introduction to SAS, we will introduce methods for creating subsets of a given data set by putting logical conditions on one or more of its variables.

DATA SET:

For this demonstration we will use the following data set. The columns represent the names, weights (in pounds), heights (in inches) and gender for twelve fictitious individuals.

```
Joe 180 72 M
Tim 225 68 M
Moe 300 70 M
Jim 184 79 M
Al 190 61 M
Ty 142 67 M
Jill 125 51 F
Ruth 110 60 F
Anna 119 59 F
Kate 127 55 F
Mary 140 70 F
Ada 100 53 F
```

OBJECTIVE:

Our objective is to write a SAS code that carries out following tasks:

1. Reads in the data from this file and stores it to a data set called *test*.
2. Creates a subset data set, call it *subset1*, which contains only the females.
3. Creates a subset data set, call it *subset2*, which contains all individuals who weigh at least 120 pounds.
4. Creates a subset data set, call it *subset3*, which contains all individuals who weigh at least 120 pounds or are at most 50 inches tall. For this subset, we want to drop the only keep the *height* and *gender* variables.
5. Creates a subset data set, call it *subset4*, which contains all males who weigh more than 160 pounds. For this subset, we want to only keep the *height* and *gender* variables.
6. Creates a subset data set, call it *subset5*, which contains no males. Note that this is subset is identical to *subset1*.

¹ Please direct any comments or suggestion to Kouros Owzar (owzar@email.unc.edu).

SAMPLE SAS CODE:

```
DATA test;
  INFILE 'a:\try.dat';
  INPUT name $ weight height gender $;
RUN;QUIT;

DATA subset1;
  SET test;
  IF (gender='F');
RUN;QUIT;

DATA subset2;
  SET test;
  IF (weight => 120);
RUN;QUIT;

DATA subset3;
  SET test;
  IF (weight => 120 OR height <= 50);
  DROP weight gender;
RUN;QUIT;

DATA subset4;
  SET test;
  IF (gender = 'M' AND weight => 160);
  KEEP weight gender;
RUN;QUIT;

DATA subset5;
  SET test;
  IF (gender ^= 'M');
RUN;QUIT;
```

THE MAIN TASKS:

1. Type the above SAS code in to the PROGRAM EDITOR window (make sure that you have downloaded try.dat).
2. Add appropriate SAS PROC PRINT commands to the above code (see first tutorial for notes on this command), to print out the six subsets along with the original data set.
3. Create the subsets manually and compare them to the corresponding subsets produced by SAS.

FINAL NOTE:

The comparison operators, =, ^= , >= , <=, > and < can be replaced by the SAS keywords, EQ, NE, GE, LE, GT and LT respectively.