
SAS-Programming – 8 (Subsetting & Combining Data Sets)

1. Starting with the Sashelp data set Fish, create a data set called Small_P perch that contains only perch that weigh less than 50 (whatever the weigh units are). Do this using a WHERE statement.
2. Repeat Problem 1 using a WHERE= data set option.
3. You are reading raw data from a data set created by Program for Problem Sets 1 (listed below). Use a subsetting IF statement to include only those subjects where the sum of Q1–Q3 is greater than or equal to 6.

```
data Questionnaire;
    informat Gender 1. Q1-Q4 $1. Visit date9.;
    input Gender Q1-Q4 Visit Age;
    format Visit date9.;
datalines;
1 3 4 1 2 29May2015 16
1 5 5 4 3 01Sep2015 25
2 2 2 1 3 04Jul2014 45
2 3 3 3 4 07Feb2015 65
;
```

4. Starting with the Sashelp data set Cars, create two temporary data sets. The first one (Cheap) should include all the observations from Cars where the MSRP (manufacturer's suggested retail price) is less than or equal to \$11,000. The other (Expensive) should include all the observations from Cars where the MSRP is greater than or equal to \$100,000. Use a KEEP= data set option to include only the variables Male, Type, Origin, and MSRP from the Cars data set. Be sure to create these two data sets in one DATA step. Use PROC PRINT to list the observations in Cheap and Expensive. Even though there are no missing values for the variable MSRP, write your program so that any observation with a missing value for MSRP will not be written to data set Cheap.
5. Run Program for Problem Sets 6 to create two data sets (FirstQtr and SecondQtr). Then create a new data set (FirstHalf) that contains all the observations from FirstQtr and SecondQtr.

Program for Problem Sets 1

```
data FirstQtr;
    input Name $ Quantity Cost;
datalines;
Fred 100 3000
Jane 90 4000
April 120 5000
;
data SecondQtr;
```

```

        input Name $ Quantity Cost;
datalines;
Ron 200 9000
Jan 210 9500
Steve 177 5400
;

```

6. Repeat Problem 5, except use PROC APPEND to combine the observations from data sets FirstQtr and SecondQtr. Because you want the resulting data set to be called FirstHalf, you will first need to make a copy of FirstQtr that is called First_Half.
7. Run Program for Problem Sets 7. Then create a new data set (Both) that contains ID, X, Y, Z, and Name. Include only those IDs that are in both data sets.

Program for Problem Sets 2

```

data First;
    input ID $ X Y Z;
datalines;
001 1 2 3
004 3 4 5
002 5 7 8
006 8 9 6
;
data Second;
    input ID $ Name $;
datalines;
02 Jim
03 Fred
001 Susan
004 Jane
;

```

8. Repeat Problem 7, except include all observations from each data set, even if there is no corresponding ID in one of the files.
9. Run Program for Problem Sets 8. Write a program to create a new data set called New_Prices, where the price of item X200 is \$410 and the price of item A123 is \$121. Caution: Item_Number is a character variable and the data set is not sorted by Item_Number.

Program for Problem Sets 3

```

data Prices;
    input Item_Number $ Price;
datalines;
A123 $123
B76 4.56
X200 400
D88 39.75
;

```