Lesson 11: Combining SAS Data Sets Horizontally

Summary

Main Points

Overview of Combining SAS Data Sets Horizontally

- When you combine data sets horizontally, you combine the observations from multiple data sets into a single observation in a new data set.

- To combine data sets horizontally, it’s important to understand the relationship between the input data sets. The relationships between input data sets can be one-to-one, one-to-many, many-to-one, many-to-many, or non-matching.

- Merging combines observations from two or more data sets into a single observation in a new data set. Match-merging is merging that is based on the values of one or more common variables.

Merging SAS Data Sets One to One

```
DATA SAS-data-set;
  MERGE SAS-data-set1 SAS-data-set2 ... ;
  BY <DESCENDING> BY-variable(s);
  <additional SAS statements>
RUN;
```

- To match-merge using the DATA step, you use the MERGE and BY statements. The MERGE statement joins observations from two or more SAS data sets into single observations. The BY variables must be common to all input data sets, and the input data sets must be sorted on the BY variables.

- If you specify only one data set in the MERGE statement, SAS treats the MERGE statement like a SET statement.

- When you match-merge data sets that have a one-to-one relationship, the output data set has the same number of observations as the input data sets.

Merging SAS Data Sets One to Many

- To match-merge data sets that have one-to-many or many-to-one relationships, you use a DATA step with the MERGE and BY statements. The DATA step is the same as the one you use to match-merge data sets that have a one-to-one relationship.

- A BY group is a group of observations that have the same value of the BY variable. When the input data sets have a one-to-many or many-to-one relationship, each BY group might contain more than one observation.

- The output data set contains all of the observations for each BY group.
Lesson 11: Combining SAS Data Sets Horizontally

Merging SAS Data Sets That Have Non-Matches

To match-merge data sets that have non-matches, you use the DATA step with the MERGE and BY statements. The DATA step is the same as the one you use for match-merging data sets with other relationships.

The output data set contains both matches (observations that contain data from both input data sets) and non-matches (observations that contain data from only one input data set).

When you specify the IN= data set option after an input data set in the MERGE statement, SAS creates a temporary variable that indicates whether the data set contributed data to each output observation. You can specify the IN= option for multiple input data sets in one MERGE statement.

To create an output data set that contains only matches or only non-matches, you can use the IN= data set option to identify which input data sets contributed data to each observation that SAS outputs. Then, you can use the subsetting IF statement to output only those observations that contain data from all the input data sets.

Merging SAS Data Sets Many to Many

You can match-merge data sets that have a many-to-many relationship in two ways: using the DATA step techniques that you’ve already learned, or using the SQL procedure.

When you use a DATA step with MERGE and BY statements to match-merge data sets that have a many-to-many relationship, the output data set does not contain all possible combinations.

When you use PROC SQL to join data sets that have a many-to-many relationship, the output data set does contain all possible combinations of matching observations.

Sample Code

Match-Merging Data Sets

data empsauc;
   merge empsau
       phonec;
   by EmpID;
run;
Lesson 11: Combining SAS Data Sets Horizontally

Match-Merging Data Sets and Selecting Observations by Which Data Set Contributed

```
data empsauc;
    merge empsau(in=Emps)
        phonec(in=Cell);
    by EmpID;
    if Emps=0 or Cell=0;
run;
```