Lesson 10: Combining SAS Data Sets Vertically

Summary

Main Points

Overview of Combining SAS Data Sets Vertically

- When you combine data sets vertically, you place the observations from one or more data sets below or above the observations in another data set. This lesson covers three methods of combining data vertically: appending, concatenating, and interleaving.

- It’s important to understand the structure and contents of your input data sets to decide which method of combining to use. You can use PROC CONTENTS or the Explorer window to look at the descriptor portion of your data sets. You can use PROC PRINT or the VIEWTABLE window to view the data portion of your data sets.

- It’s important to consider whether the data sets that you want to combine have variables in common. It’s easier to combine data sets that have common variables, but it is also possible to combine data sets that have different variables.

- Appending and concatenating data sets both place observations from one data set after the observations from another data set in a single data set. In both methods, the observations in the combined data set are in the same order that they were in each original data set.

- When you append data sets, you can only use two input data sets. Appending adds all of the observations from the second data set to the end of the first data set.

- Concatenating copies all of the observations from the first data set and all of the observations from the second data set and writes them to a new data set.

- Interleaving intersperses the observations from two or more input data sets based on the value of one or more common variables, in a new data set.

Appending SAS Data Sets

```
PROC APPEND BASE=SAS-data-set
       DATA=SAS-data-set <FORCE>;
RUN;
```

- PROC APPEND adds the observations from one SAS data set to the end of another SAS data set. BASE= names the data set to which the observations are added, and DATA= names the data set containing observations that are added to the base data set. In the PROC APPEND step, you can specify only two data sets.

- When SAS processes a PROC APPEND step, SAS does not read the observations in the base data set. Also, SAS cannot change any variable information in the descriptor portion of the base data set. The base data set has the same number of variables before and after appending.
Lesson 10: Combining SAS Data Sets Vertically

- When the DATA= data set contains variables that are not in the BASE= data set, you can use the FORCE option to force SAS to append the observations. The FORCE options causes SAS to drop the extra variables in the DATA= data set, and then to issue a warning message to the log.

Concatenating SAS Data Sets

```
DATA SAS-data-set;
  SET SAS-data-set1 SAS-data-set2 …;
  <additional SAS statements>
RUN;
```

SAS-data-set (**RENAME**=(old-name-1 = new-name-1
old-name-2 = new-name-2
…
old-name-n = new-name-n))

- You can use a DATA step to concatenate data sets. When you specify multiple data sets in the SET statement, SAS combines them into a single data set. In the combined data set, the observations appear in the order in which the data sets are listed in the SET statement.

- You can add additional DATA step statements, such as an assignment statement to create new variables in the output data set.

- Because the DATA step creates a new data set, the input data sets can contain different variables. If the data sets specified in the SET statement have a variable with the same name but different types, SAS generates a compile-time error by default.

- You can use the **RENAME**= option to change the name of one or more variables. If the **RENAME**= option is associated with an input data set in the SET statement, it renames one or more variables being read from that data set. However, the **RENAME**= option does not rename variables in the input data set. Instead, the **RENAME**= option tells SAS which slot in the PDV to use when SAS is building observations from the new data set.

- You can use multiple **RENAME**= options in one SET statement if you want to rename variables from multiple data sets.
Lesson 10: Combining SAS Data Sets Vertically

Interleaving SAS Data Sets

To interleave data sets, you use a DATA step with a BY statement. The BY statement specifies one or more BY variables. SAS uses the values of the BY variables to arrange the observations in the output data set. Any BY variable that you specify must be common to all input data sets.

By default, SAS arranges observations in ascending order unless you specify the DESCENDING option.

Before you run a DATA step that contains a BY statement, your input data sets must be sorted on the BY variables.

SAS outputs observations with duplicate BY values in the order in which the data sets are listed in the SET statement. If the duplicate values are in the same data set, SAS outputs those observations in the order in which they appear in the data set.

You could also interleave data sets by concatenating them first, and then sorting the output file.

Sample Code

Appending Data Sets

```sas
proc append base=emps
data=emps2010 force;
run;
```

Concatenating Data Sets

```sas
data empsall2;
set empscn empsjp(rename=(Region=Country));
run;
```
Lesson 10: Combining SAS Data Sets Vertically

Interleaving Data Sets

```sas
data empsname;
  set empscn empsjp(rename=(Region=Country));
  by First;
run;
```