Print

Summary: Lesson 4 - Creating Simple Queries

Topic Summaries

SAS Enterprise Guide Tools for Querying Data

- There are three tools in SAS Enterprise Guide that enable you to query and manipulate your data.
- In the Filter and Sort task, you can select variables, filter rows, and sort your data. The query results are stored in a new SAS data set.
- The Query Builder enables you to select variables, filter rows, and sort your data. In addition, the Query Builder enables you to compute new columns, join tables, group data, summarize data, and modify column attributes. You can specify the name and storage location of the output data, and you can create SAS views and reports as output instead of a SAS data set.
- The Data Explorer enables you to create different views of data in a SAS library by specifying columns to include, creating filters, and specifying sorting criteria. You can also quickly calculate basic statistics. After you explore your data you can add it to your project.

Filtering and Sorting Data

- In the Filter and Sort task, the **Variables**, **Filter**, and **Sort** tabs enable you to do the following:
  - select the variables from the input data set that you want in your output data set
  - arrange the variables in the order that you want them in the output data
  - build simple filter expressions using variables names, operators, and data values
  - sort by multiple variables
  - specify a sort sequence
- The Query Builder provides similar tabs for selecting columns, filtering rows, and sorting data. The **Select Data** tab is similar to the **Variables** tab in the Filter and Sort task. The **Filter Data** and **Sort Data** tabs in the Query Builder and the Filter and Sort task provide the same functionality.

Creating New Columns with an Expression

- In the Query Builder, the New Computed Column Wizard enables you to do the following:
  - choose from three types of computed column: summarized columns, recoded columns, or columns based on an expression
  - build expressions based on variables, operators, and functions
  - specify a name, alias, length, type, and format
- You can use functions to create new columns. A SAS function is a special routine that returns a value from arguments that you supply. Arguments can be constants, variables, expressions, or other functions. Many functions ignore missing values in calculations. It's important to understand that missing values are **ignored**, not treated as zero values.
- Computed columns are automatically added to your selected columns list on the **Select Data** tab.

Grouping and Summarizing Data

- After a column is added to the **Select Data** tab, you can select a summary statistic from the drop-down list in the **Summary** column. You can do this for numeric or character columns. By default, the new column name includes the statistic and the original column name. The new summarized column takes the place of the original column in the **Select Data** tab.
- The **Automatically select groups** check box is selected by default on the **Select Data** tab. You can select your own variable for grouping by clearing the check box and selecting the group variable.
- When data is summarized, you can filter the raw data or the summarized data or both. If filters are defined for both, the following happens:
  - the query extracts all rows that satisfy the conditions specified in the raw data filter
  - the resulting data is grouped by the assigned grouping variable, and summary statistics are computed for each unique group
  - the filters defined for summarized data are applied to the summarized values

When you filter grouped data, you build an SQL HAVING clause, which is applied only after grouping occurs.

Joining Tables

- The Query Builder supports two types of joins: inner joins and outer joins. These two types of joins differ based on whether they include nonmatching rows. The type of join determines which rows are included in the output table.
- An **inner join** creates an output table that contains only the matching rows in the input tables. This is the default type of join in the Query Builder.
- An **outer join** includes some or all of the nonmatching rows from one or both tables. There are three types of outer joins:

    - A **full outer join** returns all matching rows plus nonmatching rows from both tables.
    - A **left outer join** returns all matching rows plus nonmatching rows from the first, or left, table.
    - A **right outer join** returns all matching rows plus nonmatching rows from the second, or right, table.
- To join tables in the Query Builder, you add the tables to your query from your SAS Enterprise Guide project, from your local computer, from a SAS server, or from a SAS folder. You can join data from up to 32 different tables.
- The Query Builder looks for columns that have the same column name and type. If a name and type match is found, the table is automatically joined in the query.
- In the Join Properties window, you can specify the join type. The condition in the bottom of the window shows the operator used for comparing the values in the columns. The default is an equal sign, but other choices are also available.
- When you join multiple tables in the Query Builder, all the columns from the joined tables are available to use for other data manipulation tasks.
- One thing to consider when joining multiple tables is where the SQL code that performs the query will run. This is only an issue if you have more than one SAS server available in your SAS Enterprise Guide configuration. By default, the query executes on the server where the first table that was added to the query resides. Any data in the query that does not reside on that server is automatically copied to the server before the query runs. Typically, you want to run the query on the server where the largest table in the query resides. This minimizes the amount of data that is copied from one server to another. In the Query Options window, you can specify the server for the query.

Close