Three Approaches of Developing Summary Software in SAS

Frances Wilson
Texas Instruments Incorporated

Summaries of "MEAN" statistics over different combinations of BY variables are often needed in report processing and generating. SAS is an excellent language for calculating such statistics; however, significant manipulation of data often becomes necessary to develop a dataset in the form needed for actual printing of the report. Especially, when reports must be presented in a form consistent with pre-SAS generated forms. Three major methods of writing summary software were identified in attempts to better handle the data manipulation. The purpose of this paper is to discuss the usage, advantages and disadvantages of these three approaches.

Approach number one is a straightforward approach to the problem. The programmer uses multiple PROC MEANS with appropriate BY variables to generate the statistics. Then he must perform multiple SET's and MERGE's until he has a dataset that he can use to produce the desired output.

Approach number two involves a subtle trick to the SAS MEANS. Records are output for all desired combinations of BY variables using dummy values for summary combinations. Then, a single PROC MEANS with only one set of BY variables need be executed. The result is generally a dataset that is usable for output generation since the dummy values are assigned with the report form taken into consideration.

Approach number three is the usage of the PROC SUMMARY developed for SAS 79 in place of the PROC MEANS. One SUMMARY can replace multiple MEANS. The PROC SUMMARY provides a way to perform the calculations in one step without even having to sort by the variables.

All of the above approaches to summary software have good application. Time, cost and software readability are all important considerations in their use.

Example problem: given dataset with variables: week, month, season, year number; provide totals and means for each time period for the variable number.

**APPRAOH I**

**STRAIGHTFORWARD APPROACH**

In this approach a "PROC MEANS" would be used for each of the combinations desired.

**MEANS "BY"**

1. Week
2. Month
3. Season
4. Year

Should the information be required in a single table, then significant manipulation of the four datasets would become necessary, including several sorts. The more types of "BY" combinations needed the more complicated the manipulation. The advantages of SAS can become obscured with this approach.

**APPRAOH II**

**DUMMY RECORDS—SINGLE PROC MEANS**

Artificial values for "BY" variables are set up to represent the different combinations desired. For each original record in the dataset there are now four. One for each desired combination. A "PROC MEANS" will then be performed to yield one dataset with all of the different totals.

**APPRAOH III**

**PROC SUMMARY**

The "class" statement of the "PROC SUMMARY" provides a way to perform the calculations in one step without even having to sort by the variables.

**APPRAOH I**

**STRAIGHTFORWARD APPROACH**

In this approach a "PROC MEANS" would be used for each of the combinations desired.

**MEANS "BY"**

1. Week
2. Month
3. Season
4. Year

Should the information be required in a single table, then significant manipulation of the four datasets would become necessary, including several sorts. The more types of "BY" combinations needed the more complicated the manipulation. The advantages of SAS can become obscured with this approach.

**APPRAOH II**

**DUMMY RECORDS—SINGLE PROC MEANS**

Artificial values for "BY" variables are set up to represent the different combinations desired. For each original record in the dataset there are now four. One for each desired combination. A "PROC MEANS" will then be performed to yield one dataset with all of the different totals.

**APPRAOH III**

**PROC SUMMARY**

The "class" statement of the "PROC SUMMARY" provides a way to perform the calculations in one step without even having to sort by the variables.

**APPRAOH I**

**STRAIGHTFORWARD APPROACH**

In this approach a "PROC MEANS" would be used for each of the combinations desired.

**MEANS "BY"**

1. Week
2. Month
3. Season
4. Year

Should the information be required in a single table, then significant manipulation of the four datasets would become necessary, including several sorts. The more types of "BY" combinations needed the more complicated the manipulation. The advantages of SAS can become obscured with this approach.

**APPRAOH II**

**DUMMY RECORDS—SINGLE PROC MEANS**

Artificial values for "BY" variables are set up to represent the different combinations desired. For each original record in the dataset there are now four. One for each desired combination. A "PROC MEANS" will then be performed to yield one dataset with all of the different totals.

**APPRAOH III**

**PROC SUMMARY**

The "class" statement of the "PROC SUMMARY" provides a way to perform the calculations in one step without even having to sort by the variables.

**APPRAOH I**

**STRAIGHTFORWARD APPROACH**

In this approach a "PROC MEANS" would be used for each of the combinations desired.

**MEANS "BY"**

1. Week
2. Month
3. Season
4. Year

Should the information be required in a single table, then significant manipulation of the four datasets would become necessary, including several sorts. The more types of "BY" combinations needed the more complicated the manipulation. The advantages of SAS can become obscured with this approach.

**APPRAOH II**

**DUMMY RECORDS—SINGLE PROC MEANS**

Artificial values for "BY" variables are set up to represent the different combinations desired. For each original record in the dataset there are now four. One for each desired combination. A "PROC MEANS" will then be performed to yield one dataset with all of the different totals.

**APPRAOH III**

**PROC SUMMARY**

The "class" statement of the "PROC SUMMARY" provides a way to perform the calculations in one step without even having to sort by the variables.

**APPRAOH I**

**STRAIGHTFORWARD APPROACH**

In this approach a "PROC MEANS" would be used for each of the combinations desired.

**MEANS "BY"**

1. Week
2. Month
3. Season
4. Year

Should the information be required in a single table, then significant manipulation of the four datasets would become necessary, including several sorts. The more types of "BY" combinations needed the more complicated the manipulation. The advantages of SAS can become obscured with this approach.

**APPRAOH II**

**DUMMY RECORDS—SINGLE PROC MEANS**

Artificial values for "BY" variables are set up to represent the different combinations desired. For each original record in the dataset there are now four. One for each desired combination. A "PROC MEANS" will then be performed to yield one dataset with all of the different totals.

**APPRAOH III**

**PROC SUMMARY**

The "class" statement of the "PROC SUMMARY" provides a way to perform the calculations in one step without even having to sort by the variables.
perform a "set" and select out those calculations desired. The _type_ variable produced by the "PROC SUMMARY" is used to do this.

Type values for the above summary

Variables Considered in Summary Calculation

<table>
<thead>
<tr>
<th><em>TYPE</em></th>
<th>Week</th>
<th>Month</th>
<th>Season</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Over entire dataset</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Over season</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Over month</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Over week</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

To get the desired total select _type_ =0,1,2,4

Comparison Between the Three Approaches Using Similar Program

<table>
<thead>
<tr>
<th>Time Difference</th>
<th>Cost Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>24% Increase II over I</td>
<td>8% Increase II over I</td>
</tr>
<tr>
<td>1% Increase III over I</td>
<td>None</td>
</tr>
<tr>
<td>23% Decrease III over II</td>
<td>8% Decrease III over II</td>
</tr>
</tbody>
</table>

REFERENCES