A SIMPLE METHOD OF PRODUCING A RATIO TABLE
Diane Muffett, Purdue University

ABSTRACT

A ratio table can be used to illustrate statistics for various studies including those done on health survey and/or research data. One study might require a prevalence rate; another might need a ratio. In order to generate the numbers in table format, a computation which is essentially a cell by cell division must be done. I found the resources on this topic to be nonexistent. The purpose of this paper is to provide a resource for other programmers encountering this problem. While it will not provide answers for all the tables one might have to do, I hope that it will give the user a basic idea of how to approach a ratio table.

BACKGROUND

Health statistics, particularly those relating to diseases, are often presented in table format. The presentation may include the display of the ratio of persons with a disease to the total population. These ratios are tabled in the same format as the counts for each group. For example, a table would list the number of persons with a disease broken down by age group per year. An identical table would list the corresponding number of people in the total population. The ratio table must correspond with these two tables. In this example, the ratio is the number of persons with the disease divided by the total population. This ratio is thus the equivalent of dividing each cell in the first table by its corresponding cell in the second table.

The SAS® System's TABULATE procedure can do many things including the calculation of percentages using denominators. The denominator can be defined as the sum of a row, column, or the table total. The percentage of one cell in relation to another cell can be calculated, but the sum values for both variables will appear on the table. A cell by cell ratio cannot be calculated using the TABULATE procedure.

EXAMPLES

The records for the subpopulation of persons with the disease are on one file (DISEASE); the records for the total population are on another (PERSONS). The following examples show a portion of the code and the table produced for the population with the disease, the total population, and the ratio of persons with the disease to the total population.
The table listing the subpopulation with the disease (Table A) contains the numerator values. It used the following code:

```plaintext
PROC SORT DATA=DISEASE; BY YEAR;
PROC FORMAT;
  VALUE YR 87='1987' 
     88='1988'
     89='1989';
PROC TABULATE DATA=DISEASE FORMCHAR=' ' FORMAT=7.1;
  CLASS YEAR AGE;
  VAR DIS; /* DIS=1 FOR EACH PERSON IN FILE */
  TABLE YEAR, DIS '-- AGE;
  KEYLABEL SUM=' 
  TITLE 'NUMBER OF PERSONS WITH THE DISEASE PER 1,000 POPULATION';
```

<table>
<thead>
<tr>
<th></th>
<th>&lt;15</th>
<th>15-44</th>
<th>45-64</th>
<th>65-74</th>
<th>75+</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>60</td>
<td>518</td>
<td>1109</td>
<td>744</td>
<td>383</td>
</tr>
<tr>
<td>1988</td>
<td>93</td>
<td>365</td>
<td>1357</td>
<td>711</td>
<td>401</td>
</tr>
<tr>
<td>1989</td>
<td>53</td>
<td>535</td>
<td>1176</td>
<td>815</td>
<td>432</td>
</tr>
</tbody>
</table>

The total population table (Table B) contains the denominator values. It was produced using this code:

```plaintext
PROC SORT DATA=PERSONS; BY YEAR;
PROC FORMAT;
  VALUE YR 87='1987' 
     88='1988'
     89='1989';
PROC TABULATE DATA=PERSONS FORMCHAR=' ' FORMAT=7.1;
  CLASS YEAR AGE;
  VAR POP; /* POP=1 FOR EACH PERSON IN FILE */
  TABLE YEAR, POP '-- AGE;
  KEYLABEL SUM=' 
  TITLE 'NUMBER OF PERSONS (IN THOUSANDS) IN TOTAL POPULATION';
```

<table>
<thead>
<tr>
<th></th>
<th>&lt;15</th>
<th>15-44</th>
<th>45-64</th>
<th>65-74</th>
<th>75+</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>26382</td>
<td>53963</td>
<td>20816</td>
<td>7222</td>
<td>3582</td>
</tr>
<tr>
<td>1988</td>
<td>26406</td>
<td>54048</td>
<td>21108</td>
<td>7388</td>
<td>3616</td>
</tr>
<tr>
<td>1989</td>
<td>26813</td>
<td>54392</td>
<td>21352</td>
<td>7554</td>
<td>3722</td>
</tr>
</tbody>
</table>

1147
The ratio table (Table C) is equal to Table A divided by Table B. It is produced by merging the two files; summarizing the analysis variables; using the resulting sums to compute the ratio; and carrying the ratio values into the TABULATE procedure. This is accomplished using the following code:

```sas
PROC SORT DATA=DISEASE; BY YEAR PERSON;
PROC SORT DATA=PERSONS; BY YEAR PERSON;
DATA PERSONS;
MERGE PERSONS (IN=INPR) DISEASE;
BY YEAR PERSON;
IF INPR;
PROC FORMAT;
VALUE YR 87='1987'
88='1988'
89='1989';
PROC SUMMARY DATA=PERSONS NWAY;
CLASS YEAR AGE;
VAR DIS POP;
OUTPUT OUT=SUMPERS SUM=SUMDIS SUMPOP;
DATA SUMPERS;
SET SUMPERS;
/* RATIO IS CALCULATED USING WHOLE NUMBERS - ROUND NUMBERS FIRST */
RATIO=(ROUND(SUMDIS,1)/ROUND(SUMPOP,1)) * 1000;
PROC TABULATE DATA=PERSONS FORMCHAR=' ' FORMAT=7.1;
CLASS YEAR AGE;
VAR RATIO;
FORMAT YEAR YR ;
TABLE YEAR, RATIO= ' ' AGE;
KEYLABEL SUM=' ';
TITLE 'PREVALENCE RATES OF PERSONS WITH THE DISEASE PER 1,000 POPULATION';
```

<table>
<thead>
<tr>
<th>TABLE C</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREVALENCE RATES OF PERSONS WITH THE DISEASE PER 1,000 POPULATION</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1987</td>
</tr>
<tr>
<td>1988</td>
</tr>
<tr>
<td>1989</td>
</tr>
</tbody>
</table>

1148
CONCLUSION

This is a very basis and simple method of producing a ratio table. It cannot be used in all cases: more complex tables involving multiple breakdowns may require separate calculations of the ratios. I hope that this has given you enough information to start with a simple table and the foundation for you to attempt more complex tables.

REFERENCE


* SAS is a registered trademark of SAS Institute, Inc., Cary, NC 27511-8000, USA.

The author may be contacted at:
Affirmative Action Office
Purdue University
Freehafer Hall, 2nd Floor
West Lafayette, IN 47907