1. Problem Statement

This tutorial topic illustrates procedures for transposing data, i.e., converting observations into variables or variables into observations. This concept is illustrated in figure 1. The observation to variable transposition is represented by the transition from la to lb. Variables to observations is represented by lb to la. The transition from la to lc and vice-versa requires transposing subsets of the data separately. The data structures illustrated by la and lc will be referred to here as linear data sets and rectangular data sets, respectively.

![Figure 1](image)

The ability to transpose data is important in that it provides a measure of independence between data set structure and the structure required by applications programs. For example, if data is needed in a rectangular format for analysis, but the data is sparse, the data may be stored in the transposed linear format to save space.

The applicability of transposing data is illustrated for two examples. Each example is accomplished using PROC TRANSPOSE and then ARRAYS (in a data step).

2. Row by Column Reports

For this example, the data is linear and a row by column report in which one variable determines the rows and another the columns must be produced. Specifically, there is an INCOME value for each EMPLOYEE in each YEAR. A report in which EMPLOYEES are the rows and YEARS are the columns is required. This is accomplished by transposing the data from its linear form to a rectangular form (with each YEAR value becoming a separate variable).

The use of PROC TRANSPOSE is illustrated first (figures 2 and 3). The data is read in (lines 4-6, figure 2) and the EMPLOYEE (row) sums are computed (lines 16-20, figure 2). These two data sets are then interleaved (lines 22-26, figure 2) with the resulting data set transposed to the required rectangular form (lines 33-36, figure 2). PROC PRINT is then used (lines 38-42, figure 2) to print the row by column report. Note that the SUMBY feature is used to produce the YEAR (column) sums.

The same report can be generated using ARRAYS (figures 4 and 5) to transpose the data. As before, the data is read in (lines 4-6, figure 4). Then, in a single data step (lines 21-36, figure 4) the data is transposed and the EMPLOYEE (row) sums are calculated. The report is then printed using PROC PRINT as before.

Several comments about this example are in order. First, the report could have been produced using PUT statements in the same data step that transposed the data. Second, the PROC TRANSPOSE method, although simpler, required more DATA/PROC steps than the ARRAY method. Third, the ARRAY method requires the number of columns (or the maximum) be known with the column headings "hardcoded" in the program. With TRANSPOSE, the number of columns and their headings is transparent to the program.

3. Multiple Output Frequencies

For this example, the one-way frequency distributions for three variables must be generated and output to a data set. The input is illustrated in figure 6. For each attended at the conference, data on whether or not they are giving a paper (PAPER), whether is was invited (INVITED), and whether they are enjoying themselves (ENJOY) are collected.

PROC FREQ can be used on these three variables to produce the frequencies. Unfortunately, since only the frequencies for the last table request in a TABLES statement is output to a data set, only the frequency distribution for one of them can be output. Three TABLES statements with output data sets could be used to circumvent this problem. Another way is to transpose the data so that the three one-way frequency distributions become one two-way frequency distribution. This is accomplished by taking the rectangular data set and creating 3 new observations from each input observation, thus creating a linear data set. The two variables in this linear data set represent: 1) an indicator specifying which original variable the observation represents; and 2) the value of that original variable.

PROC TRANSPOSE is illustrated first (figures 7 and 8). Each observation is transposed into three observations (lines 4-7, figure 7). The two-way frequency is then generated and output to a data set (lines 14-15, figure 7), which is then transposed back to rectangular form (lines 20-23, figure 7) for printing.
The use of ARRAYS is illustrated next (figures 9 and 10). In the first data step, each observation is transposed to create three observations (lines 4-10, figure 9). The two-way frequencies are generated (lines 18-19, figure 9). The output data set is then transposed back to rectangular form in a following data step (lines 24-37, figure 9).

Several comments about this example are in order. First, the ARRAY method requires hard-coding the variable names in several places (lines 6, 27, 28, figure 9) while for the TRANSPOSE method they only need to be specified once (line 7, figure 7). Second, the use of TRANSPOSE required nearly four times as long as the data step using ARRAYS. Thus, for large data sets in which TRANSPOSE's data independence is not needed, the use of ARRAYS in lieu of TRANSPOSE should be considered. For small data sets, TRANSPOSE may be preferred because it is simpler to code.
THIS EXAMPLE ILLUSTRATES TRANSPOSING DATA TO PRODUCE ROW BY COLUMN REPORTS USING PROC TRANSPOSE.

DATA ORIG;
INPUT EMPLOYEE YEAR $ INCOME;
CARDS;
NOTE: DATA SET WORK.ORIG HAS 8 OBSERVATIONS AND 3 VARIABLES. 680 OBS/TRK
NOTE: THE DATA STATEMENT USED 0.04 SECONDS AND 172K.

PROC MEANS NOPRINT;
CREATE EMPLOYEE SUMS;
BY EMPLOYEE;
VAR INCOME;
OUTPUT OUT=TOTALS SUM=INCOME;
NOTE: DATA SET WORK.TOTALS HAS 3 OBSERVATIONS AND 2 VARIABLES. 953 OBS/TRK
NOTE: THE PROCEDURE MEANS USED 0.07 SECONDS AND 172K.

DATA ORIG2;
SET ORIG TOTALS; BY EMPLOYEE;
COMPANY='SUGI';
IF YEAR='ALLYEARS' THEN YEAR='ALLYEARS';
NOTE: DATA SET WORK.ORIG2 HAS 11 OBSERVATIONS AND 4 VARIABLES. 595 OBS/TRK
NOTE: THE DATA STATEMENT USED 0.06 SECONDS AND 180K.

PROC PRINT;
TITLE TRANSPOSING DATA USING PROC TRANSPOSE TO PRODUCE
ROW-BY-COLUMN REPORTS
TITLE2 PRINT OF DATA ORIG2 - INPUT DATA WITH INTERLEAVED SUMS
TITLE3 FINAL REPORT;
NOTE: THE PROCEDURE PRINT USED 0.11 SECONDS AND 172K
AND PRINTED PAGE 1.

PROC TRANSPOSE OUT=RECT;
BY COMPANY EMPLOYEE;
VAR INCOME;
ID YEAR;
NOTE: DATA SET WORK.RECT HAS 3 OBSERVATIONS AND 7 VARIABLES. 340 OBS/TRK
NOTE: THE PROCEDURE TRANSPOSE USING PROC TRANSPOSE USED 0.69 SECONDS AND 172K.

PROC PRINT DATA=RECT(DROP=NAME_);
BY COMPANY;
SUM BY COMPANY;
TITLE3 FINAL REPORT;
NOTE: THE PROCEDURE PRINT USING PROC TRANSPOSE USED 0.11 SECONDS AND 172K
AND PRINTED PAGE 2.

FIGURE 2

FIGURE 3
THIS EXAMPLE ILLUSTRATES TRANSPOSING DATA TO PRODUCE ROW
BY COLUMN REPORTS USING ARRAYS;
DATA ORIG;
INPUT EMPLOYEE YEAR INCOME;
CARDS;
NOTE: DATA SET WORK.ORIG HAS 8 OBSERVATIONS AND 3 VARIABLES. 680 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.06 SECONDS AND 172K.
15 PROC PRINT;
16 TITLE TRANSPOSING DATA USING ARRAYS TO PRODUCE;
17 TITLE2 ROW-BY-COLUMN REPORTS;
18 TITLE3 PRINT OF INPUT DATA ORIG;
19
NOTE: THE PROCEDURE PRINT USED 0.10 SECONDS AND 172K.
AND PRINTED PAGE 1.
21 DATA RECT;
22 SET ORIG BY EMPLOYEE;
23 ARRAY INCS (1) INCOME78-INCOME80;
24 RETAIN INCOME78-INCOME80;
25 COMPANY='SUGI';
26 I=YEAR-77;
27 INCS=INCOME;
28 IF LAST.EMPLOYEE THEN
29 DO /*CREATE OUTPUT RECORD*/;
30 ALLYEARS=SUM(OF INCOME78-INCOME80);
31 OUTPUT;
32 DO OVER INCS /*INITIALIZE TO MISSING*/;
33 INCS=.;
34 END;
35 END /*CREATE OUTPUT RECORD*/;
36 KEEP COMPANY EMPLOYEE INCOME78-INCOME80 ALLYEARS;
37 NOTE: DATA SET WORK.RECT HAS 3 OBSERVATIONS AND 6 VARIABLES. 397 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.07 SECONDS AND 188K.
38 PROC PRINT;
39 BY COMPANY;
40 ID EMPLOYEE;
41 VAR INCOME78-INCOME80 ALLYEARS;
42 SUM INCOME78-INCOME80 ALLYEARS;
43 SUMBY COMPANY;
44 TITLE3 FINAL REPORT;
45
NOTE: THE PROCEDURE PRINT USED 0.10 SECONDS AND 172K.
AND PRINTED PAGE 2.
NOTE: SAS INSTITUTE INC.
SAS CIRCLE
BOX 8000
CARY, N.C. 27511

Figure 4

Figure 5
### SAMPLE DATA FOR MULTIPLE OUTPUT FREQUENCIES EXAMPLE

#### PRINT OF FIRST TWENTY OBSERVATIONS

<table>
<thead>
<tr>
<th>OBS</th>
<th>ID</th>
<th>PAPER</th>
<th>INVITED</th>
<th>ENJOY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### SAMPLE DATA FOR MULTIPLE OUTPUT FREQUENCIES EXAMPLE

#### PRINT OF FREQUENCIES

<table>
<thead>
<tr>
<th>PAPER FREQUENCY</th>
<th>CUM FREQ</th>
<th>PERCENT</th>
<th>CUM PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>680</td>
<td>68.806</td>
<td>68.806</td>
</tr>
<tr>
<td>1</td>
<td>307</td>
<td>31.194</td>
<td>100.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INVITED FREQUENCY</th>
<th>CUM FREQ</th>
<th>PERCENT</th>
<th>CUM PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>910</td>
<td>92.199</td>
<td>92.199</td>
</tr>
<tr>
<td>1</td>
<td>77</td>
<td>7.801</td>
<td>100.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ENJOY FREQUENCY</th>
<th>CUM FREQ</th>
<th>PERCENT</th>
<th>CUM PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>201</td>
<td>20.365</td>
<td>20.365</td>
</tr>
<tr>
<td>1</td>
<td>786</td>
<td>79.635</td>
<td>100.000</td>
</tr>
</tbody>
</table>

*NOTE: DATA SET WORK.LINEAR HAS 680 OBS AND PRINTED PAGE 1.*

*NOTE: DATA SET WORK.LINEAR HAS 6 OBSERVATIONS AND 4 VARIABLES. 529 OBS/TK*
TRANSPOSING DATA USING PROC TRANSPOSE TO PRODUCE MULTIPLE OUTPUT FREQUENCIES

FIRST SIX OBSERVATIONS OF DATA LINEAR

<table>
<thead>
<tr>
<th>OBS</th>
<th>ID</th>
<th><em>NAME</em></th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>PAPER</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>INVITED</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>ENJOY</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>PAPER</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>INVITED</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>ENJOY</td>
<td>1</td>
</tr>
</tbody>
</table>

TRANSPOSING DATA USING PROC TRANSPOSE TO PRODUCE MULTIPLE OUTPUT FREQUENCIES

DATA SET FREQS

<table>
<thead>
<tr>
<th>OBS</th>
<th>COUNT</th>
<th>PERCENT</th>
<th>VALUE</th>
<th><em>NAME</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>201</td>
<td>6.7882</td>
<td>0</td>
<td>ENJOY</td>
</tr>
<tr>
<td>2</td>
<td>910</td>
<td>30.7389</td>
<td>0</td>
<td>INVITED</td>
</tr>
<tr>
<td>3</td>
<td>680</td>
<td>22.9651</td>
<td>0</td>
<td>PAPER</td>
</tr>
<tr>
<td>4</td>
<td>786</td>
<td>26.0651</td>
<td>1</td>
<td>ENJOY</td>
</tr>
<tr>
<td>5</td>
<td>77</td>
<td>2.6005</td>
<td>1</td>
<td>INVITED</td>
</tr>
<tr>
<td>6</td>
<td>307</td>
<td>10.3681</td>
<td>1</td>
<td>PAPER</td>
</tr>
</tbody>
</table>

TRANSPOSING DATA USING PROC TRANSPOSE TO PRODUCE MULTIPLE OUTPUT FREQUENCIES

DATA SET MULTFREQ - MULTIPLE OUTPUT FREQUENCIES

<table>
<thead>
<tr>
<th>OBS</th>
<th>VALUE</th>
<th><em>NAME</em></th>
<th><em>LABEL</em></th>
<th>ENJOY</th>
<th>INVITED</th>
<th>PAPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>COUNT</td>
<td>FREQUENCY</td>
<td>COUNT</td>
<td>201</td>
<td>910</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>COUNT</td>
<td>FREQUENCY</td>
<td>COUNT</td>
<td>786</td>
<td>77</td>
</tr>
</tbody>
</table>

*THIS EXAMPLE ILLUSTRATES TRANSPOSING DATA TO PRODUCE MULTIPLE OUTPUT FREQUENCIES USING ARRAYS.
DATA LINEAR:
SET SAVE.CONFRNE;
DO OVER VARS: VALUE=VARS;
OUTPUT;
END;

NOTE: DATA SET WORK.LINEAR HAS 2961 OBSERVATIONS AND 6 VARIABLES. 366 OBS
NOTE: THE DATA STATEMENT USED 0.22 SECONDS AND 180K.

PROC PRINT DATA=LINEAR(OBS=6);
TITLE TRANSPOSING DATA USING ARRAYS TO PRODUCE MULTIPLE OUTPUT FREQUENCIES;
TITLE3 FIRST SIX OBSERVATIONS OF DATA LINEAR;
NOTE VALUES OF VARIABLES PAPER, INVITED AND ENJOY;
NOTE: THE PROCEDURE PRINT USED 0.10 SECONDS AND 172K
AND PRINTED PAGE 1.

PROC FREQ;
TABLES VALUE*IVAR/NOPRINT SPARSE OUT=FREQS;
NOTE: DATA SET WORK.FREQS HAS 6 OBSERVATIONS AND 4 VARIABLES. 529 OBS/TR
NOTE: THE PROCEDURE FREQ USED 0.33 SECONDS AND 174K.

PROC PRINT;
TITLE3 DATA SET FREQS;
NOTE: THE PROCEDURE PRINT USED 0.09 SECONDS AND 172K
AND PRINTED PAGE 2.

DATA MULTFREQ;
*CONVERT BACK TO RECTANGULAR FORM;
SET FREQS;
BY VALUE;
ARRAY VARS (IVAR) PAPER INVITED ENJOY; *MUST SPECIFY;
RETAIN PAPER--ENJOY;
VARS=COUNT;
IF LAST.VALUE THEN
DO /*CREATE OUTPUT RECORD*/;
OUTPUT;
DO OVER VARS; /*INITIALIZE TO MISSING*/
   VARS=.;
END;
END /*CREATE OUTPUT RECORD*/;
KEEP VALUE PAPER--ENJOY;

NOTE: DATA SET WORK.MULTFREQ HAS 2 OBSERVATIONS AND 4 VARIABLES. 529 OBS/TR
NOTE: THE DATA STATEMENT USED 0.07 SECONDS AND 180K.

PROC PRINT;
TITLE3 DATA SET MULTFREQ - MULTIPLE OUTPUT FREQUENCIES;
NOTE: THE PROCEDURE PRINT USED 0.10 SECONDS AND 172K
AND PRINTED PAGE 3.
TRANSPOSING DATA USING ARRAYS TO PRODUCE MULTIPLE OUTPUT FREQUENCIES

FIRST SIX OBSERVATIONS OF DATA

<table>
<thead>
<tr>
<th>OBS</th>
<th>ID</th>
<th>PAPER</th>
<th>INVITED</th>
<th>ENJOY</th>
<th>IVAR</th>
<th>VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

DATA SET FREQS

<table>
<thead>
<tr>
<th>OBS</th>
<th>VALUE</th>
<th>PAPER</th>
<th>INVITED</th>
<th>ENJOY</th>
<th>IVAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>680</td>
<td>910</td>
<td>201</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>307</td>
<td>77</td>
<td>78.</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>786</td>
<td>26.5451</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

TRANSPOSING DATA USING ARRAYS TO PRODUCE MULTIPLE OUTPUT FREQUENCIES

DATA SET MULTIFREQ

<table>
<thead>
<tr>
<th>OBS</th>
<th>VALUE</th>
<th>PAPER</th>
<th>INVITED</th>
<th>ENJOY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>680</td>
<td>910</td>
<td>201</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>307</td>
<td>77</td>
<td>786</td>
</tr>
</tbody>
</table>

FIGURE 10