
Using Proc Transpose to manipulate Saskatchewan Covered Population Data

Tong Zhu

Saskatchewan Cancer Agency

April 27, 2011



Overview

- Introduction
- Problem statement-Data structure
- The desired result
- 1. Import CSV file
- 2. Using Proc Transpose
- Results

Introduction

- Saskatchewan Cancer Agency (SCA)
Epidemiology
 - Surveillance
 - Prevention
 - Screening
 - Clinical research
- We use covered population (COVPOP) from Sask. Health to calculate
 - Incidence
 - Mortality
 - Prevalence

The problem (original data)

Region	Male <1	1 to 4	5 to 9	10 to 14
1	340.13	1388.95	1621.17	1694.84
2	316.09	1222.57	1482.03	1575.69
3	229	967.5	1306	1402.5
4	1716.46	6913.69	7730.06	8034.89
5	290.77	1258.27	1544.54	1646.06
6	2030.22	8043.08	9351.25	9783.36
7	239.81	995.89	1217.35	1290.15
8	254.91	955.11	1230.04	1322.56
9	588.92	2410.87	2853.03	2826.92
10	702.69	2651.07	2930.53	2972.03
11	343	1156	1179	1146
12	122	504	499	504
13	38	139	144	133

Starting with the end in mind... (desired result)

Region	Agegroup	Sex	Year	CovPop
1	<1	Male	2010	340.13
1	1-4	Male	2010	1388.95
1	5-9	Male	2010	1621.17
1	10-14	Male	2010	1694.84
1	15-19	Male	2010	1923.83
1	<1	Female	2010	349.49
1	1-4	Female	2010	1418.07
1	5-9	Female	2010	1617.69
1	10-14	Female	2010	1709.5
1	15-19	Female	2010	1781.74

Step1. Import to SAS from excel

```
proc import out=work.covPop_plan2010
  datafile='H\SKPOP\Health Region 2010.csv'
  dbms=csv replace;
  getnames=no;
run;
```

Covered population after importing to SAS

VAR1	VAR2	VAR3	VAR4	VAR5
1	340.13	1388.95	1621.17	1694.84
2	316.09	1222.57	1482.03	1575.69
3	229	967.5	1306	1402.5
4	1716.46	6913.69	7730.06	8034.89
5	290.77	1258.27	1544.54	1646.06
6	2030.22	8043.08	9351.25	9783.36
7	239.81	995.89	1217.35	1290.15
8	254.91	955.11	1230.04	1322.56
9	588.92	2410.87	2853.03	2826.92
10	702.69	2651.07	2930.53	2972.03
11	343	1156	1179	1146
12	122	504	499	504
13	38	139	144	133

Step 2. Transpose!

Sort first

```
proc sort data=covPop_plan2010;  
    by var1;  
run;
```

....then transpose

```
proc transpose data=covPop_plan2010  
    out=HRcvspop2010;  
    by var1;  
run;
```

Variables after transpose

- Col1 /*The count of the population*/
- VAR1 /*Health Regions*/
- _NAME_ /*Name of the former variables-VAR3-VAR44*/

Result: Dataset after transpose

VAR1	_NAME	COL1
1	VAR3	340.13
1	VAR4	1388.95
1	VAR5	1621.17
1	VAR6	1694.84
1	VAR7	1923.83
1	VAR8	2023.11
1	VAR9	1966.46
1	...	
1	...	
1	VAR44	141.33
2	VAR3	316.09
2	...	
2	VAR44	138.00

Transpose Syntax

```
PROC TRANSPOSE <DATA=input-data-set> <DELIMITER=delimiter> <LABEL=label> <LET>  
<NAME=name> <OUT=output-data-set> <PREFIX=prefix> <SUFFIX=suffix>;
```

```
  BY <DESCENDING> variable-1  
  <...<DESCENDING> variable-n>  
  <NOTSORTED>;
```

```
  COPY variable(s);
```

```
  ID variable;
```

```
    IDLABEL variable;
```

```
  VAR variable(s);
```



Thanks