SUMMARIZING DATA USING PROC REPORT

Theresa Famolaro, University of California San Francisco, Institute for Health and Aging

ABSTRACT

Many data sets are composed of multiple records for one unique identifier. For example, in a public health database, a patient may have had multiple clinic visits to a single health care provider. The short-term solution would be to utilize PROC FREQ to display a cross-tab on the frequency of visits for each patient. However, this method is often very cumbersome and confusing as output for the end user, and does not make a very efficient report bot in appearance and usage. The client wants to view this information in a manner they can understand. To meet these needs the developer utilized PROC REPORT as the method to summarize the number of visits by each unique patient and then display this information by provider. Some of the tools of PROC REPORT discussed are the COLUMN statement, DEFINE statement, and how to create COMPUTED variables.

THE SAS CODE FOR PROC REPORT

PROC REPORT DATA = *SAS data set*
PANELS = *number of panels*
HEADLINE
HEADSKIP
SPACING = space between columns
COLUMN column-list;
DEFINE item (/define option list);
COMPUTE computed variable;
ENDCOMP;
RBREAK BEFORE/AFTER (Option list)

SUMMARY OF COMMON COMMANDS

COLUMN assigns the arrangement of report variables and of headers.
DEFINE is the statement used to characterize an individual variable. Variables can be defined as DISPLAY, ANALYSIS, GROUP, OR COMPUTED variables. Different options can be assigned for analysis and formatting.
COMPUTE Defines a COMPUTED variable in the report.
ENDCOMP Finishes a series of data statements that are related from a COMPUTE statement.
RBREAK Performs calculations at a break or controls the location and content of break lines. DOUBLE OVERLINE, DOUBLE UNDERLINE, SUMMARIZE are some of the options.

PREPARING DATA FOR PROC REPORT

Researchers and policy makers are usually following particular variables very closely, and often want reports made available as needed. However, it has been my experience that the end user is in need not only of accuracy of data, but also ease of readability. PROC REPORT aids in development and design of an easy to read data report. PROC REPORT contains many options and deviations which can not be explained here at this time, however an attempt will be made to cover all of the basics to do a report and then refer the reader to the appropriate resources. PROC REPORT can be used in a windows and non-windows environment. In order to lay out a more comprehensive understanding of the SAS code, the non-windows environment will be discussed.

For this example, the problem consists in presenting data to nurse consultants for 16 public health counties across the State of California. Race/ethnicity is a common descriptive variable that carries great importance in additional funding, and developing health education promotional materials. PROC REPORT is the procedure used to format and output a report of unique clients by race/ethnicity for each county in the program.

PROC REPORT displays columns of individual variables. In order to prepare data for the COLUMNS statement in PROC REPORT, some re-coding of existing variables is usually necessary. In the following program subset, a unique client identifier is created so those clients with multiple records will not be counted twice. The existing race variable is re-coded into multiple race variables, which are INDIAN, BLACK, CHINESE, FILIPINO, HISPANIC, JAPANESE, OASIAN, WHITE, OTHER, AND, UNKNOWN. In this next segment of programming, we are using PROC MEANS as way to sum variables by unique identifier. In order to sum the number of unique observations for each race by COUNTY, race is then re-coded once more, and PROC MEANS by COUNTY is executed once more. For descriptive purposes, a format for COUNTY, a data macro and TITLE statement are produced.

An Example of a Basic Program:

```
DATA TEMP1;
      SET TEMP;

*/Assigns a unique identifier to each observation*/

    UNIQUEID = TRIM(COUNTY)!!CLIENTID;

/*Re-code race variable*/

      IF RACE = 1 THEN INDIAN = 1;
        ELSE INDIAN = 0;
      IF RACE = 2 THEN BLACK = 1;
        ELSE BLACK = 0;
      IF RACE = 3 THEN CHINESE = 1;
        ELSE CHINESE = 0;
      IF RACE = 4 THEN FILIPINO = 1;
        ELSE FILIPINO = 0;
    IF RACE = 5 THEN HISPANIC = 1; ELSE
HISPANIC = 0;
    IF RACE = 6 THEN JAPANESE = 1; ELSE
JAPANESE = 0;
    IF RACE = 7 THEN OASIAN = 1; ELSE
OASIAN = 0;
    IF RACE = 8 THEN WHITE = 1; ELSE
 WHITE = 0;
    IF RACE = 9 THEN OTHER = 1;  ELSE
 OTHER = 0;
    IF RACE = 10 THEN UNKNOWN = 1; ELSE
UNKNOWN = 0;
    IF RACE NE 0 THEN NUMCLNT = 1; ELSE
NUMCLNT = 0;

      PROC SORT;
        BY UNIQUEID COUNTY SITE;

 DATA TEMP2;
      SET TEMP1;

/*Sum Data by Unique Identifier*/

    PROC MEANS NOPRINT;
      BY UNIQUEID COUNTY SITE;

    VAR NUMCLNT INDIAN BLACK CHINESE
    FILIPINO HISPANIC JAPANESE OASIAN
    WHITE OTHER UNKNOWN;

    OUTPUT OUT = TEMP2

    SUM (NUMCLNT INDIAN BLACK
    CHINESE FILIPINO HISPANIC JAPANESE
    OASIAN WHITE OTHER UNKNOWN) =

    NUMCLNT INDIAN BLACK CHINESE
    FILIPINO HISPANIC JAPANESE OASIAN
    WHITE OTHER UNKNOWN;

 DATA TEMP3;
      SET TEMP2;

/*Recount race variables for unique clients*/

    IF NUMCLNT NE 0 THEN NUMCLNT = 1;
    IF INDIAN NE 0 THEN INDIAN = 1;
      ELSE INDIAN = 0;
    IF BLACK NE 0 THEN BLACK = 1;
      ELSE BLACK = 0;
    IF CHINESE NE 0 THEN CHINESE = 1;
      ELSE CHINESE = 0;
    IF FILIPINO NE 0 THEN FILIPINO = 1;
      ELSE FILIPINO = 0;
    IF HISPANIC NE 0 THEN HISPANIC = 1;
      ELSE HISPANIC = 0;
    IF JAPANESE NE 0 THEN JAPANESE =  1;
      ELSE JAPANESE = 0;
    IF OASIAN NE 0 THEN OASIAN = 1;
      ELSE OASIAN = 0;
    IF WHITE NE 0 THEN WHITE = 1;
      ELSE WHITE = 0;
    IF OTHER NE 0 THEN OTHER = 1;
      ELSE OTHER = 0;
    IF UNKNOWN NE 0 THEN UNKNOWN = 1;
      ELSE UNKNOWN = 0;
    PROC SORT;
      BY COUNTY;

  DATA TEMP4;
    SET TEMP3;

/*Summarize data by county*/

    PROC MEANS NOPRINT;
      BY COUNTY;

    VAR NUMCLNT INDIAN BLACK CHINESE
    FILIPINO HISPANIC JAPANESE OASIAN
    WHITE OTHER UNKNOWN;

    OUTPUT OUT = TEMP4

    SUM (NUMCLNT  INDIAN BLACK
    CHINESE FILIPINO HISPANIC JAPANESE
    OASIAN WHITE OTHER UNKNOWN) =

    NUMCLNT INDIAN BLACK CHINESE
    FILIPINO HISPANIC JAPANESE OASIAN
    WHITE OTHER UNKNOWN;

  DATA TEMP5;
    SET TEMP4;

/*Create formats for county variables*/

      PROC FORMAT;
      VALUE $CTY
            '09' =  'EL DORADO'
            '10' = 'FRESNO'
            '12' = 'HUMBOLDT'
            '16' = 'KINGS'
            '17' = 'LAKE'
            '20' = 'MADERA'
            '21' = 'MARIN'
```

```
                '30' = 'ORANGE'
                '31' = 'PLACER'
                '65' = 'LA - EAST'
                '36' = 'SAN BERNARDINO'
                '37' = 'SAN DIEGO 37'
                '67' = 'SAN DIEGO 67'
                '42' = 'SANTA BARBARA'
                '50' = 'STANISLAUS'
                '54' = 'TULARE'
                '57' = 'YOLO';

      FORMAT COUNTY $CTY. ;

/*Create Macro for date*/

      %MACRO FDATE(FMT);
         %GLOBAL FDATE;

        DATA _NULL_;
          CALL SYMPUT
   ("FDATE",LEFT(PUT("&SYSDATE"D,&FMT)));
        RUN;
       %MEND FDATE;

      %FDATE (WORDDATE.)

/*Title Statement and Footnote Statement*/

  TITLE1 "                    &FDATE";
  TITLE2 "                          ";
  TITLE3 '--------------------';
  TITLE4  'CLIENT ETHNICITY';
  TITLE5 '--------------------';
  TITLE6 "                          ";
FOOTNOTE1 "REPORT IS BASED ON
UNDUPLICATED COUNT OF CLIENTS.";

/*Invoke Proc Report*/

PROC REPORT
DATA = TEMP
PANELS = 20
HEADLINE
HEADSKIP
SPACING = 1;

COLUMN (COUNTY NUMCLNT INDIAN PI BLACK
PB CHINESE PC FILIPINO FP HISPANIC PH
JAPANESE PJ OASIAN PO WHITE PW OTHER
POTH UNKNOWN PU TOTAL);

DEFINE COUNTY/GROUP 'C/O/U/N/T/Y' FORMAT =
$CTY. WIDTH = 15 RIGHT;
DEFINE INDIAN/ANALYSIS 'I/N/D/I/A/N' WIDTH = 3
RIGHT;
DEFINE BLACK/ANALYSIS 'B/L/A/C/K' WIDTH = 3
RIGHT;
DEFINE CHINESE/ANALYSIS 'C/H/I/N/E/S/E' WIDTH
= 3 RIGHT;
DEFINE FILIPINO/ANALYSIS 'F/I/L/I/P/I/N/O' WIDTH
= 3 RIGHT;
DEFINE HISPANIC/ANALYSIS 'H/I/S/P/A/N/I/C'
WIDTH = 3 RIGHT;
DEFINE JAPANESE/ANALYSIS 'J/A/P/A/N/E/S/E'
WIDTH = 3 RIGHT;
DEFINE OASIAN/ANALYSIS 'O/T/H/ /A/S/I/A/N'
WIDTH = 3 RIGHT;
DEFINE WHITE/ANALYSIS 'W/H/I/T/E' WIDTH = 5
RIGHT;
DEFINE OTHER/ANALYSIS 'O/T/H/E/R' WIDTH = 3
RIGHT;
DEFINE UNKNOWN/ ANALYSIS  'U/N/K/N/O/W/N'
WIDTH = 3 RIGHT;
DEFINE NUMCLNT/ANALYSIS  'T/O/T/A/L' WIDTH =
5 RIGHT NOPRINT;
DEFINE PI/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PB/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PC/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PF/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PH/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PJ/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PO/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE PW/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE POTH/ COMPUTED
FORMAT=PERCENT10. WIDTH = 5 center '%';
DEFINE PU/ COMPUTED FORMAT=PERCENT10.
WIDTH = 5 center '%';
DEFINE TOTAL/ COMPUTED 'T/O/T/A/L' WIDTH = 5
RIGHT;

COMPUTE PI;
      IF INDIAN.SUM = 0 THEN PI = 0; ELSE
      PI = INDIAN.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE PB;
      IF BLACK.SUM = 0 THEN PB = 0; ELSE
      PB = BLACK.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE PC;
      IF CHINESE.SUM = 0 THEN PC = 0; ELSE
      PC = CHINESE.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE PF;
      IF FILIPINO.SUM = 0 THEN PF = 0; ELSE
      PF = FILIPINO.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE PH;
      IF HISPANIC.SUM = 0 THEN PH = 0; ELSE
      PH = HISPANIC.SUM / NUMCLNT.SUM;
ENDCOMP;
```

```
COMPUTE PJ;
      IF JAPANESE.SUM = 0 THEN PJ = 0; ELSE
      PJ = JAPANESE.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE PO;
      IF OASIAN.SUM = 0 THEN PO = 0; ELSE
      PO = OASIAN.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE PW;
      IF WHITE.SUM = 0 THEN PW = 0; ELSE
      PW = WHITE.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE POTH;
      IF OTHER.SUM = 0 THEN POT = 0; ELSE
      POTH = OTHER.SUM / NUMCLNT.SUM;
ENDCOMP;


COMPUTE PU;
     IF UNKNOWN.SUM = 0 THEN PU = 0;
     ELSE
     PU = UNKNOWN.SUM / NUMCLNT.SUM;
ENDCOMP;

COMPUTE TOTAL;
      TOTAL=NUMCLNT.SUM;
ENDCOMP;

RBREAK AFTER / DOL SUMMARIZE;

RUN;
```

Conclusions

PROC REPORT is an excellent format to summarize data. PROC REPORT combines tools from PROC TABULATE, PROC MEANS, and PROC PRINT into one simple format. It is an outstanding method in presenting descriptive statistics that are easy to understand for the end user. By saving a REPORT format or by inserting a unique data set into the data step, the same report can be written for multiple data sets.

References

SAS Institute Inc., Reporting from the Field: SAS® Software Experts Present Real-World Report-Writing Applications, First Edition, Carey, NC: SAS Institute Inc., 1994

SAS Institute Inc., SAS® Guide to the REPORT Procedure Usage and Reference Version 6, First Edition, Carey, NC: SAS Institute, Inc., 1990

SAS Institute Inc., SAS® Macro Language: Reference Version 6, First Edition, Carey, NC: SAS Institute, Inc., 1997

About the Author

Theresa Famolaro is Research Associate at the Institute for Health and Aging at the University of California San Francisco. Currently, she manages the data for the Preventive Health Care for the Aging Program from the State of California, Department of Health Services. She has been a SAS programmer for 2 years.

Acknowledgements

Thanks to Leslie Ross, Ph.D. of the UCSF Institute for Health and Aging for her priceless support.

Grant funding

The California Department of Health Services Preventive Health Care for the Aging Program.

Contact information

Theresa Famolaro theresa@itsa.ucsf.edu

STATE OF CALIFORNIA - DEPARTMENT OF HEALTH SERVICES
PREVENTIVE HEALTH CARE FOR THE AGING

CLIENT ETHNICITY

AUGUST 1998 - JUNE 1999
CONTRACTOR: ALL CONTRACTORS

COUNTY	INDIAN	%	BLACK	%	CHINESE	%	FILIPINO	%	HISPANIC	%	JAPANESE	%	OTHASIAN	%	WHITE	%	OTHER	%	UNKNOWN	%	TOTAL
EL DORADO	7	3%	0	0%	0	0%	0	0%	6	3%	0	0%	2	1%	190	92%	1	0%	0	0%	206
FRESNO	2	0%	60	11%	2	0%	14	3%	193	35%	15	3%	8	1%	261	47%	4	1%	0	0%	559
HUMBOLDT	15	4%	2	1%	0	0%	1	0%	8	2%	0	0%	3	1%	345	92%	0	0%	0	0%	373
KINGS	4	2%	7	3%	0	0%	6	3%	76	32%	1	0%	5	2%	135	56%	5	2%	0	0%	239
LAKE	2	1%	2	1%	0	0%	0	0%	12	4%	1	0%	0	0%	304	94%	2	1%	0	0%	323
MADERA	2	1%	3	1%	1	0%	4	2%	30	14%	3	1%	0	0%	166	79%	0	0%	0	0%	209
MARIN	1	0%	13	3%	2	1%	1	0%	41	11%	2	1%	7	2%	311	82%	1	0%	0	0%	377
ORANGE	0	0%	10	1%	11	1%	13	2%	118	15%	12	2%	24	3%	570	75%	3	0%	1	0%	762
PLACER	3	1%	2	1%	3	1%	2	1%	15	4%	14	4%	2	1%	349	89%	1	0%	0	0%	391
SAN BERNARDINO	1	0%	42	12%	2	1%	7	2%	107	30%	4	1%	4	1%	182	52%	4	1%	0	0%	353
SAN DIEGO 37	1	0%	11	3%	5	1%	50	13%	91	23%	5	1%	11	3%	219	55%	6	2%	0	0%	399
SANTA BARBARA	1	0%	5	1%	0	0%	9	2%	52	13%	3	1%	1	0%	332	80%	11	3%	0	0%	414
STANISLAUS	1	0%	7	2%	1	0%	2	1%	24	7%	2	1%	16	4%	304	84%	3	1%	1	0%	361
TULARE	4	1%	7	2%	0	0%	0	0%	96	23%	3	1%	44	11%	261	62%	4	1%	0	0%	419
YOLO	0	0%	8	2%	8	2%	4	1%	30	8%	9	3%	1	0%	293	82%	4	1%	0	0%	357
LA - EAST	1	0%	25	6%	32	7%	15	3%	118	26%	4	1%	12	3%	240	54%	1	0%	0	0%	448
SAN DIEGO 67	1	0%	1	0%	0	0%	23	9%	59	24%	21	9%	2	1%	135	55%	4	2%	0	0%	246
	46	1%	**	3%	67	1%	**	2%	1E3	17%	99	2%	**	2%	4597	71%	54	1%	2	0%	6436

REPORT IS BASED ON UNDUPLICATED COUNT OF CLIENTS FOR WHOM RACE HAS BEEN CODED ON AT LEAST

ONE ENCOUNTER IN THE REPORTING PERIOD