Paper 203-26

Creating a Consolidated Report from Several SAS® Outputs Using DATA _NULL_ and PUT Statements

Anne Horney and Gail F. Kirk,
Cooperative Studies Program Coordinating Center, Perry Point, Maryland

ABSTRACT

We often find it necessary to combine output values from several procedures into one table or page of a report. When we need a p-value from one procedure, and means and standard errors from another and frequencies from a third procedure, the separate pieces of information must be brought together in one DATA step before using PUT statements to print the table. The problem of how to get the data/output/information ready to print is particularly difficult for inexperienced programmers. We will show examples using several different methods such as PROC FREQ, PROC MEANS, PROC UNIVARIATE and DATA steps. In addition to creating PUT tables these methods can also be used to create delimited ASCII files that can be merged into word processing or spreadsheet packages.

INTRODUCTION

Our goal is to produce a table which describes a continuous variable (Age) and two categorical variables (Gender and Marital Status) by treatment group. As is so often the case in SAS programming, there is more than one way to accomplish the task. We will use SAS procedures and DATA step programming to show some ways to meet the goal. Statistics for the continuous variable will be n, mean, standard deviation and the probability of no difference between treatment groups; for the discrete variables statistics will be n, percent-of-group, and probability of no difference between groups. See the second example for methods of working with probabilities.

Our sample dataset `form1` contains five variables:

`ID`	Subject identifier
`Treatmnt`	Treatment A or B
`Age`	Subject's age in years
`Gender`	1=male, 2=female
`MarStat`	1=never married, 2=married, 3=divorced/widowed

The 51 observations have been sorted by treatment group; none of the data are missing. Assume the formats `gender.` and `marital.` exist.

EXAMPLE ONE.

N, MEAN, AND STANDARD DEVIATION USING PROC MEANS

One method of calculating statistics such as means and standard deviations is to use the procedure (PROC) MEANS, which can create an output data set containing the calculated values. The output dataset is named in the OUTPUT statement; the desired statistics are requested and named. When the CLASS statement is used, the dataset contains statistics for each value of the class variable (here, treatmnt) and for the total sample. The data step AGEA changes the value of treatmnt for the total sample from ' ' (missing) to 'C' so that the data can be sorted to put the total last.

SAS Code:

```
proc means data=form1;
  class treatmnt;
  var age;
  output out=age
    n=n_age mean=m_age std=s_age;
  run;

data agea; set age;
  if treatmnt=' ' then treatmnt='C';
  keep treatmnt n_age m_age s_age;
  run;

proc sort; by treatmnt;
  run;
```

N AND PERCENT USING ARRAYS IN A DATA STEP

We can count the occurrences of each category of a discrete variable and calculate percentages in a DATA step. An array is defined for each categorical variable of interest. The array GEN will be used to sum the occurrences of 1(male) and 2(female) for each treatment group and overall; likewise, MARITAL will hold the marital status counts. For each two-dimensional array the rows are the categories and the columns are the treatment groups. The percentages will be calculated later in the DATA _NULL_ step.

SAS Code:

```
  data counts(keep=gen1-gen9 mar1-mar12);
    set form1 end=eof;
*columns: Treament A, B, total;
*Rows: Male, Female, Total;
    array gen{3,3}  gen1-gen9;
*Rows: Nvr Married,Married,Divorced,total;
    array marital{4,3}  mar1-mar12;

    retain gen1-gen9 mar1-mar12 0;

    if treatmnt='A' then trt=1;
    if treatmnt='B' then trt=2;

    if gender in (1,2) then do;
      gen{gender,trt}+1;
      gen{gender,3}+1;     *row total;
      gen{3,trt}+1;        *column total;
      gen{3,3}+1;          *overall total;
    end;

    if 1 le marstat le 3 then do;
      marital{marstat,trt}+1;
      marital{marstat,3}+1;   *row total;
```

```
    marital{4,trt}+1;        *column tot;
    marital{4,3}+1;          *overall tot;
  end;

  if eof then output;
  run;
```

DATA _NULL_ AND PUT

We bring the pieces together and "put" out the data in tabular form.

SAS Code:

```
 data _null_; set agea(in=ina )
                   counts(in=inc);
  array gen{3,3}  gen1-gen9;
  array marital{4,3}  mar1-mar12;
  retain col 0;

  file print header=h1;

*Age*;
  if ina then do;
    col+17;
    if treatmnt='A' then put
      / @1 'Age (years)' @;
    put @col n_age 2.
        +2 m_age 4.1
        +2 s_age 4.1 @;
  end;
  if treatmnt='C' then put;

  if inc then do;
*Gender*;
    put /// @20 'N     %'
             @37 'N     %'
             @54 'N     %'
           / @19 '__________'
             @36 '__________'
             @53 '__________'
           / @1 'Gender';
    do row=1 to 2;
      put @3 row :gender. @;
      col=2;
      do rx=1 to 3;
        col=col+17;
        if gen{3,rx} gt 0 then
          pc=gen{row,rx}/gen{3,rx} * 100;
        else pc=.;
        put @col gen{row,rx} 2.
          + 3 pc 5.1 @;
      end;
      put;
    end;

*Marital Status*;
    put / @1 'Marital Status';
    do row=1 to 3;
      put @3 row :marital. @;
      col=2;
      do rx=1 to 3;
        col=col+17;
        if marital{4,rx} gt 0 then
          pc=marital{row,rx}/marital{4,rx}
            * 100;
        else pc=.;
        put @col marital{row,rx} 2.
          +3 pc 5.1 @;
      end;
      put;
    end;
```

```
    end;
    return;
    h1:
    title 'Demographic Information by'
      'Treatment Group';
    put //   @33 'Treatment Group'
         //   @23 'A' @40 'B' @56 'Total'
          / @1 'Demographic'
          / @1 'Information'
            @17 ' N  Mean  S.D.'
            @34 ' N  Mean  S.D.'
            @51 ' N  Mean  S.D.'
          / @1  '___________'
            @17 '______________'
            @34 '______________'
            @51 '______________';
  run;
```

EXAMPLE TWO. ANOTHER WAY TO PRODUCE THE TABLE (and add some probability values)

N, MEAN, AND STANDARD DEVIATION USING PROC UNIVARIATE

The procedure UNIVARIATE produces many of the same statistics as PROC MEANS, and the OUTPUT data set is similar. In addition, UNIVARIATE calculates medians and percentiles, although we will not be using those values here. Note that the data set must be sorted into BY variable order. Note also that the procedure must be executed twice: once for by-group processing, once for overall.

We will also get the treatment p-value in this second example.

SAS Code:

```
 proc univariate data=form1 noprint;
   by treatmnt;
   var age;
   output out=age
     n=n_age mean=m_age std=s_age;
 run;
 proc univariate data=form1 noprint;
   var age;
   output out=tage
     n=n_age mean=m_age std=s_age;
 run;
```

T-TEST P-VALUE USING ANOVA PROCEDURE

Since the TTEST procedure does not have an OUTPUT option, we will use the ANOVA procedure, which calculates the equivalent probability value.

SAS Code:

```
 proc anova data=form1 outstat=p_age;
   class treatmnt;
   model age=treatmnt;
 run;
```

N, PERCENT, AND CHI-SQUARE PROBABILITY USING PROC FREQ

Counts and percentages, as well as chi-square probabilities, for discrete variables can be obtained by use of the procedure FREQ along with the procedure's two types of output statements. Note that PROC FREQ is performed separately for the one-way tables (totals).

SAS Code:

```
Proc freq data=form1;
  table gender*treatmnt
      /chisq out=gender outpct;
  output out=cgender(keep=p_pchi) chisq;
  run;

proc freq data=form1;
  table marstat*treatmnt
      /chisq out=marital outpct;
  output out=cmarital(keep=p_pchi) chisq;
  run;

proc freq data=form1;
  table gender / out=tgender;
  table marstat / out=tmarital;
  run;

data gen; set gender
              tgender;
  by gender;
  if _n_=1 then set cgender;
  if treatmnt=' ' then treatmnt='C';
  run;

data mar; set marital
              tmarital;
  by marstat;
  if _n_=1 then set cmarital;
  if treatmnt=' ' then treatmnt='C';
  run;
```

DATA _NULL_ AND PUT

Again we bring the pieces together and put out the data table.

SAS Code:

```
data _null_;
  set age(in=in_a)
       tage(in=in_t)
       p_age(in=inp
             where=(_type_='ANOVA'))
       gen(in=in_g)
       mar(in=in_m);
  retain col 0;

  file print header=h1;
  if in_g and treatmnt='A' and gender=1
    then put
          // @20 'N     %'
             @37 'N     %'
             @54 'N     %'
             @68 'probability'
           / @19 '__________'
             @36 '__________'
             @53 '__________'
             @68 '___________'
           /  @1 'Gender' @;

  if in_m and treatmnt='A' and marstat=1
    then put // @1 'Marital Status' @;

  if treatmnt='A' then  do;
    if in_a then put / @1 'Age (years)' @;
    If in_g then put
      / @3 gender gender. @;
    If in_m then put
      / @3 marstat marital. @;
  end;

  if in_a or in_t then do;
    col=col+17;
    put @col n_age 2.
             +2 m_age 4.1
             +2 s_age 4.1 @;
  end;

  if in_g or in_m then do;
    if treatmnt='A' then col=2;
    col+17;
    if treatmnt in ('A','B') then put
      @col count 2. + 3 pct_col 5.1 @;
    else if treatmnt = 'C' then
      put @col count 2. +3 percent 5.1 @ ;
  end;

  if inp then put @71 prob 5.3;
  if in_g and treatmnt='C' and gender=1
    then put @71 p_pchi 5.3 @;
  if in_m and treatmnt='C' and marstat=1
    then put @71 p_pchi 5.3 @;

  return;
  h1:
  title 'Demographic Information by'
    'Treatment Group';
  put // @33 'Treatment Group'
       // @23 'A' @40 'B' @56 'Total'
        / @1  'Demographic'
        / @1  'Information'
          @17 ' N  Mean  S.D.'
          @34 ' N  Mean  S.D.'
          @51 ' N  Mean  S.D.'
          @68 'probability'
        / @1  '___________'
          @17 '______________'
          @34 '______________'
          @51 '______________'
          @68 '___________';
run;
```

CONCLUSION

Using DATA step arrays and/or output data sets from SAS procedures will require some extra programming steps. The time is well spent, though, when you compare the programming time to the alternative of hand-transcribing, and checking, and checking again, and verifying that the correct numbers from the appropriate SAS output listings have been put in the correct destination slots.

ACKNOWLEDGMENTS

SAS and all other SAS Institute Inc. products are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

Anne Horney
(410) 642-2411 ext. 5298
Rebecca.Horney@med.va.gov

Gail F. Kirk
(410) 642-2411 ext. 5296
Gail.Kirk@med.va.gov

Mailing address for both:
CSPCC (151E)
VA Medical Center
P. O. Box 1010
Perry Point, MD 21902

APPENDIX A. Sample Data Set form1

TREATMNT	ID	AGE	GENDER	MARSTAT
A	1	21	1	1
A	2	36	1	2
A	3	45	2	3
A	4	40	2	1
A	5	44	1	1
A	6	32	1	2
A	7	25	2	1
A	8	33	2	3
A	9	34	1	2
A	10	27	1	3
A	11	54	1	1
A	12	37	2	2
A	13	33	1	2
A	14	44	1	3
A	15	22	2	2
A	16	32	2	3
A	17	25	2	2
A	18	33	2	1
A	19	34	1	2
A	20	27	1	3
A	21	54	1	1
A	22	37	2	2
A	23	33	1	3
A	24	44	1	3
A	25	22	2	2
B	26	32	2	3
B	27	25	2	2
B	28	33	2	2
B	29	34	1	1
B	30	27	1	3
B	31	54	1	1
B	32	37	2	2
B	33	33	1	2
B	34	44	1	3
B	35	22	2	3
B	36	32	2	3
B	37	25	2	2
B	38	33	2	1
B	39	34	1	2
B	40	27	1	3
B	41	54	1	1
B	42	37	2	2
B	43	33	1	3
B	44	44	1	3
B	45	22	2	2
B	46	32	2	3
B	47	45	2	1
B	48	32	1	2
B	49	24	1	3
B	50	33	1	3
B	51	33	1	3

APPENDIX B. Example One Output Data Sets

Example One: Data Set "age"

OBS	TREATMNT	_TYPE_	_FREQ_	N_AGE	M_AGE	S_AGE
1		0	51	51	34.2941	8.73223
2	A	1	25	25	34.7200	9.11738
3	B	1	26	26	33.8846	8.50565

Example One: Data Set "agea" after sorting

OBS	TREATMNT	N_AGE	M_AGE	S_AGE
1	A	25	34.7200	9.11738
2	B	26	33.8846	8.50565
3	C	51	34.2941	8.73223

Example One: Data Set "counts"

OBS	GEN1	GEN2	GEN3	GEN4	GEN5	GEN6	GEN7	GEN8	GEN9	MAR1	MAR2	MAR3	MAR4	MAR5	MAR6	MAR7	MAR8	MAR9	MAR10	MAR11	MAR12
1	14	14	28	11	12	23	25	26	51	7	5	12	10	9	19	8	12	20	25	26	51

APPENDIX C. Example Two Output Data Sets

Example Two: Data Set "age"

OBS	TREATMNT	N_AGE	M_AGE	S_AGE
1	A	25	34.7200	9.11738
2	B	26	33.8846	8.50565

Example Two: Data Set "tage"

OBS	N_AGE	M_AGE	S_AGE
1	51	34.2941	8.73223

Example Two: Data Set "p_age"

OBS	_NAME_	_SOURCE_	_TYPE_	DF	SS	F	PROB
1	AGE	ERROR	ERROR	49	3803.69	.	.
2	AGE	TREATMNT	ANOVA	1	8.89	0.11458	0.73644

Example Two: Data Set "gender"

OBS	GENDER	TREATMNT	COUNT	PERCENT	PCT_ROW	PCT_COL
1	1	A	14	27.4510	50.0000	56.0000
2	1	B	14	27.4510	50.0000	53.8462
3	2	A	11	21.5686	47.8261	44.0000
4	2	B	12	23.5294	52.1739	46.1538

Example Two: Data Set "cgender"

OBS	P_PCHI
1	0.87719

Example Two: Data Set "tgender"

OBS	GENDER	COUNT	PERCENT
1	1	28	54.9020
2	2	23	45.0980

Example Two: Data Set "marital"

OBS	MARSTAT	TREATMNT	COUNT	PERCENT	PCT_ROW	PCT_COL
1	1	A	7	13.7255	58.3333	28.0000
2	1	B	5	9.8039	41.6667	19.2308
3	2	A	10	19.6078	52.6316	40.0000
4	2	B	9	17.6471	47.3684	34.6154
5	3	A	8	15.6863	40.0000	32.0000
6	3	B	12	23.5294	60.0000	46.1538

Example Two: Data Set "cmarital"

OBS	P_PCHI
1	0.55800

Example Two: Data Set "tmarital"

OBS	MARSTAT	COUNT	PERCENT
1	1	12	23.5294
2	2	19	37.2549
3	3	20	39.2157

APPENDIX D. GOAL: Tabulated Data (Example Two)

Demographic Information by Treatment Group

Demographic Information	Treatment Group A			Treatment Group B			Total			probability
	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	
Age (years)	25	34.7	9.1	26	33.9	8.5	51	34.3	8.7	0.736

	A N	A %	B N	B %	Total N	Total %	probability
Gender							
Male	14	56.0	14	53.8	28	54.9	0.877
Female	11	44.0	12	46.2	23	45.1	
Marital Status							
Single	7	28.0	5	19.2	12	23.5	0.558
Married	10	40.0	9	34.6	19	37.3	
Divorced	8	32.0	12	46.2	20	39.2	