

Chapter 6. SAS Reporting

6.1. Produce Tabular Reports: PROC TABULATE

- Produce a variety of tabular reports, displaying frequencies and descriptive statistics.
- Similar to PROC PRINT, PROC MEANS, PROC FREQ, etc., but PROC TABULATE produces prettier reports.

General Syntax

```
proc tabulate data=dataset;  
  class list-of-categorical-variables / <options>;  
  var list-of-numeric-variables / <options>;  
  table page-variable, row-variable, column-variable / <options>;  
run;
```

- CLASS: Specify *categorical* variables to be used for dividing observations into groups.
- VAR: Specify *numeric* variables of which you will get the summary statistics.

- TABLE

- Tell SAS how to organize a table.
- Specify the dimensions of the table up to 3 dimensions.
- Separate each dimension of the table by putting a comma (,) between variable names.
- If 2 dimensions are specified, then you get rows and columns;
If only 1 is specified, then that becomes, by default, the column dimension.
- One TABLE statement defines only one table, but it is possible to use multiple TABLE statements in one procedure.
- Use an asterisk (*) between variable names if including multiple variables in one dimension.

- Missing data

- By default, observations are excluded from tables if they have missing values for variables listed in CLASS statement.
- If you want to keep these observations, simply add missing option:

```
proc tabulate data=dataset MISSING;
```

- Keyword
 - By default, PROC TABULATE produces simple *counts* of observations in each category.
 - For other statistics (listed below), include keyword in TABLE Statement.
 - Include an asterisk (*) after/before variable names.

Option	Description	Option	Description
MAX	Maximum	ALL	Add a row, column, or page showing the total.
MIN	Minimum	N	Number of non-missings
MEAN	Mean	NMISS	Number of missings
MEDIAN	Median	SUM	Sum
MODE	Mode	PCTN	Percentage of observations for the group
STDDEV	Standard deviation	PCTSUM	Percentage of a total sum represented by the group

- Customizing table
 - `FORMAT=`: Change the format of all data cells in the table.
 - Associate different format with each of the variables (`*f=format`)
 - `KEYLABEL`: Allow to provide a label for any of the keywords used by the procedure.
 - `TABLE` option
 - `BOX=`: Allow to write text in the upper left corner of the table (usually empty).
 - `MISSTEXT=`: Specify a value for SAS to print in empty data cells.

Example

Raw
Data

Obs	obs	Gender	Type	Agegroup	White blood cell	Red blood cell	Cholesterol
1	1	Female	AB	Young	7710	7.40	258
2	2	Male	AB	Old	6560	4.70	.
3	3	Male	A	Young	5690	7.53	184
4	4	Male	B	Old	6680	6.85	.
5	5	Male	A	Young	.	7.72	187

SAS Code

```
* 1-dimensional table;
proc tabulate data=blood;
  class Gender;
  table Gender;
run;
```

Output

Gender	
Female	Male
N	N
440	560

```
* 2-dimensional table;
proc tabulate data=blood;
  class Gender Type;
  table Gender, Type;
run;
```

	Type			
	A	AB	B	O
	N	N	N	N
Gender				
Female	178	20	34	208
Male	234	24	62	240

```

* Crossing, grouping, and concatenating;
proc tabulate data=blood;
  class Gender Type Agegroup;
  var wbc;
  table Gender All, mean*wbc*(Type Agegroup
All);
run;

```

	Mean						
	White blood cell						
	Type				Agegroup		All
	A	AB	B	O	Old	Young	
Gender							
Female	7218.13	7420.56	6716.07	7049.63	7105.98	7121.36	7112.43
Male	7051.01	6893.00	6990.53	6930.43	6939.35	7061.66	6987.54
All	7123.45	7142.89	6900.12	6987.06	7011.56	7089.08	7042.97

```

* Customizing your table;
proc tabulate data=blood format=comma9.2;
  class Gender AgeGroup;
  var rbc wbc Chol;
  table (Gender=' ' ALL)*(AgeGroup=' ' All),
    rbc*(n*f=3. mean*f=5.1)
    wbc*(n*f=3. mean*f=comma7.)
    chol*(n*f=4. mean*f=7.1);
  keylabel ALL = 'Total';
run;

```

		Red blood cell		White blood cell		Cholesterol	
		N	Mean	N	Mean	N	Mean
Female	Old	242	5.5	234	7,106	208	195.9
	Young	167	5.5	169	7,121	141	212.3
	Total	409	5.5	403	7,112	349	202.5
Male	Old	309	5.4	306	6,939	279	199.1
	Young	198	5.5	199	7,062	167	203.1
	Total	507	5.5	505	6,988	446	200.6
Total	Old	551	5.5	540	7,012	487	197.7
	Young	365	5.5	368	7,089	308	207.3
	Total	916	5.5	908	7,043	795	201.4

6.2. Produce Simple Outputs: PROC REPORT

- Produce output that is similar to PROC PRINT, PROC MEANS, PROC FREQ, etc., but more visually appealing.

General Syntax

```
proc report data=dataset NOWINDOWS;  
... <options>;  
run;
```

- Without any options, it generates the same output as PROC PRINT.
 - Except there is no observation number (obs).
 - PROC PRINT prints the variable names as column headings;
PROC REPORT uses variable labels if they exist.
- If you have at least one character variable in your report, then, by default, SAS produces a detail report with one row per observation.
- If the report includes only numeric variables, then, by default, PROC REPORT will *sum* those variables.

- Report window
 - If you have already run PROC REPORT, you need to close the interactive Report window before re-running it.
 - NOWINDOWS or NOWD: Turn off the report output and send it to the output screen.
 - WINDOWS: Turn the default back on.
 - HEADLINE: Place an underline underneath column headings
 - HEADSKIP: Place a blank line underneath column headings.
 - Using HEADLINE and HEADSKIP together: Create a blank line underneath the underline.
 - SPLIT= " ": Tell SAS that you want to split the comments between the words (blank). Otherwise, other characters (slashes) are possible as line breaks.
 - MISSING: By default, observations are excluded from reports if they have missing values for variables listed in ORDER, CROUP, ACROSS statement. Use MISSING option to keep missing observations.

- Statements

Statement	Description
COLUMN	List specific variables that you want to include in the report.
WHERE	Print observations that meet specific condition.
DEFINE	Specify options to specific variables.
BREAK	Add a break for each unique value of the variable you specified.
RBREAK	Report statistics at the top/bottom of report
COMPUTE	Create a compute block.
ENDCOMPUTE	All variables used to compute the new variable need to be listed in the COLUMN statement.

- DEFINE options

General Syntax

```
proc report data=dataset nowindows;
    column list-of-variables;
    define variable-name / <options> 'column-header';
run;
```

Option	Description
ANALYSIS	Calculate statistics for the variable. This is the default usage for numeric variables, and the default statistic is sum.
DISPLAY	Create one row for each observation in the dataset. This is the default usage for character variable.
ACROSS	Create a column for each unique value of the variable. Combine observations by the variable and provide a sum for numeric variable or a frequency for character variable.
GROUP	Create one row for each unique value of the variable. By default, grouping by character variables produces the sum of numeric values.
ORDER	Order the rows by ascending order of the variable (default). ORDER DESCENDING: Order by descending order.
CENTER / LEFT / RIGHT	Center, left, or right alignment.
FORMAT=	Apply standard or user-defined formats.
PAGE	Put the variable on a separate page.
WIDTH=	Provide extra space. Character: The default spacing is the length of the variable. Numeric: The default spacing is 9.
COMPUTED	Create a new variable whose value you calculate is a compute block.

- (R)BREAK

General Syntax

```
proc report data=dataset;  
  column list-of-variables;  
  define variable-name / <options> 'column-header';  
  break location variable-name / <options>;  
  rbreak location / <options>;  
run;
```

- LOCATION: BEFORE or AFTER, depending on whether you want the break to precede or follow the particular section of the report.
- Options: PAGE (Start a new page),
SUMMARIZE (Insert summary statistics for numeric variables)
- BREAK: One break for every unique value of the variable you specify. The variable must be listed in a DEFINE statement with either a GROUP or ORDER option.
- RBREAK: Produce only one break at the beginning or end.

- Adding statistics

Option	Description	Option	Description
MAX/MIN	Maximum / Minimum	N	Number of non-missings
MEAN	Mean	NMISS	Number of missings
MEDIAN	Median	SUM	Sum
MODE	Mode	PCTN	Percentage of observations for the group
STD	Standard deviation	PCTSUM	Percentage of a total sum represented by the group

General Syntax

```
proc report data=dataset;
  column list-of-variables;
  column var-1 N var-2, statistic ... var-k, statistics;
run;
```

- COMPUTE & ENDCOMPUTE

General Syntax

```
proc report data=dataset;
  column list-of-variables;
  define new-variable-name / computed;

  compute new-variable-name / <options>;
    computing statements
  endcomp;
run;
```

Example

Raw Data

```
* Dataset: National parks and monuments in the USA;
data park;
  input Name $21. Type $ Region $ Museums Campings;
  datalines;
Dinosaur          NM West 2 6
Ellis Island       NM East 1 0
Everglades         NP East 5 2
Grand Canyon      NP West 5 3
Great Smoky Mountains NP East 3 10
Hawaii Volcanoes   NP West 2 2
Lava Beds          NM West 1 1
Statue of Liberty  NM East 1 0
Theodore Roosevelt NP . 2 2
Yellowstone        NP West 2 11
Yosemite          NP West 2 13
;
run;
```

SAS Code

```
* PROC REPORT with only numeric
variables;
proc report data=park nowindows;
  column Museums Campings;
run;
```

Output

Museums	Campings
26	50

```

* ORDER;
proc report data=park nowindows;
  column Region Name Museums
  Campings;
  define Region / order;
run;

```

Region	Name	Museums	Campings
East	Ellis Island	1	0
	Everglades	5	2
	Great Smoky Mountains	3	10
	Statue of Liberty	1	0
West	Dinosaur	2	6
	Grand Canyon	5	3
	Hawaii Volcanoes	2	2
	Lava Beds	1	1
	Yellowstone	2	11
	Yosemite	2	13

```

* GROUP;
proc report data=park nowindows;
  column Region Type
  Museums Campings;
  define Region / group;
  define Type / group;
run;

```

Region	Type	Museums	Campings
East	National monument	2	0
	National park	8	12
West	National monument	3	7
	National park	11	29

```

* ACROSS & GROUP;
proc report data=park nowindows;
  column Region
  Type, (Museums Campings);
  define Region / group;
  define Type / across;
run;

```

	Type			
	National monument		National park	
Region	Museums	Campings	Museums	Campings
East	2	0	8	12
West	3	7	11	29

```

* (R)BREAK options;
proc report data=park nowindows;
  column Name Region
           Museums=museums_mean
  Campings;
  define Region / order;
  define museums_mean / mean
    'Mean number of museums'
  format=4.2;

  break after Region / summarize;
  rbreak after / summarize;
run;

```

Name	Region	Mean number of museums	Campings
Ellis Island	East	1.00	0
Everglades		5.00	2
Great Smoky Mountains		3.00	10
Statue of Liberty		1.00	0
	East	2.50	12
Dinosaur	West	2.00	6
Grand Canyon		5.00	3
Hawaii Volcanoes		2.00	2
Lava Beds		1.00	1
Yellowstone		2.00	11
Yosemite		2.00	13
	West	2.33	36
		2.40	48

```

* Produce N for each Region and (MEAN
and STD) for each Region x Type
category;
proc report data=park nowindows;
  column Region N Type ,
           (Museums Campings), (mean std);
  define Region / group;
  define Type / across;
run;

```

		Type							
		National monument				National park			
		Museums		Campings		Museums		Campings	
Region	N	mean	std	mean	std	mean	std	mean	std
East	4	1	0	0	0	4	1.4142136	6	5.6568542
West	6	1.5	0.7071068	3.5	3.5355339	2.75	1.5	7.25	5.5602758

```

* Compute variables;
proc report data=park nowindows;
  column Name Region Museums
  Campings Facilities Note;
  define Museums/ analysis sum
  noprint;
  define Campings/ analysis sum
  noprint;
  define Facilities/ computed

  "Campings/and/Museums";
  define Note / computed;

  compute Facilities;
    Facilities = Museums.sum +

  Campings.sum;
  endcomp;

  compute Note / char length=10;
    if campings.sum=0
    then Note ="No Camping";
  endcomp;
run;

```

Name	Region	Museums	Campings	Campings and Museums	Note
Dinosaur	West	2	6	8	
Ellis Island	East	1	0	1	No Camping
Everglades	East	5	2	7	
Grand Canyon	West	5	3	8	
Great Smoky Mountains	East	3	10	13	
Hawaii Volcanoes	West	2	2	4	
Lava Beds	West	1	1	2	
Statue of Liberty	East	1	0	1	No Camping
Theodore Roosevelt		2	2	4	
Yellowstone	West	2	11	13	
Yosemite	West	2	13	15	

6.3. Writing Simple Custom Reports

- Useful either when reporting your result as filled in as a complete sentence or when you want one page per observation.
- FILE statement: Create a report.
- PUT statement
 - List, column, or formatted style
 - No need to worry about putting \$ after character variable.
 - Control spacing with the same pointer controls that INPUT statement uses.
(cf. 2.4. Modifiers and Pointers)
 - In addition to printing variables, you can insert text strings by simply enclosing them in quotation marks.

Example

Raw
Data

Obs	Name	Type	Region	Museums	Campings
1	Dinosaur	National monument	West	2	6
2	Ellis Island	National monument	East	1	0
3	Everglades	National park	East	5	2
4	Grand Canyon	National park	West	5	3
5	Great Smoky Mountains	National park	East	3	10
6	Hawaii Volcanoes	National park	West	2	2
7	Lava Beds	National monument	West	1	1
8	Statue of Liberty	National monument	East	1	0
9	Theodore Roosevelt	National park		2	2
10	Yellowstone	National park	West	2	11
11	Yosemite	National park	West	2	13

SAS
Code

```
data _NULL_;
  set park;

  Total = Museums + Campings;

  file "C:\Users\jl4201\Desktop\P6110\SAS\Chapter 6\Report.txt" print;
  title "National parks and monuments in the USA";

  put @5 Name "is a " Type "in " Region "region."
    / @5 "The number of Museums in " Name "is " Museums ","
    / @3 "and the number of camp grounds in " Name "is " Campings "."
    / @5 "That is, there are " Total "facilities in " Name "." //;

  PUT _PAGE_;

run;
```

Output	<pre>1 National parks and monuments in the USA 21:59 Wednesday, September 27, 2017 Dinosaur is a National monument in West region. The number of Museums in Dinosaur is 2 , and the number of camp grounds in Dinosaur is 6 . That is, there are 8 facilities in Dinosaur . National parks and monuments in the USA 2 21:59 Wednesday, September 27, 2017 Ellis Island is a National monument in East region. The number of Museums in Ellis Island is 1 , and the number of camp grounds in Ellis Island is 0 . That is, there are 1 facilities in Ellis Island National parks and monuments in the USA 11 21:59 Wednesday, September 27, 2017 Yosemite is a National park in West region. The number of Museums in Yosemite is 2 , and the number of camp grounds in Yosemite is 13 . That is, there are 15 facilities in Yosemite .</pre>
---------------	--
