

22S:166 Computing in Statistics

Proc tabulate

Lecture 21
Nov. 7, 2008

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Proc tabulate

- displays descriptive statistics in tabular format
- can create variety of tables ranging from simple to complex and highly customized
- computes many of same statistics reported from *proc means* and *proc freq*
- flexibility in classifying values of variables and establishing a hierarchical relationship between variables
- mechanism for labeling and formatting variables and procedure-generated statistics

Example 1

from
http://ftp.sas.com/techsup/download/sample/base/tabulate/tabformat_classvar

PROC TABULATE Sample

USAGE: User would like to format the CLASS variables and ANALYSIS variables.

METHOD: Use a FORMAT statement to format the CLASS variables. Use the format modifier on the TABLE statement to format the analysis variables.

DATE CREATED: 2-19-97

SAMPLE CODE:

```
data sales;
  input name $ region $ product $ sales;
  cards;
SMITH  A  CANDY  22000.
SMITH  A  CHIPS  10000.
JONES  A  CANDY  25000.
JONES  A  CHIPS  5000.
JOHNSON B  CANDY  12000.
JOHNSON B  CHIPS  15000.
ADAMS  B  CANDY  10000.
ADAMS  B  CHIPS  8000.
;

proc format;          /* Create user-defined format */
  value $fmtx 'A'='CARY'
              'B'='RALEIGH';

proc tabulate data=sales;
  /*-----*/
  /* Use FORMAT stmt. to assign format to CLASS variable */
  /* Use *F= to assign a format to an ANALYSIS variable */
  /*-----*/
  format region $fmtx.;
  class name region;
  var sales;
  table region*name, sales*(sum n)*f=comma8.;
run
```

SAMPLE OUTPUT:

```

-----
|           |           | SALES |           |
|           |           |-----|           |
|           |           | SUM   | N         |
|-----+-----+-----+-----|
| REGION | NAME |           |           |
|-----+-----+-----+-----|
| CARY   | JONES | 30,000 | 2         |
|           | SMITH | 32,000 | 2         |
|-----+-----+-----+-----|
| RALEIGH | ADAMS | 18,000 | 2         |
|           | JOHNSON | 27,000 | 2         |
|-----+-----+-----+-----|
-----

```

Example 2

PROC TABULATE Sample

USAGE: User has data derived from a multiple choice questionnaire. They would like to get frequency counts of the response for each question.

METHOD: Manipulate the data so that TABULATE receives one CLASS variable for responses instead of four. Also, create a new answer variable. Place both variables on the CLASS statement.

DATE CREATED: 2-19-97

SAMPLE CODE:

```

data old;
input q1 $ q2 $ q3 $ q4 $;
cards;
A B C D
E F A E
C B B A
B A D E
E F A B
A A A C
F E A E
;

data new;
  set old;
  q='Question 1'; ans=q1; output;
  q='Question 2'; ans=q2; output;
  q='Question 3'; ans=q3; output;
  q='Question 4'; ans=q4; output;
drop q1-q4;
run;

proc tabulate data=new format=1.0;
class q ans;
table q=' ', ans='CHOICES'*n' / misstext='0';
run;

```

SAMPLE OUTPUT:

```

-----
|           | CHOICES | | | | | |
|           |-----|
|           | A|B|C|D|E|F|
|-----+-----+-----+-----|
|Question 1 | 2|1|1|0|2|1|
|-----+-----+-----+-----|
|Question 2 | 2|2|0|0|1|2|
|-----+-----+-----+-----|
|Question 3 | 4|1|1|1|0|0|
|-----+-----+-----+-----|
|Question 4 | 1|1|1|1|3|0|
|-----+-----+-----+-----|
-----

```

Example 3

```
options ls=72;

data timerec;
  input employee $ week $ phase $ hours;
  cards;
Chen 11SEP89 Analysis 8
Chen 11SEP89 Analysis 7
Chen 11SEP89 Coding 2.5
Chen 11SEP89 Testing 8
Chen 11SEP89 Coding 8.5
Chen 11SEP89 Testing 6
Chen 11SEP89 Coding 4
Stewart 11SEP89 Coding 8
Stewart 11SEP89 Testing 4.5
Stewart 11SEP89 Coding 4.5
Stewart 11SEP89 Coding 10.5
Stewart 11SEP89 Testing 10
;
run;

proc tabulate data=timerec format=8.1;
  class employee week phase;
  var hours;
  table week, employee all, sum*hours=' *(phase all);
  table week, employee all, pctsum*hours=' *(phase all);
  keylabel sum='Total Hours'
  pctsum='Percentage of Hours';
  title 'Summary of Project Hours';
run;
```

Summary of Project Hours

```
week 11SEP89
```

	Total Hours			
	phase			
	Analysis	Coding	Testing	All
employee				
Chen	15.0	15.0	14.0	44.0
Stewart	.	23.0	14.5	37.5
All	15.0	38.0	28.5	81.5

```
week 11SEP89
```

	Percentage of Hours			
	phase			
	Analysis	Coding	Testing	All
employee				
Chen	18.4	18.4	17.2	54.0
Stewart	.	28.2	17.8	46.0
All	18.4	46.6	35.0	100.0