

PROC UNIVARIATE and PROC TABULATE - A Powerful Duo to Produce Descriptive Tables including Non-parametric Estimates

Barbara Schneider, Department for Medical Statistics, University of Vienna

ABSTRACT

Comprehensive tables including non-parametric estimates (e.g. median quantiles, interquartile range) are of major concern in descriptive analyses.

PROC UNIVARIATE within the SAS® System is an excellent tool for computing various parametric and non-parametric estimates resulting in extensive output listings. In addition attractive comprehensive overview tables of some of these estimates are very helpful. PROC TABULATE allows overview tables in various styles but unfortunately does not support non-parametric estimates. In this presentation a SAS macro is suggested combining these two PROCs to produce the desired tables.

OBJECTIVES

In any study we will usually want to summarize some of the data in a simple way. Sometimes this will be as far as the statistical analysis goes.

For categorial data frequency tables or bar diagrams can be used to condense information. Within the SAS System this can be performed applying PROC FREQ, PROC CHART or PROC GCHART.

The distribution of ordinal or quantitative data can be visualized using e.g. box-and-whisker plots. In addition to graphical presentation measures of location and variation (parametric and non-parametric) have to be enclosed.

The goal is to produce descriptive tables of the following shape:

VARIABLES	STATISTICS			
VARIABLE1				
VARIABLE2				
...				
VARIABLEn				

Though PROC TABULATE within the SAS System allows this shape of table it unfortunately does not support non-parametric estimates such as median, quantiles, interquartile range etc. PROC UNIVARIATE computes parametric and non-parametric estimates but does not automatically produce tables of the desired style. The following program creates a comprehensive table using various DATA STEP commands, the appreciated output facility of PROC UNIVARIATE and PROC TABULATE for displaying the table.

Emphasis was laid on keeping the program as general as possible.

EXAMPLE

To illustrate the program the FITNESS data set is used. Various measurements on men involved in a physical fitness course are taken:

Variables Label

AGE	Age in years
GROUP	Experimental group
MAXPULSE	Maximum heart rate
OXYGEN	Oxygen consumption
RSTPULSE	Heart rate while resting
RUNPULSE	Heart rate while running
RUNTIME	Min. to run 1.5 miles
WEIGHT	Weight in kg

PROGRAM

```
/* Definition and customizing block -- begin */
```

```
options ls=120 nolabel nonumber;
title "TAB.1";
```

```
%let indata = sasuser.fitness; /* input data set */
%let vars = age weight oxygen runtime; /*list
of variables-order to be displayed in the final table*/
```

```
%let by=group; /* %let by=; if there is no BY- GROUP
processing */
```

```
/* statistics to be computed – syntax suitable for PROC
UNIVARIATE */
```

```
%let stats =
N=N MEAN=MEAN STD=STD MEDIAN=MEDIAN
Q1=Q1 Q3=Q3 QRANGE=QRANGE MIN=MIN
MAX=MAX;
```

```
/* statistics to be displayed in the final table */
```

```
%let statvar1 = n;
%let statvar = mean std median q1 q3 qrangle min max;
/* Definition and customizing block -- end */
```

```
%let cvars = "&vars"; /* used in the following data step -
for variable selection */
```

```
data trans; /* for each observation variables (columns)
are transposed into observations (rows) */
set &indata;
array w &vars;
do over w;
wert=w;
variable=scan(&cvars,_i_);
index=_i_; /* necessary to keep the sequence of the
variable- list */
output;
end;
run;
```

```
proc sort data=trans;
by &by index;
run;
```

```
proc univariate data=trans noprint;
var wert;
output out=o &stats;
by &by variable notsorted;
run;
```

```
data o;
format /* formats for statistical estimates */
n 7.
mean 7.1
median 7.1
std 7.1
q1 7.1
q3 7.1
qrange 7.1
min 7.1
max 7.1
;
set o;
run;
```

```
title2 "DESCRIPTIVE STATISTICS";
/*
```

```
proc print data=o noobs;
format variable $varfmt.;
var variable &statvar1 &statvar;
by &by;
run;
*/
```

```
/* Display of the final table */
proc tabulate data=o format=7.2;
var &statvar1 &statvar;
class variable &by;
table variable*&by,&statvar1*(f=7.) &statvar*(f=7.2)
/ rts=20;
keylabel sum=" ";
run;
```

```
title1;
title2;
```

REMARK

This program can be easily implemented into the SAS/AF[®] modul - suitable for users who are not familiar with SAS programming tools

REFERENCES

SAS[®] Language: Reference, Version 6, First Edition
SAS[®] Procedures Guide, Version 6, Third Edition

SAS, SAS/GRAPH and SAS/AF are registered trademarks or trademarks of SAS Institute Inc.in the USA and other countries. ® indicates USA registration.

AUTHOR CONTACT

Dipl.Ing. Dr. Barbara Schneider
Institut für Medizinische Statistik
Universität Wien
Schwarzspanierstr. 17
A - 1090 Wien
Austria
+43 1 40480 238
barbara.schneider@univie.ac.at

TAB. 1
DESCRIPTIVE STATISTICS

		N	MEAN	STD	MEDIAN	Q1	Q3	QRANGE	MIN	MAX
VARIABLE	GROUP									
age	0	10	42.00	2.94	42.50	40.00	44.00	4.00	38.00	47.00
	1	10	48.30	3.16	48.50	45.00	51.00	6.00	44.00	54.00
	2	11	52.27	2.97	52.00	50.00	54.00	4.00	48.00	57.00
oxygen	0	10	49.27	6.79	47.39	44.81	54.30	9.49	39.44	60.06
	1	10	46.56	4.55	47.02	44.75	50.39	5.63	37.39	51.86
	2	11	46.39	4.44	46.08	45.12	48.67	3.56	39.20	54.63
runtime	0	10	10.36	1.66	10.46	8.65	11.63	2.98	8.17	13.08
	1	10	10.64	1.33	10.29	10.08	10.95	0.87	8.95	14.03
	2	11	10.74	1.27	10.50	9.63	11.50	1.87	8.92	12.88
weight	0	10	80.55	6.64	81.31	75.98	85.84	9.86	68.15	89.47
	1	10	78.34	7.94	78.53	73.03	83.12	10.09	66.45	91.63
	2	11	73.81	9.32	73.71	67.25	79.38	12.13	59.08	91.63

