CC03
PRODUCING SIMPLE AND QUICK GRAPHS WITH PROC GPLOT

Sheng Zhang, Xingshu Zhu, Shuping Zhang, Weifeng Xu, Jane Liao, and Amy Gillespie
Merck and Co. Inc, Upper Gwynedd, PA

Abstract

PROC GPLOT is a widely used procedure in SAS®/GRAPH to produce scatter, line, box, and Kaplan-Meier plots. In this paper we present step by step procedures to generate several common graphics. We highlight the use of the INTERPOL symbol option. Combining various options with PROC GPLOT, we explore the flexibility and broad functionality of this SAS/GRAPH procedure to produce graphics simply and quickly.

Introduction

SAS graph is often used to visually represent the relationship among data values. There are many useful procedures to accomplish these, including PROC GPLOT, PROC GCHART, or PROC BOXPLOT. Among those, PROC GPLOT is a popular and important procedure for generating quality graphs. It allows users to enhance the graphic appearance by controlling the axes and the positions of many elements in the graph panel. This paper starts by presenting a simple SAS code for creating basic scatter plots. The code is then extended for a more enhanced variety of SAS plots, including line, box, mean plot with bars, and Kaplan-Meier plots, following step by step procedures.

Dataset

Data being used for illustration purposes in this paper is a time corresponding antibody titer measurements dataset, called ANALYZE.sas7bdat. A snap shot of the partial sample dataset is listed below:

an	treatment	age	weight	interval	censor	phase	time	titer
1	Drug 50mg	31	79	389	1	2	383	403.50
2	Placebo	46	63	864	0	5	714	330.51
3	Placebo	51	71	806	1	5	690	339.38
4	Drug 50 mg	34	84	200	1	1	110	256.93
5	Placebo	48	73	217	1	1	208	195.09
6	Drug 50 mg	46	78	455	1	3	331	469.77
7	Drug 50 mg	39	91	492	1	4	291	473.79

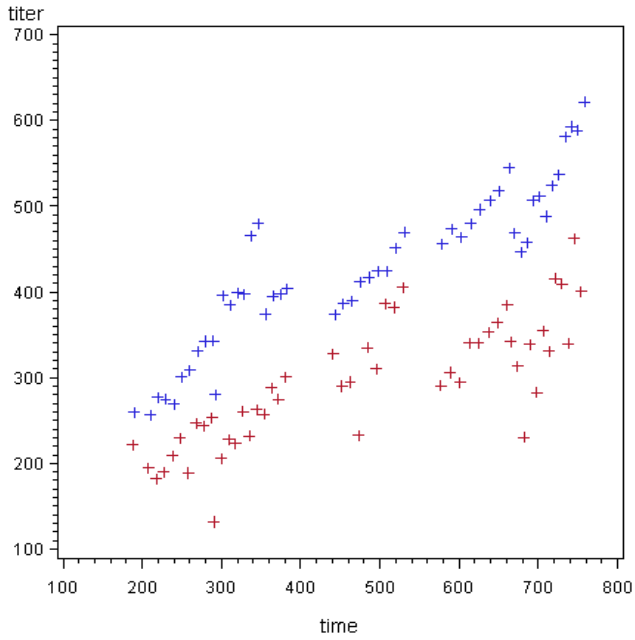
The ANALYZE dataset contains nine variables and 100 records. The labels for each variable are as follows:

- an – Subject identification number.
- treatmnt – Treatment groups description.
- age – Subject age
- weight – Subject weight
- interval – Time in the trial
- censor – Censored variable
- phase – Treatment phase
- time – Time in hours corresponding with the titer measurement

- titer – Antibody titer

(1) Scatter plot

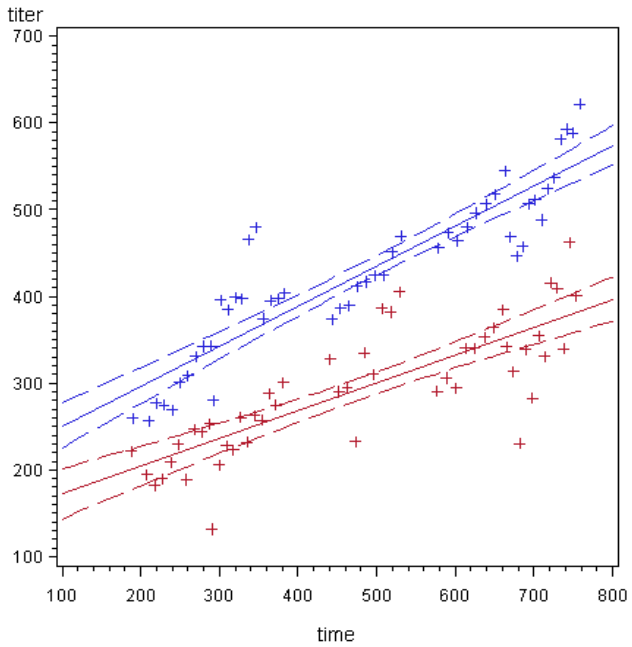
A scatter plot puts ordered pairs in a coordinate plane, showing the correlation between two variables. The scatter plot can be generated using the SAS procedure PROC GPLOT. There are three variables used in the PLOT statement, y-variable "Titer", x-variable "Time", and group-variable "Treatmnt", where ANALYZE is the name of the dataset.




```
Goptions reset=all;
proc gplot data= ANALYZE;
   plot titer*time=treatmnt;
run;
quit;
```

Note: Graphic option statements have carry-over effects. It is always recommended that option values are reset to default by using 'Goptions reset=all' before any SAS plot procedures.

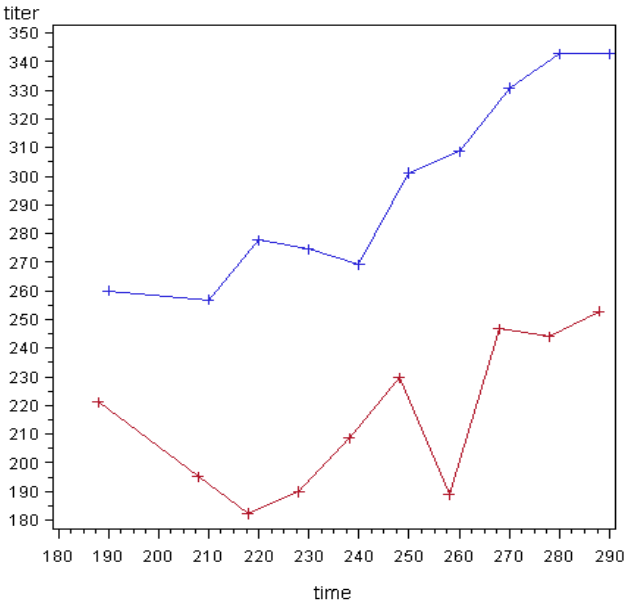
One of most important applications of scatter plots is to explore the relationship of two variables, for example, to fit the regression line with 95% C.I. limit. This can be accomplished by adding INTERPOL=RLCLM option in the SYMBOL statement, as shown below.




```
symbol1 interpol=RLCLM;
proc gplot data=ANALYZE;
   plot titer*time=treatmnt;
run;
quit;
```

(2) Line Plot

A line plot looks like a scatter plot in that it plots each individual data point on the graph. The difference is that a line graph connects data points by means of different interpolation methods. By adding the SYMBOL statement with different INTERPOL= option in PROC GPLOT, many useful graphs can be generated. INTERPOL= can also be substituted by letter I=. Our second example with INTERPOL= option is the line plot.




```
symbol1 interpol=join;
proc gplot data=ANALYZE;
   where phase eq 1;
   plot titer*time=treatmnt;
run;
```

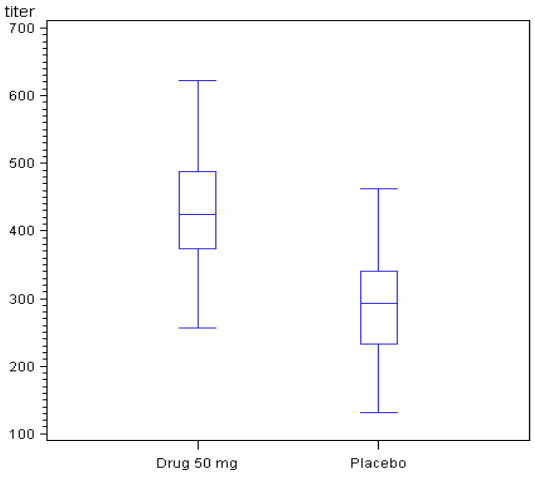
Note:
Here we subset dataset ANALYZE with phase equal to 1 to decrease data points density, and to better illustrate different interpolation methods.

Listed below are some additional options can be used to connect data points.

- INTERPOL=JOIN connects data points with straight line.
- INTERPOL=SPLINE connects data points with smooth line.
- INTERPOL=SM<nn> connects the data points with smooth line in degree of smoothness nn from 0 to 99. The bigger the nn is, the smoother the line.

(3)Box Plot

A box plot displays numerical data through five summary statistics (minimum, Q1, median, Q3, maximum). It shows the differences between groups by indicating the spread, skewness and outliers of the data. The SAS procedure PROC GPLOT with INTERPOL=BOX option in symbol statement draws the box plot of "titer" in ANALYZE dataset.




```
symbol interpol=boxt;
proc gplot data=ANALYZE;
   plot titer*treatmnt;
run;
quit;
```

Notes:
•OFFSET = (a, b) option may use in AXIS statement for X-axis to enhance the graph.
•Box plot can also be generated by different SAS procedures, such as PROC UNIVARIATE and PROC BOXPLOT.

(4) Mean plot with Bar

A variety of error bars around the means can be plotted easily with the INTERPOL option in PROC GPLOT. These include standard error bars, one standard deviation bars, or 95% confidence interval bars, as shown below.

The option INTERPOL=HILOTJ requires the outputs of three observations (representing lower, mean, upper) for each time point within a treatment group. The following code shows, in a step by step procedure, how a plot of the mean with standard error bars can be drawn.

1. output a dataset containing mean and stand error for each phase by treatment:

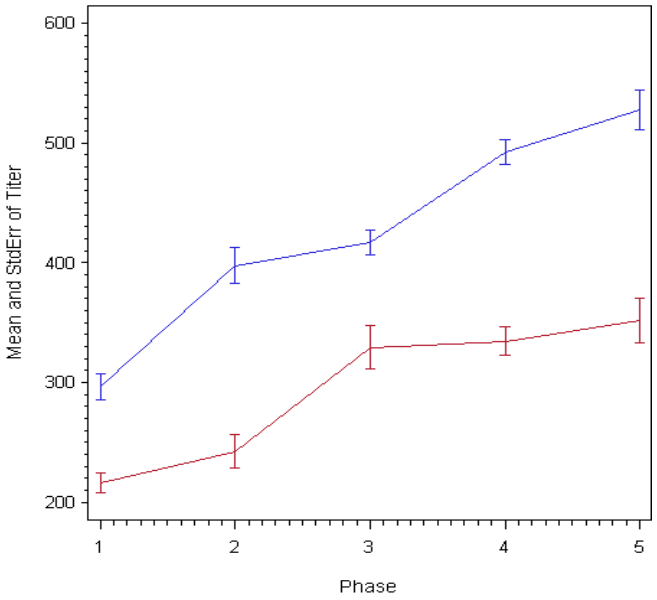
```
ods output 'Summary statistics'=stats;
proc sort data=ANALYZE out=temp_;
   by treatmnt phase;
proc means data=temp_ mean std stderr clm;
   var titer;
   by treatmnt phase;
run;
ods output close;
```

2. create a dataset containing lower, mean, upper values for each phase by treatment:

```
data plotds (keep=treatmnt phase y);
   set stats;
   y=titer_mean;              output;
   y=titer_Mean-titer_StdErr; output;
   y=titer_Mean+titer_StdErr; output;

proc sort data=plotds;
   by treatmnt phase;
run;
```

3. draw the plot using PROC GPLOT with option INTERPOL=HILOTJ:

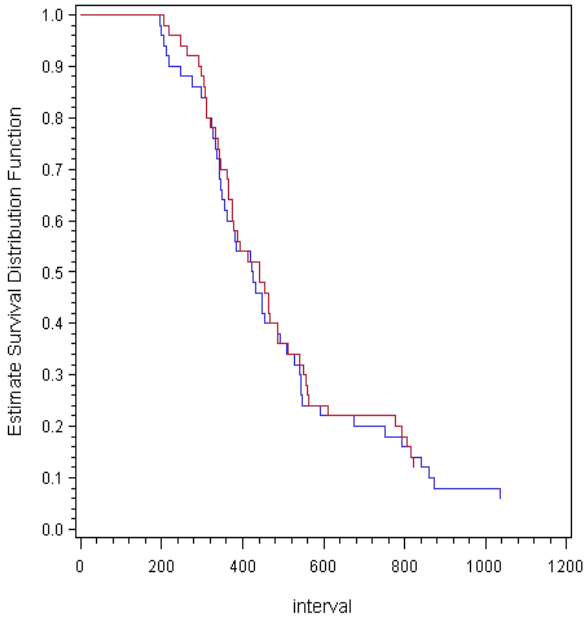



```
symbol interpol=HILOTJ;
proc gplot data=plotds;
   plot y*phase=treatmnt;
run;
```

Note:
Users can make other types of error bar plots around mean, including the one standard deviation bar or 95% C.I. bar by changing "titer_StdErr" to "titer_StdDev" or "titer_LCLM and titer_UCLM" in step 2.

(5) KM plot

A Kaplan-Meier (KM) plot is a common method to describe and graph survival characteristics. To calculate the survival function of the data, the input dataset must contain one observation per patient, and have time variables and censor variables. In our sample dataset, the time variable is Interval, and the censor variable is Censor. The KM plot can then be plotted by using the PROC GPLOT with INTERPOL=STEPLJ option in symbol statement. The code is very similar to the one used in the previous example (4). The TEMP dataset will be further described below in detail.




```
symbol i=STEPLJ;
proc gplot data=temp;
   plot survival*interval=treatmnt;
run;
```

The input dataset TEMP in the above code comes from the output of PROC LIFETEST procedure with dataset ANALYZE. The TEMP dataset has the following data structure.

Temp dataset from PROC LIFETEST

Treatmnt	Interval	Survival	SDF_LCL	SDF_UCL
Drug 50 mg	0	1.000	1.000	1.000
Drug 50 mg	197	0.980	0.866	0.997
Drug 50 mg	……	……	……	……
Drug 50 mg	1037	0.060	0.016	0.149
Drug 50 mg	1440	.	.	.
Placebo	0	1.000	1.000	1.000
Placebo	204	0.980	0.866	0.997
Placebo	……	……	……	……
Placebo	822	0.120	0.049	0.226
Placebo	1434	.	.	.

```
proc lifetest data=ANALYZE;
   by treatmnt;
   time interval*censor(0);
   SURVIVAL out=temp;
run;
```

PROC LIFETEST uses TIME statement to run a survival analysis and the temporary dataset Temp is created by the SURVIVAL statement. The dataset Temp contains all data points needed for the KM plot.

Conclusion

We have shown, in a step by step fashion, how to generate different types of SAS plots by using the PROC GPLOT procedure. Overall, five different examples have been presented by utilizing different INTERPOL options in the SYMBOL statement. The results are summarized in the table below.

The plot	Symbol INTERPOL =
Regression line with 95% CI limits	RLCLM
Straight line or Smooth line	JOIN or SPLINE, SMnn
Box plot	BOXT
Line with bars	HOLOTJ
KM plot	STEPLJ

Together with an annotate dataset and the Interpol graph option, users can easily generate customized SAS graphs with PROC GPLOT.

References

1. SAS Institute Inc., SAS 9.1.3 Help and Documentation , Cary, NC: SAS Institute Inc., 2000-2004
2. Introduction to SAS/GRAPH Software: http://www.d.umn.edu/math/docs/saspdf/gref/c01.pdf

Acknowledgements

Authors would like to thank Maryann Williams for taking the time to review and comment on the manuscript.

Trademark Information

Contact

Author Name: Sheng Zhang
Company : Merck & Co., Inc.
Address : 351 Sumneytown Pike
City: North Wales
State: PA ZIP: 19454
Email:sheng_zhang@merck.com

Author Name: Xingshu Zhu
Company : Merck & Co., Inc.
Address : 351 Sumneytown Pike
City: North Wales
State: PA ZIP: 19454
Email:xingshu_zhu@merck.com