A SAS Macro to Generate Caterpillar Plots

Guochen Song, i3 Statprobe, Cary, NC

ABSTRACT

Caterpillar plots are widely used in meta-analysis and it only requires a click in software like Winbugs. However, there is no SAS procedures that generate such graphs directly, and SAS users usually get around of this by applying PROC GPLOTS twice. This paper presents a macro to draw this type of graph completely using the annotate facility.

INTRODUCTION

In meta-analysis, caterpillar plots are widely used as graphical presentation of the results. The graph usually includes point estimates, confidence intervals, text description for these statistics from several studies and label of the studies. These type of presentations can also be used for subgroup analysis or (e.g., gender, study-center, or blood pressure groups at baseline) within one clinical trial. It is also sometimes referred as forest graph.

In software like Winbugs, generating caterpillar graph is a standard option and can be done by click the pull down menu. However, there are no existing SAS procedures in SAS Graph can complete this task in a straightforward way. It is commonly seen that two PROC GPLOTS are applied in turn, one draws the lines and one draws the dots, and a PROC GREPLAY then will put the two graphs together. These exercise requires the options and axis of the two PROC GPLOTS to be exactly the same, and if for some reason (e.g., values are out of range) the systematic parameters of the two PROC GPLOTS varies a little, the result will not be as expected. This paper presents a macro to draw this type of graph through the annotate facility, which will have fewer restrictions comparing to the two PROC GPLOTS approach.

CODE, RESULTS AND EXPLANATION

Any data in the format of estimation, confidence interval (CI) boundary, and description can be presented in this type of graph. The mocking data contains the following variables:

- lb: Lower bound of CI
- ub: Upper bound of CI
- mid: Point Estimation
- Var: Text description (optional)
- Pop: weight of the mid-point (optional, if missing, set as a constant)

Note that the variable mid does not have to be equal to (lb+ub)/2. For example, if the axis is log based, the point estimation will not be exactly the mid-point of the CI.

The following data step makes a mock dataset for this presentation, and it does not have actual meaning.

```sas
data test;
  lb=1.4;
  ub=2;
  mid=1.8;
  pop=1;
  var='Raleigh';
  output;
  lb=1.4;
  ub=4;
  pop=0.8;
  var='Dhuram';
  mid=2.8;
  output;
  lb=1;
  ub=3;
  mid=2;
```
The macro starts from here the following line. We’ll put the input dataset and the output RTF graph file as parameters.

```
%macro drawgrf(inpdata=test,outf=testt);
```

First set up the Goptions.  Note that this can be set before the macro as well.

```
goptions gunit=pct htext=2  htitle=3 device=emf gsfmode=replace gsfname=grafout
   colors=(black green yellow blue)  reset=symbol ftext='Helvetica';
filename grafout "u:\data\sas pharm\testg.emf";
```

The following dataset step generate a new variable t which will be the text description of the statistics. It is in the format of estimation (lower bound, upper bound).

```
data test;
   set &inpdata;
   t=put(mid,3.2)||'('||put(lb,3.1)||','||put(ub,3.1)||')';
run;
```

Invoke the annotating facility:

```
%annomac;
```

The data step is the key part of this paper and it generates the dataset which annotates the graph.

```
data mydata;
   %system(2,2,2);
   length text $200.;
   do k=1 to 5;
      set test point=k;  *READ IN THE DATASET;
      y=k*10+15;
      *draw the confidence interval line;
      %line(x1=lb,y1=y,x2=ub,y2=y,width=1,lintyp=1,colin=black);
      *study description;
      %label(x1=-0.3,y1=y,txt=var,ang=0,rot=0,hgt=1.7,font='Helvetica',pos=>);
      *statistics description;
      %label(x1=4.2,y1=y,txt=t,ang=0,rot=0,hgt=1.7,font='Helvetica',pos=>);
      *draw estimation points;
      %slice (x1=mid,y1=y,rad=pop,pattern=solid,ang=0,rot=360);
   end;
   *a vertical line;
   %line(x1=1,y1=15,x2=1,y2=75,width=0.2,lintyp=1,colin=blue);
   *axis;
   %line(x1=0 ,y1=15,x2=4,y2=15,width=2,lintyp=1);
   do j=0 to 4 by 0.5;
```
The data step used a pointer to read in data instead of directly using “SET”. This is not actually necessary, but because each annotate macro includes an output statement in the end, the number of observations of the dataset grows each time when an annotate facility macro is invoked. Without the pointer, the dataset will be much larger, although the graph is just fine (i.e., without the pointer, the dataset will repeat itself many times, which is similar to drawing the exact same graph at the same place multiple times). Note that a “STOP” statement is required when a pointer is used, or the data step will not know where to end. Also, one needs to change the number “5” in statement “do k=1 to 5” to the number of observations of the dataset, which can also be easily changed to a macro variable through PROC SQL.

The code used several annotate macros: %system is to tell SAS which coordinate system should be used. Using (2,2,2) will tell SAS to use the actual value of data to draw the graph. This is critical for this procedure because this will ensure that any type of data (even if some data are negative or very large) will be in the graph area. %line is to draw line segments, %label is to add texts, %slice is to draw a filled circle (and we used it to draw different size of dots) and %bar2 is to draw a frame.

Unfortunately, it requires some experiments of the place of axis, text descriptions, titles etc to get the desired results. However, after one is familiar with the parameters, the task is quite straightforward.

After the data step is completed, the graph is essentially done and one can use the GANNO procedure to draw the graph using this data. The datasys option tells SAS to use the dataset as the coordinate system. The Macro ends here.

```
proc ganno annotate=mydata datasys;
run;
%mend;
```

The following statement invokes the macro and the result generated from this statement is shown below:

```
%drawgrf;
```
CONCLUSION

As shown above, the annotate facility can be utilized to generate the caterpillar graph and the procedure is quite straightforward.

REFERENCES


CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Guochen Song
Enterprise: i3 Statprobe
Address: 1001 Winstead Dr
City, State ZIP: Cary, 27513
Work Phone: (919)-677-6040
Fax: 
E-mail: Guochen.song@i3statprobe.com
Web: www.ingenix.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.
Other brand and product names are trademarks of their respective companies.