

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
 Maths 2024; SP-9(1): 270-275
 © 2024 Stats & Maths
<https://www.mathsjournal.com>
 Received: 05-11-2023
 Accepted: 09-12-2023

Garrepally Prasad
 Senior Statistical Programmer,
 Sigma-Chi consulting (India)
 Pvt. Ltd., Madhurawada,
 Visakhapatnam, Andhra
 Pradesh, India

Pendkar Pooja
 Senior Statistical Programmer,
 Sigma-Chi consulting (India)
 Pvt. Ltd., Madhurawada,
 Visakhapatnam, Andhra
 Pradesh, India

Mokara Kiran Kumar
 Senior Statistical Programmer,
 Sigma-Chi consulting (India)
 Pvt. Ltd., Madhurawada,
 Visakhapatnam, Andhra
 Pradesh, India

Corresponding Author:
Garrepally Prasad
 Senior Statistical Programmer,
 Sigma-Chi consulting (India)
 Pvt. Ltd., Madhurawada,
 Visakhapatnam, Andhra
 Pradesh, India

Different approach to create Mean SD plot using multiple procedures

Garrepally Prasad, Pendkar Pooja and Mokara Kiran Kumar

Abstract

Graphical Representation of Mean Measurement over Time with error (e.g., SD, SE, and confidence intervals like LCLM/UCLM 95% Lower/Upper Confidence Limit for mean) Clinical domain are used to compare the outcomes of different studies or treatment groups over a period of time. I introduce a different way to create Means with standard deviation plot by using multiple procedures like PROC GPLOT, PROC SGPLOT, PROC TEMPLATE and PROC SGRENDER. This article will demonstrate detailed steps of creating Means plot as well as compare the advantages and disadvantages of different methods. To address the challenges using the PROC GPLOT procedure in order to accomplish more comprehensive graphic outputs specifically targeting error bar plots. This graph of the mean responses over time for patients in two arms of a clinical trial. (A Treatment is a group of participants who receive an intervention or who receive no intervention, such as an experimental group and the control group. This article shows how to create this and other graphs that visualize the mean response by time for groups in a clinical trial with different methods.

Keywords: SAS SD Means Proc plots

Introduction

In clinical trial analysis handling and presenting the data is very important role. There are different ways by which data can be presented, including tables, listings, and graphs. Often times, the graphic format is the preferred one because of its direct and immediate visual impact. "A picture may convey an idea more quickly and effectively than the written words". Graphics can help to identify patterns and trends during preliminary investigations of the data, including relationships among treatments or over time. In this article, we start with simple SAS codes with different procedures for creating basic mean error bar plots. These codes are then extended for an enhanced variety of SAS plots, including multiple treatment groups and visits over a period of time with one of inferential statistics, the SD, the SE or CI ^[1-3].

Datasets

Only one datasets is used in the examples to present the graph with different procedures.

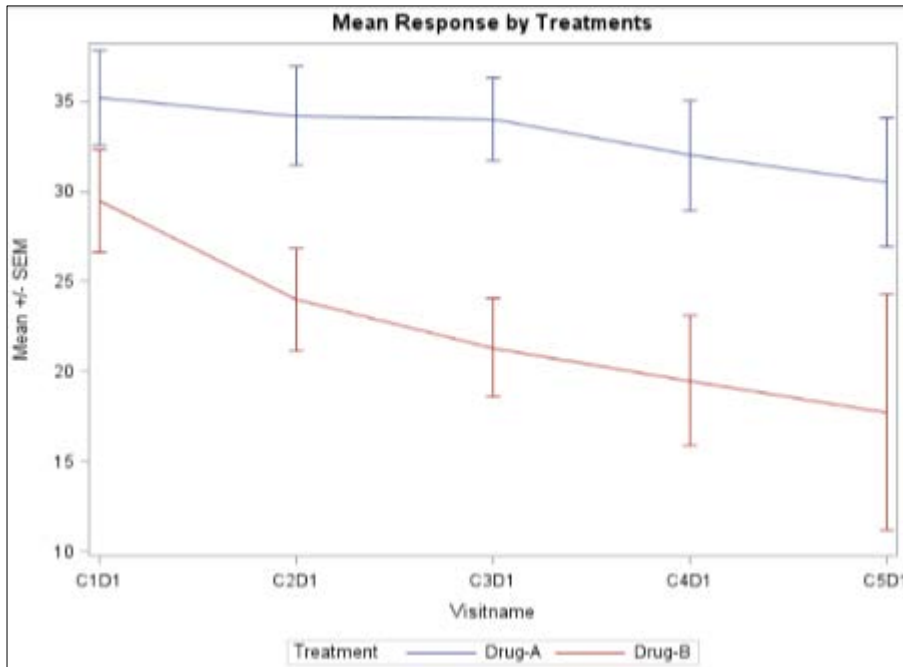
Data adrs;

```
input trt$ Subject $ res1-res5;
label Visit = 'Visitname' trt = "Treatment";
Visit='C1D1'; res=res1; output;
Visit='C2D1'; res=res2; output;
Visit='C3D1'; res=res3; output;
Visit='C4D1'; res=res4; output;
Visit='C5D1'; res=res5; output;
drop res1-res5;
datalines;
Drug-A 001 35 38 35 34.
Drug-A 002 42 40 41 39 38
Drug-A 003 40 37 36 35 33
Drug-A 004 31 31 30 31 30
```

```
Drug-A 005 28 25 28 21 21
Drug-B 006 25 20 15 10 18
Drug-B 007 39 34 28 25.
Drug-B 008 36 29 26 20 11
Drug-B 009 28 25 27 33 36
Drug-B 010 20 14 12 10 06
Drug-B 011 29 22 20 19.;
run;
```

Use the VLINE statement for mean and variation

The VLINE statement in PROC SGPLOT can summarize data across groups. When you use the RESPONSE= and STAT= option, it can display the mean, median, count, or percentage of a response variable. You can add "error bars" to the graph by using the LIMITSTAT= option. Following Cheng, the error bars indicate the standard error of the mean (SEM). The following statements create the line plot shown at the top of this article [4-7]:



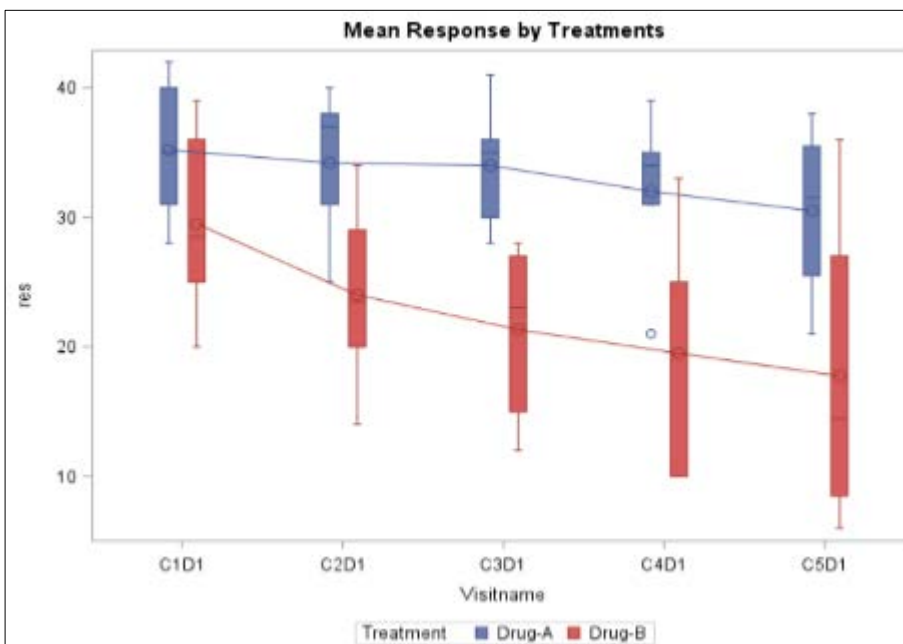
```
title "Mean Response by Treatments";
proc sgplot data=adrs;
vline Visit / response=res group=trt stat=mean
limitstat=stderr;
yaxis label='Mean +/- SEM';
run;
```

confidence limits for the mean (LIMITSTAT=CLM) or indicate the extent of one or more standard deviations (LIMITSTAT=STDDEV) [7].

Use the VBox statement for mean and variation by Visit

The VBox statement in PROC SGPLOT can summarize data across groups. When you use the CATEGORY= option is used to specify the time-like variable and the GROUP= option is used to specify the arms of the study.

VLNE statement computes the mean and standard error for Y for each value of Visit and Treatment variables. This graph shows the standard error of the mean, but you could also show



```
title "Mean Response by Treatments";
proc sgplot data=adrs;
vbox res / category=Visit group=trt groupdisplay=cluster
connect=mean clusterwidth=0.35;
run;
```

Graph summarized statistics

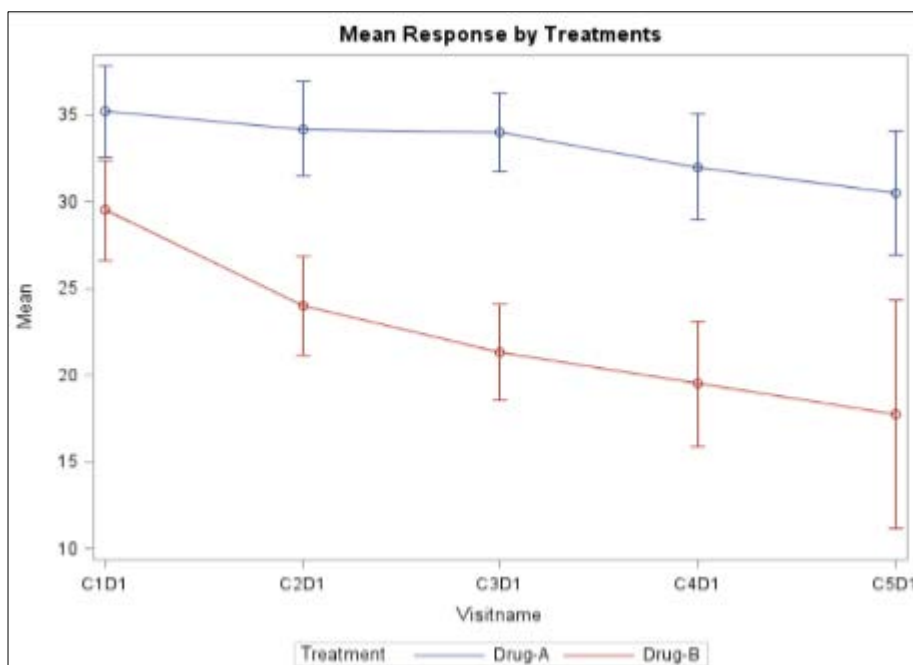
In the previous sections, the VLINE and VBOX statements automatically summarized the data for each time point and for each arm of the study. This is very convenient, but the SGPLOT statements support only a limited number of statistics such as the mean and median. For more control over the statistics, you can use PROC MEANS or PROC UNIVARIATE to summarize the data and then use the SERIES statement to plot the statistics and (optionally) use the SCATTER statement to plot error bars for the statistic [7, 8].

PROC MEANS supports dozens of descriptive statistics, but, for easy comparison, I will show how to create the same graph by using summarized data. The following call to PROC MEANS creates an output data set that contains statistics for each visit/arm combination.

```
proc means data=adrs N mean stderr stddev lclm uclm
NDEC=2;
class trt Visit;
var res;
output out=MeanOut N=N mean=Mean stderr=SEM
stddev=SD lclm=LCLM uclm=UCLM;
run;
/*Mean and +/- SEM plot*/
data Summary;
set MeanOut(where=(trt^=" " & Visit^=""));
LowerSEM = Mean - SEM;
UpperSEM = Mean + SEM;
run;
data Summary_SD;
set MeanOut(where=(trt^=" " & Visit^=""));
LowerSD = Mean - SD;
UpperSD = Mean + SD;
run;
```

Simple Graph

Creating simple a graph of summary statistics that is similar to the VLINE graph using series and scatter options.

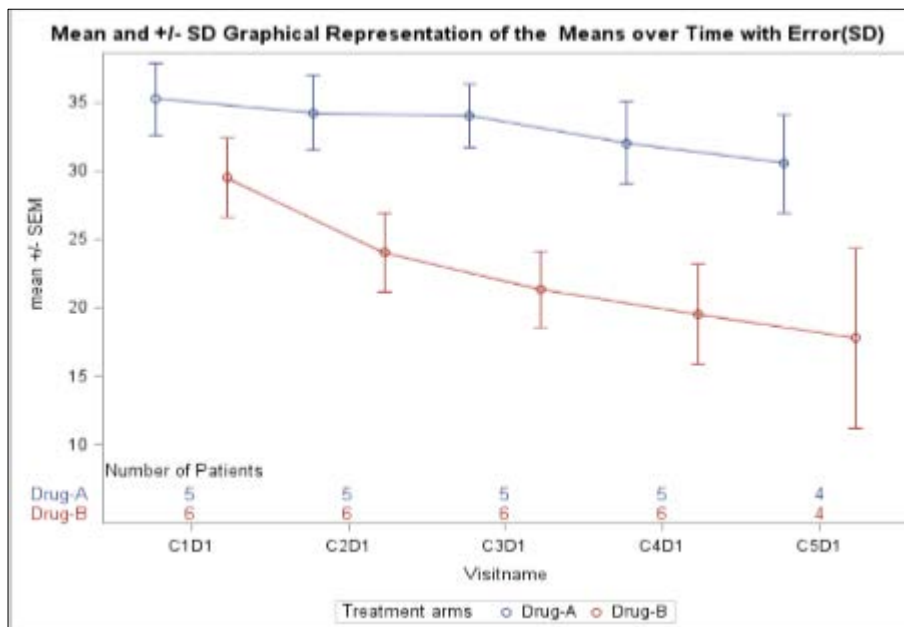


```
proc sgplot data=Summary;
series x=Visit y=Mean / group=trt;
scatter x=Visit y=Mean / group=trt
yerrorlower=LowerSEM yerrorupper=UpperSEM;
run;
```

Adding tabular information to a mean-versus-time graph

You can augment a mean-versus-time graph by adding additional information about the study at each time point. In Cheng's paper, much of the code was devoted to adding

information about the number of patients that were measured at each time point. In SAS 9.4, we can use the XAXISTABLE statement to add one or more rows of information to a graph. The output from PROC MEANS includes a variable named N, which gives the number of nonmissing measurements at each time. The following statements add information about the number of patients. The CLASS= option subsets the counts by the arm, and the COLORGROUP= option displays the text in the group colors [9, 10].



```
Title 'Mean and +/- SD Graphical Representation of the Means over Time with Error(SD)';
proc sgplot data=Summary;
scatter x=Visit y=Mean / group=trt groupdisplay=cluster
clusterwidth=0.9
yerrorlower = Lower SEM yerrorupper = UpperSEM
markercharattrs=(size=9) name='nm';
series x=Visit y=Mean / group=trt attrid=result
groupdisplay=cluster;
yaxis values = (10 to 50 by 10) label="Mean +/- SE Change in result" labelattrs=(size=8pt);
xaxis label='Visitname';
xaxistable N / location=inside class=trt colorgroup=trt
title="Number of Patients"
valueattrs=(size=10) labelattrs=(size=10);
yaxis label='mean +/- SEM';
keylegend 'nm'/title="Treatment arms";
run;
```

PROC template

To create better graphs, SAS developed Graph Template Language (GTL). GTL has many significant advantages over traditional SAS/GRAPH procedures like more control over visual attributes of the graph, less annotation and minimal coding. Statistical Graphics (SG) procedures like PROC SGPLOT, SGPANEL etc., which require minimal coding use GTL to create commonly used graphs. But, writing graphs in GTL has more advantages compared to SG procedures and traditional SAS/GRAPH procedures.

GTL code might look a little lengthy and intimidating, but once you understand how to read GTL code, it will be very useful to create not only simple graphs but also complex graphs. The purpose of this paper is to help the programmer understand and utilize various elements of GTL to create wonderful graphs.

We use two procedures to create graphs in GTL.

1. Proc template
2. Proc sgrnder

By itself, Proc template doesn't create the graph. It creates a template and we apply this template to the data using proc sgrnder

Here is the basic code that is needed in writing GTL code

```
proc template;
define statgraph template-name;
begingraph;
...SAS statements...
endgraph;
end;
run;
proc sgrnder data= template= template-name; run;
```

The common elements and their corresponding proc template statements in a SAS graph are:

Element	GTL statement
Title	Entrytitle
Footnote	Entryfoonote
Plot	Depends on the graph. E.g. scatterplot
Legend	Discretelegend
Layout	Single cell: Layout overlay Multi cell: Layout datalattice; Layout datapanel

Title and footnotes

If we have to represent a title in two lines, then we write two entrytitle statements.

E.g. entrytitle "Graphical Representation of the Means over Time with Error(SD or SE)" entrytitle "over time";

Common attriubtes of titles and footnotes that we look for in creating these graphs are alignment, text size and text wieght.

E.g. entrytitle textattrs=(size=12pt weight=normal) halign = center 'Systolic BP over time';

Option	Value
Halign	Center left right
Textattrs	Size Weight Color

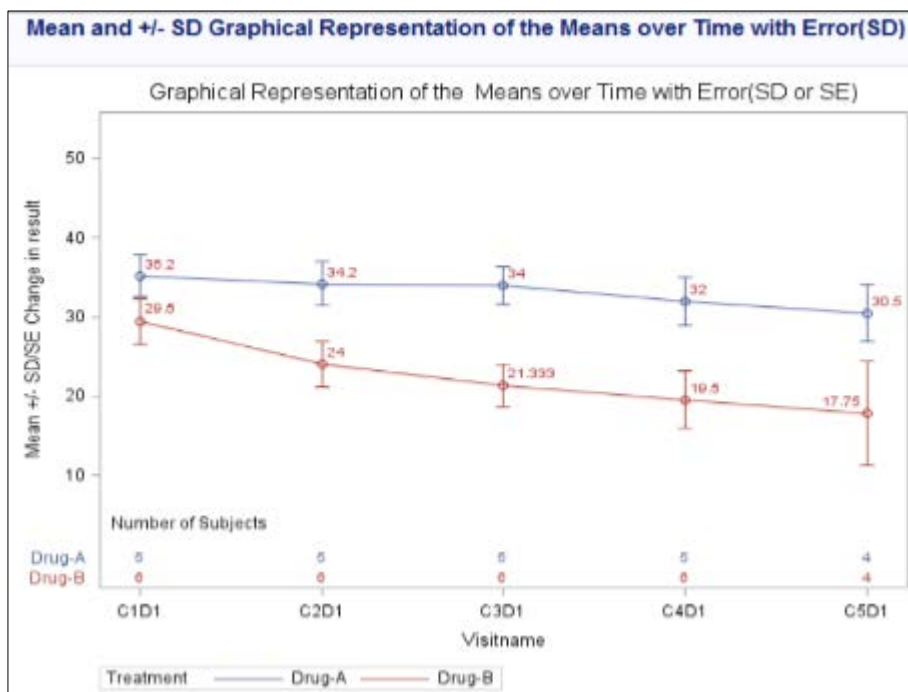
Layout

It is a single celled graph. So, we used 'layout overlay' statement. We provided axis options in this 'layout overlay' statement.

X-axis and Y-axis attributes can be manipulated using 'xaxisopts' and 'yaxisopts' for single celled graphs. OFFSETMIN: gaps at the beginning of the x-axis or y-axis OFFSETMAX: gaps at the end of the x-axis or y-axis Offsetmin and offsetmax helps us to keep required distance

from the plot area to the cell wallGrid is shown on y-axis using 'griddisplay' option. There are many

ways to provide tick values. We should choose the one that fits our data. Some important tick value options under 'linearopts' are



```
proc template;
define statgraph temp;
begingraph;
entrytitle textattrs=(size=12pt weight=normal) haln = center
'Graphical Representation of the Means over Time with
Error(SD or SE)';
layout overlay/
xaxisopts=(offsetmin=0.05 offsetmax=0.05)
yaxisopts=(griddisplay=on gridattrs=(thickness= 0.1
color=lightgrey)
linearopts=(tickvaluepriority=true
TICKVALUESEQUENCE=(START=10 END=50
INCREMENT=10)) offsetmin=0.15 offsetmax=0.1
label='Mean +/- SD/SE Change in result');
seriesplot x= Visit y= mean /group =trt name= 'Treat' ;
scatterplot x=Visit y=mean/ group=trt yerrorlower=
LowerSEM
yerrorupper=UpperSEM markerattrs=(symbol= graphdata4)
datalabel= mean datalabelattrs=(color=brown);
drawtext textattrs=(size=9pt) "Number of Subjects"
/anchor=bottomleft width=22 widthunit=percent
xspace=wallpercent yspace=wallpercent x=1 y=11
justify=center;
innermargin/align=bottom pad=0.8;
axistable x=Visit value=n / class=trt colorGroup=trt ;
endinnermargin;
discretelegend 'Treat' / title= "Treatment" titleattrs= (size=9pt
weight=normal) location= outside haln=left valn=bottom
valueattrs= (size=9pt) border=true;
endlayout;
endgraph;
end;
run;
proc sgrender data=Summary template=temp;
run;
```

Layout

1. Grid is shown on y-axis using 'griddisplay' option.

2. There are many ways to provide tick values. We should choose the one that fits our data. Some important tick value options under 'linearopts' are:

Option	Description
Tickvaluelist	Specifies the order of the tick values for a linear axis as list.
Tickvaluerotation	Specifies how the tick values are rotated on the X and X2 axes.
Tickvaluesequence	Specifies the tick values for a linear axis by start, end, and increment.
Viewmax	Specifies the maximum data value to include in the display.
Viewmin	Specifies the minimum data value to include in the display.
Tickvaluepriority	Specifies whether an axis tick specification can extend the axis data range

We may have to use these options individually or in a combination.

To increase the range of axis from the available data values, we can use 'tickvaluepriority=true' in combination with 'TICKVALUESEQUENCE' or 'TICKVALUELIST'.

E.g.1: tickvaluepriority=true TICKVALUESEQUENCE=(START=10 END=50 INCREMENT=10)

E.g.2: tickvaluepriority=true TICKVALUELIST= (12 13 14 15 16) If there are many tick values in an axis which either looks too close, If there are many tick values in an axis which either looks too close or some of tick values are missing on the axis, then you either use TICKVALUEROTATION to present values at an angle which provides space or we can decrease of tick values (e.g. tickvalueattrs=(size=7)) or a combination of both [11, 12].

Plot

We can create more than 1 plot in a LAYOUT OVERLAY statement. In this graph, we created line plot using 'seriesplot' and scatter plot with error bars using 'scatterplot'. Mean values are shown using 'datalabel= mean' in scatter plot

statement. Marker symbols in scatter plot can be modified to our requirement from the default value using 'markerattrs=(symbol=)';

Axis table

Number of subjects at each time point are shown on x-axis using 'axistable'. It can also be used on X2/Y/Y2 axis. 'Innermargin- endinnermargin' block is used to present axis table within a cell. If we want to present this outside the wall, then we must produce multi-cell graph where the axistable is drawn in a new cell. In this paper, presence of axis table outside the cell wall is explained in Panel box plot ^[11-13].

Legend

We used the name of the scatter plot 'treatment' in 'discretelegend' statement and the position details on where we want the legend to be.

Option	GTL Value
Location	Outside inside
Halign	Center left right
Valign	Center top bottom
Across	Positive- integer specifies the number of legend entries that are placed horizontally before the next row begins
Border	True False

Conclusion

The graphs created in this paper not only helps those who already know how to create graphs using traditional or SG procedures but also to those who are entirely new to graph programming. These plots may also be created in different ways. Proc SGPLOT and GTL can create high quality, customizable graphs. Proc SGPLOT is quick, easy to use, and easy to remember. GTL is more verbose but highly structured and flexible.

References

1. Nuzzo R. Statistical errors. *Nature*. 2014;506:150-1521.
2. Wasserstein RL, Lazar NA. The ASA's statement on p-values: 3. context, process, and purpose. *The American Statistician*. 2016;70:129-133.
3. Johnson VE. Revised standards for statistical evidence. *Proc Natl Acad Sci USA*. 2013;110:19313-19317.
4. Burnham KP, Anderson DR. *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Verlag; c2002.
5. Cohen AC. Estimating Parameters of Logarithmic-Normal Distributions by Maximum Likelihood. *Journal of the American Statistical Association*. 1951;46:206-212.
6. Fisher RA. *Statistical Methods for Research Workers*. 14th ed. New York: Hafner Publishing; c1973.
7. Rousseeuw PJ, Croux C. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*. 1993;88:1273-1283.
8. Velleman PF, Hoaglin DC. *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: Duxbury Press; c1981.
9. Littell RC, Freund RJ, Spector PC. *SAS System for Linear Models, Third Edition*. SAS Institute Inc; c1991.
10. Searle SR. *Linear Models*. John Wiley & Sons, New York; c1971.
11. Searle SR, Casella G, McCulloch CE. *Analysis of Variance Components*. John Wiley & Sons, New York; c1992.
12. www.sas.com
13. www.support.sas.com