

Paper 221-26

How the SAS[®] System Can Be Used for Teaching Statistics

Ass. Prof. Dipl.Ing. Dr. Barbara Schneider
 Institute for Medical Statistics, University of Vienna, Vienna, Austria

Abstract

This presentation offers ideas and examples for using the SAS[®] System in teaching statistics to students in medicine and to physicians. Two parts will be discussed. The first part describes features already available in the SAS[®] System. The second part introduces an educational application. This application is a visual demonstration of elementary statistical notion. Topics such as the influence of changing the parameters of a distribution on the shape of the distribution and the variability of a point estimator are discussed.

The demonstration is developed on a PC (Windows) using Base SAS, SAS/GRAPH[®], SAS/INSIGHT[®] and SAS/AF[®]. SAS Version 8.1

SAS code will be shown.

The presentation is aimed at people who are interested or involved in teaching statistics

Objectives

Teaching statistics can be a very tough job especially when the students are non-statisticians or mathematicians. For example physicians are interested in results from data analyses. But before using a software package one has to be familiar with at least some basic methodological concepts and ideas of probability theory and statistics. Students very often experience a course in statistics as too theoretical. Therefore it is desirable to make the course more concrete. Visualization is an excellent tool to obtain more insight on those

topics that are difficult to illustrate with pen and paper.

An application that serves as a companion for a statistical course must be user friendly. Because the attention of the application must be focused as much as possible on the content of the demonstration, the handling of the package must be very simple.

The student is desired to take an active role. He has to determine the reading time and learning time himself.

He should be able to try out things that are too difficult to try out by pen and paper or things that require too much arithmetical work. In this context simulations together with graphical presentations are powerful techniques to explain statistical concepts clearly.

Examples

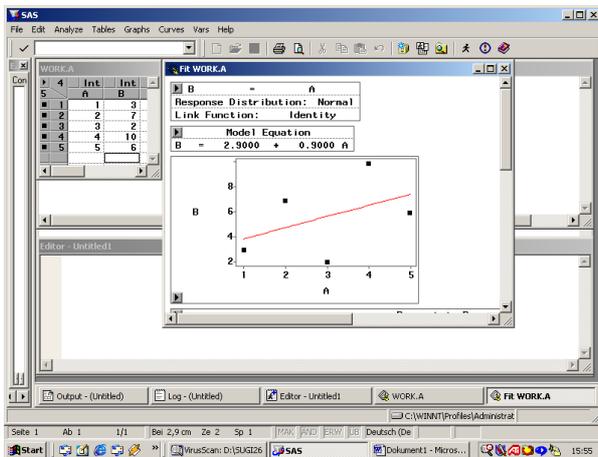
The first example introduces a feature already available in the SAS[®] System.

Example 1:

This example demonstrates how SAS/INSIGHT can be used right now in teaching basic concepts of linear regression analysis without further programming effort.

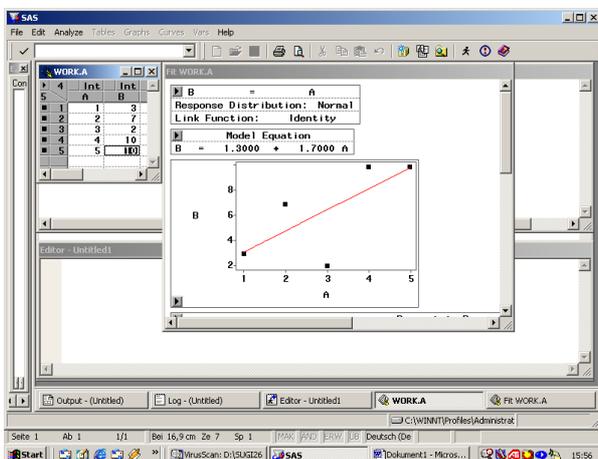
Open SAS/INSIGHT and create a new data set (see Fig.1, WORK.A): A : independent variable, B: dependent variable. Choose Analyze from the menu and select Fit(YX). Load B into the Y-Variable field and A into the X-Variable field then click OK. Resize the windows as shown in Fig.1. Make sure to have the model equation displayed.

Fig. 1



Then change the value of variable B in observation 5 from 6 to 10 and press the enter key. The students are able to observe how the regression line and simultaneously the model equation change (Fig. 2). So, students can repeat this step several times exploring the “rules of linear regression” in a dynamic way.

Fig. 2



The following examples show how SAS/AF® FRAME technology can be used for teaching

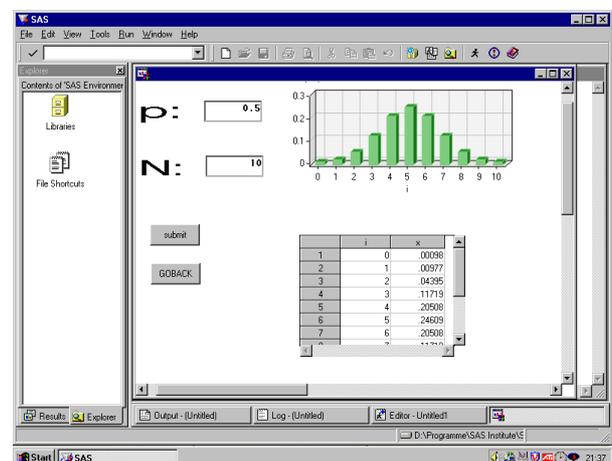
Example 2:

The influence of changing the parameters of a binomial distribution on the shape of the distribution:

A SAS program is built to construct the binomial distribution. This program is 'wrapped up' into an AF frame to make it available to students not familiar with the SAS programming language. The frame (Fig. 3) consists of 2 Version 6 (V6) input fields to transfer the parameters of a binomial distribution to the SCL code (SAS Component Language). Graphic text controls left to the input fields describe the fields. p: is the probability of the outcome of the Bernoulli trial. N: represents the number of independent Bernoulli trials. 2 pushbuttons (V8 pushbutton controls) are used to submit the SCL code or to go back to another frame (not discussed in the paper). A chart control displays the graph (probability function) in the data table object below the graph, the probability function is displayed numerically (i: number of outcomes, x: probability).

To run the application you simply have to fill in the 2 parameters and press the submit button.

Fig. 3



To control the application SCL code has to be written:

```
pushbutton1: /* refers to the pushbutton label submit */
```

```
call notify('TAB1','_SET_DATASET_', ''); /* resets the data table */
```

```

submit continue; /* in this block the SAS
code which produces the probability function is
wrapped up */
options nonumber nodate;

title;

%let p=&P; /* the macro variables are set to
the value of the input field */

%let n= &N;

data b;

label x="P(X=i)"
y="P(X<=i)"
y1="P(X>=)"
;
do i=0 to &n;

y=probbnml(&p,&n,i);
if i=0 then do;
x=probbnml(&p,&n,i);
end;
if i>0 then do;
x=probbnml(&p,&n,i)-probbnml(&p,&n,i-1);
end;
y1=x+1-probbnml(&p,&n,i) ;
output;
end;
run;

data b1;
label i="i" x="P(X=i)";
format i 5. x 6.5 ;
set b;
keep i x;
run;

title "p=&p, N=&n";
endsubmit; /* the submit block ends */

/* the following statements display the
generated dataset into the chart control and the
data table object */
CHART1.DataSet='work.b';
CHART1.viewmode='3D';

CHART1.xvariable='i';
CHART1.yvariable='x';

call notify('TAB1','_SET_DATASET_', 'work.b1');

return;

```

This application can easily be extended for the comparison of two binomial distributions.

The same technical concept can be used for demonstrating any other distribution (e.g. normal distribution)

Example3:

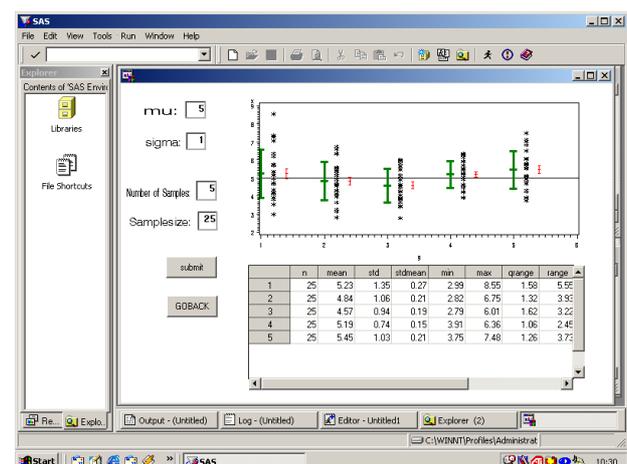
Demonstrating the variability of the mean of the normal distribution :

A simulation is used to explain how to interpret the standard error of the mean.

The SAS program that facilitates the simulation is again wrapped up into an AF frame application (Fig.4). The frame consists of 4 input field controls, facilitating the same functionality as in example 2, labeled with graphic text controls left to the input fields. mu: expectation of the normal distribution, sigma: the standard deviation, number of samples generated and the sample size within each sample. Further there are 2 pushbuttons and a graph output control which displays the generated graph. A (V6) data table object shows descriptive evaluation of the samples.

To start the application fill out the input fields and press submit. The generated samples are displayed in the graph output control. Sample values are black, means \pm standard deviation green and mean \pm standard error red.

Fig. 4



The following SCL code controls the application.

```

pushbutton1:  /* submit */

call notify('TAB1','_SET_DATASET_', '');

submit continue;

goptions device=win nodisplay goutmode=replace;

%let m=&M;

%let s= &S;

data n;
do i=1 to &NS;
do j=1 to &SL;
x=&m+&s*rannor(0);
output ;
end;
end;
run;

proc univariate data=n normal noprint;
var x;
output out=o n=n mean=mean std=std
stdmean=stdmean min=min max=max
qrange=qrange range=range probn=probn;
by i;
run;

data o;
format
n 5.
mean 7.2
std 7.2
stdmean 7.2
min 7.2
max 7.2
qrange 7.2
range 7.2
probn 5.3
;
set o;
run;

data n ;
set n;
keep g1 x g g2;
g=i;
g1=i+0.2;
g2=i+0.4;
run;

symbol1 c=green v=none i=std1t w=3;
symbol2 c=black v=star i=none ;
symbol3 c=red v=none i=std1mt ;

proc gplot data=n;
plot x*g=1 x*g1=2 x*g2=3
/ overlay vref=&m cvref=black caxis=black
ctext=black;
run;

```

```

endsubmit;

Goutput1._setGraph('work.gseg.gplot.grseg');

call notify('TAB1','_SET_DATASET_', 'work.o');

return;

term:

submit continue; /* this block is
necessary to reset the goptions especially for
the nodisplay option */
goptions reset=all;
endsubmit;

return;

```

Conclusion

SAS software, especially SAS/AF FRAME technology is an excellent tool for implementing some didactic concepts for teaching statistics and probability theory.

SAS, Base SAS, SAS/GRAPH, SAS/AF software are registered trademarks of SAS Institute Inc., Cary, NC, USA

REFERENCES

SAS[®] Language: Reference, Version 8
SAS[®] Procedures Guide, Version 8
SAS/GRAPH[®] Software: Reference, Version 8, Volumes 1 and 2
SAS[®] Component Language: Reference, Version 8
SAS/AF[®] Software: Changes and Enhancements in Version 8, Course Notes

Address for correspondence:

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