

GRAPHICAL DISPLAY OF DATA – A NONPARAMETRIC APPROACH

Shi-Tao Yeh, GlaxoSmithKline, King of Prussia, PA

ABSTRACT

A nonparametric method does not depend on the specification of a probability distribution. Nonparametric methods are almost as powerful as parametric methods. SAS[®] software system provides several nonparametric methods and procedures to display your data in a graphics format.

The nonparametric methods discussed in this paper are: 1) histogram, 2) kernel density estimates, 3) cumulative density function, 4) box plots, 5) multivariate scatter plots, and 6) SAS 9.1 experimental ODS Graphics in KDE procedure. Some sample code and illustrations are also provided in the paper.

The SAS products used in this paper are SAS BASE[®], SAS/STAT[®], and SAS/GRAPH[®] on a UNIX[®] platform and the SAS System for Windows[®] V9.1.

INTRODUCTION

The traditional statistical influence and hypothesis test is called the parametric test because it depends on the specification of a probability distribution, whereas the nonparametric method does not. The parametric statistics and statistical inferences are based on the sampling distribution of a particular statistic. Normal distribution is a well established parametric sampling distribution family. If the data collected, or the variables of interest, are not known, then a nonparametric method or distribution free approach may be a good starting point for the data analysis.

SAS software provides procedures for nonparametric methods; such as the BOXPLOT procedure, histogram plot in the UNIVARIATE procedure, the KDE procedure and the LOESS procedure. These procedures compute the statistic estimates and save it as SAS datasets or ODS tables for further processing into high resolution graphics. The BOXPLOT procedure creates a box-and-whisker plot of measurements. The HISTOGRAM plot in the UNIVARIATE procedure creates histograms with features of superimposing parametric and nonparametric density curve estimates. The KDE procedure computes either a univariate or bivariate kernel density estimation. It computes the density estimates by smoothing the probability density function from observed data. The LOESS procedure performs a nonparametric method for estimating local regression surfaces. It requires no assumption about the parametric form of the regression surface. These procedures not only provide comprehensive support for modern nonparametric methods within the SAS system, but also create a powerful tool for data analysis and high resolution graphic tasks.

The nonparametric graphical methods and displays demonstrated in this paper are: 1) histogram, 2) kernel density estimates, 3) cumulative density function, 4) box plots, 5) multivariate scatter plots, and 6) experimental ODS Graphics.

The data used for the illustrations are from hypothetical clinical trials.

HISTOGRAMS WITH DIFFERENT MIDPOINT SIZES

A histogram is a graphical display of the distribution of variables of interest. For empirical data, a histogram provides the simplest display of the shape of the density function across the range of a variable.

In a histogram, each class interval is proportionally represented by a simple block or bar that is placed over the midpoint of the class interval. The purpose of a histogram is to graphically summarize the distribution of a univariate data set. A histogram can illustrate: 1) the location of the data, 2) the scale of the data, 3) skewness of the data, 4) presence of outliers, and 5) presence of multiple modes in the data.

A histogram is constructed by dividing the range of the data into equal-sized class intervals. For each interval, the number of data points are counted and represented by a block or bar. A histogram consists of all bars from the

data set. The class intervals can be defined arbitrarily. The scale of the class interval has an impact on histogram midpoint size and shape of the histogram. Figure 1 illustrates the histogram with different midpoint sizes. The kernel density function bandwidth parameter c is selected as 0.2, 0.5 and 1. Each bandwidth parameter c produces a kernel density function.

There are three types of y-axis presentations in a histogram: 1) frequency count of data point in each interval class, 2) the frequency data point count in a class divided by the total number of observations, and the y-axis represented by a percentage, and 3) the data point count in the class divided by the number of observations times the class width. From a probabilistic point of view, the type 3 y-axis is most akin to the probability density function.

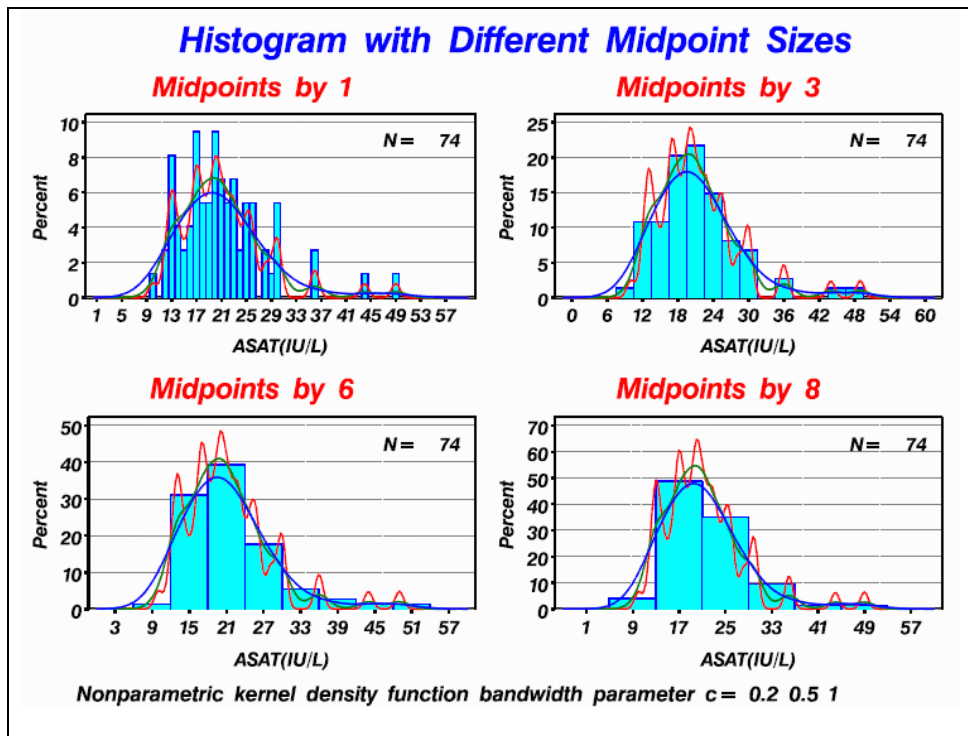


Figure 1. Histograms with Different Midpoint Sizes

Sample SAS code for Figure 1 is as follows:

```

%macro sethsg(mpoint=,name=);
  goptions reset=all device=ps600c gsfname=psname rotate=landscape gsfmode=replace
    ftext=swissbi vorigin=1.2in horigin=1in vsize=6.2in hsize=9.2in ;
  title h=4.5 c=red "Midpoints by &mpoint";
  symbol1 c=red l=1 ;
  symbol2 c=green l=2 ;
  symbol3 c=black l=3 ;
  proc univariate noprint;
    var asat;
    histogram asat/
      midpoints=9 to 60 by &mpoint
      font=swissbi
      waxis = 25
      kernel(c=0.2 0.5 1 w=20)
      wbarline = 10
      cbarline = blue
      cfill = cyan
      height = 6
      name="&name"
    ;
  inset n='N=' (3.) /noframe position=ne height=6;
run;
quit;
%mend;

```

```

%sethsg(mpoint=6, name = hist1);
run;
%sethsg(mpoint=1, name = hist2);
run;
%sethsg(mpoint=8, name = hist3);
run;
%sethsg(mpoint=3, name = hist4);
run;
title;
proc gslide;
note move=(15,96)pct h=2.5 c=blue 'Histogram with Different Midpoint Sizes';
note move=(5,1)pct h=1.5 'Nonparametric kernel density function bandwidth
parameter c= 0.2 0.5 1';
%macro setxy;
%do i=0 %to 2 %by 2;
%let x1=%eval(&i * 23 );
%let y1=%eval(&x1 + 45);
%eval(1*&i+1)/ ulx=&x1 uly=50 urx=&y1 ury=50
llx=&x1 lly=5 lrx=&y1 lry=5
%eval(1*&i+2)/ ulx=&x1 uly=94 urx=&y1 ury=94
llx=&x1 lly=51 lrx=&y1 lry=51
%end;
5/ ulx=0 uly=100 urx=100 ury=100
llx=0 lly=0 lrx=100 lry=0
%mend;
%macro histn;
%do i=1 %to 4;
&i:hist%eval(&i)
%end;
5:gslide
%mend;
filename pfname 'temp.ps';
proc greplay igout=gseg nofs tc=cat1x1;
tdef conv1x1
%setxy;
template conv1x1;
treplay
%histn;; quit;
x ps2pdf temp.ps hslab01.pdf;run;quit;
x rm temp.ps;run; quit;
run;

```

HISTOGRAM WITHOUT DISPLAYING BARS

You can use the 'WIDTH' statement to control the thickness of the lines. 'WIDTH=n' statement specifies the width (in positive integer pixels) of the line. The WIDTH statements are added to the following sample code.

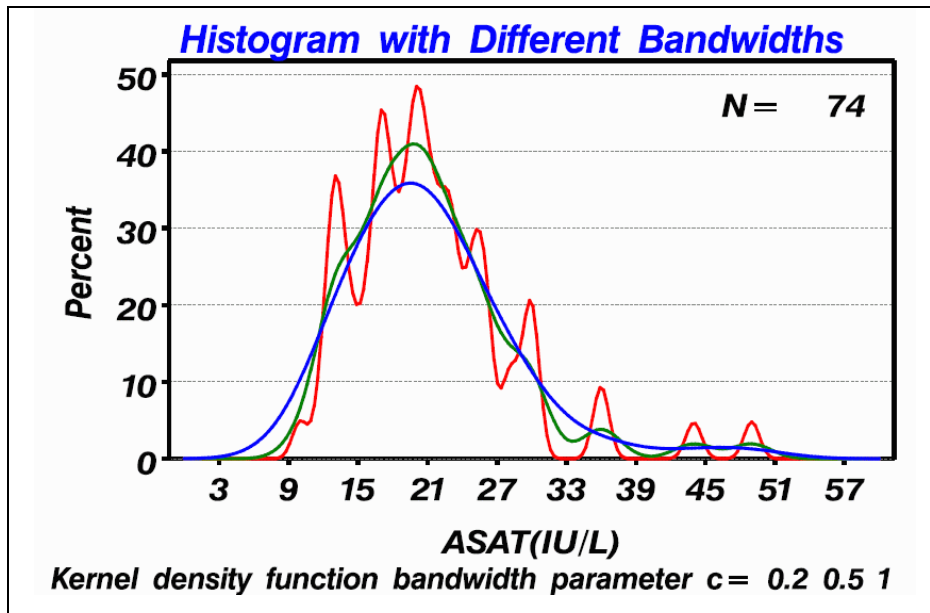


Figure 2. Histogram without Displaying Bars

TWO-WAY COMPARATIVE HISTOGRAM

The UNIVARIATE procedure provides the following features for producing a two-way comparative histogram.

| | |
|---|--|
| * | Specifies two classification variables |
| * | Specifies the order of the component histograms |
| * | Specifies the number of rows and columns for the graph |
| * | Specifies the distance between the component histogram tiles |
| * | Specifies the scale, values and labels of the vertical axis |
| * | Specifies the midpoints for histogram intervals |
| * | Inserts summary statistics to the graph |
| * | Accept SAS/GRAPH statement for graphical enhancements |
| * | Suppresses the tables of descriptive statistics. |
| * | Superimposes kernel density estimates on histograms |

```

goptions reset=all device=ps600c gsfname=psname rotate=landscape gsfmode=replace
  ftext=swissbi vorigin=1.2in horigin=1in vsize=6.2in hsize=9.2in ;
symbol1 i=j c=red;
filename psname 'temp.ps';
proc univariate data=final noprint;
  var asat;
  class trx_o visit;
  histogram asat/
    nrows = 4
    ncols = 2
    intertile = 1
    kernel( c = 1 w=15 )
    vscale = count
    vaxis = 0 to 40 by 10
    vaxislabel='Count'
    midpoints=5 to 70 by 6
    font=swissbi
    waxis = 20
    wbarline = 10

```

```

cbarline = blue
cfill = cyan
height = 3.5
name="hist1"
;
inset n='N=' (3.) /noframe position=ne height=2.5;
format trx_o $trx. visit $visits.;
run; quit; title;
proc gslide;
note move=(28,96)pct h=2.8 c=blue
"Two-Way Comparative Histogram";
note move=(0,1)pct h=1.9 'Worst case (highest) value is selected for On-therapy
Period for each subj.';
proc greplay igout=gseg nofs tc=cat1x1;
tdef conv1x1
1/ ulx=0 uly=95 urx=100 ury=95
llx=0 lly=5 lrx=100 lry=5
2/ ulx=0 uly=100 urx=100 ury=100
llx=0 lly=0 lrx=100 lry=0;
template conv1x1;
treplay
1:hist1
2:gslide;
quit;
x ps2pdf temp.ps hslab02.pdf;run;quit;
x rm temp.ps;run;
quit;
run;

```

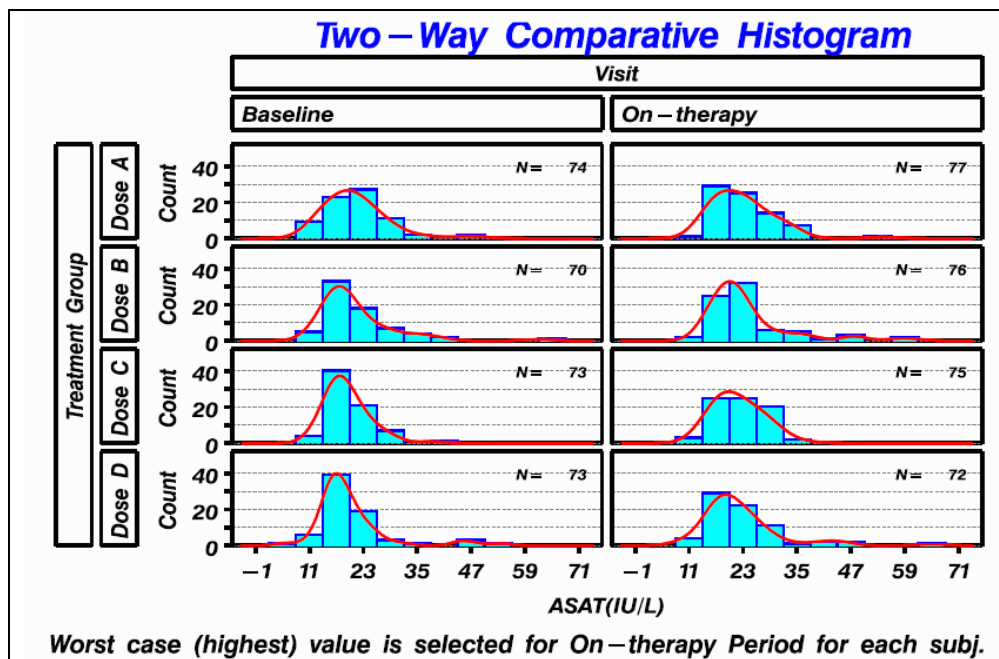


Figure 3. Sample Display of Two-Way Comparative Histogram

KERNEL DENSITY ESTIMATES

The KDE procedure utilizes the kernel density estimation method to compute nonparametric estimates of univariate and bivariate probability density functions. This type of nonparametric technique attempts to estimate the density function, $f(x)$, of a data distribution without assuming a particular form for that distribution. The output from the kernel density estimate can be treated as a smoothed histogram. Smoothing involves a trade-off between

fidelity of the data (low bias), and smoothness (low variance). The smoothing and the choice of bandwidth is an important issue in the application of kernel density estimation. The KDE procedure provides several methods for automatic bandwidth selection. The KDE procedure creates the SAS data sets to store the density estimates for subsequent plotting or analysis tasks. The following sample SAS code produces Figure 4.

```

proc kde data=ecg out=ol;
  by trt;
  var base_chg;
run;

goptions reset=all device=ps600c gsfname=psname rotate=landscape gsfmode=replace
  ftext=swiss vorigin=1.2in horigin=1in hsize=9.2in vsize=6.2in hpos=90 vpos=60 ;
filename psname 'temp.ps';
title1 j=1 h=10pt 'Protocol: XYZ999';
title2 j=1 h=10pt 'Population: Intent-to-Treat';
title3 j=c h=14pt 'Density Estimates of QTC Maximum Change from Baseline by
Treatment and Gender';
footnote1 h=13pt j=1 'Note: Normal(Increase < 30 msec), High(Increase 30 - 60
msec), Concern(Increase > 60 msec)';
footnote2 h=10pt j=1 "USER ID:/kdqtc02w.sas &sysdate &systemtime";
symbol1 i=j c=blue l=1 width=20;
symbol2 i=j c=red l=12 width=20;
symbol3 i=j c=green l=23 width=20;
symbol4 i=j c=brown l=33 width=20;
axis1 order=( -45 to 90 by 15 ) label=(h=14pt 'Maximum Change from
Baseline(msec)') width=16
  value=(h=12pt );
axis2 order=(0 to 0.035 by 0.005) width=16 value=(h=13pt) label=(h=13pt a=90
'Density');
legend1 label=none value=(h=13pt);
proc gplot data=ol;
  plot density*base_chg=trt /overlay
    href=0 30 60
    lvref=2
    vaxis=axis2
    haxis=axis1
    legend=legend1
  ;

run;
x ps2pdf temp.ps kdqtc02w.pdf;run;
x rm temp.ps;run;
quit;
run;

```

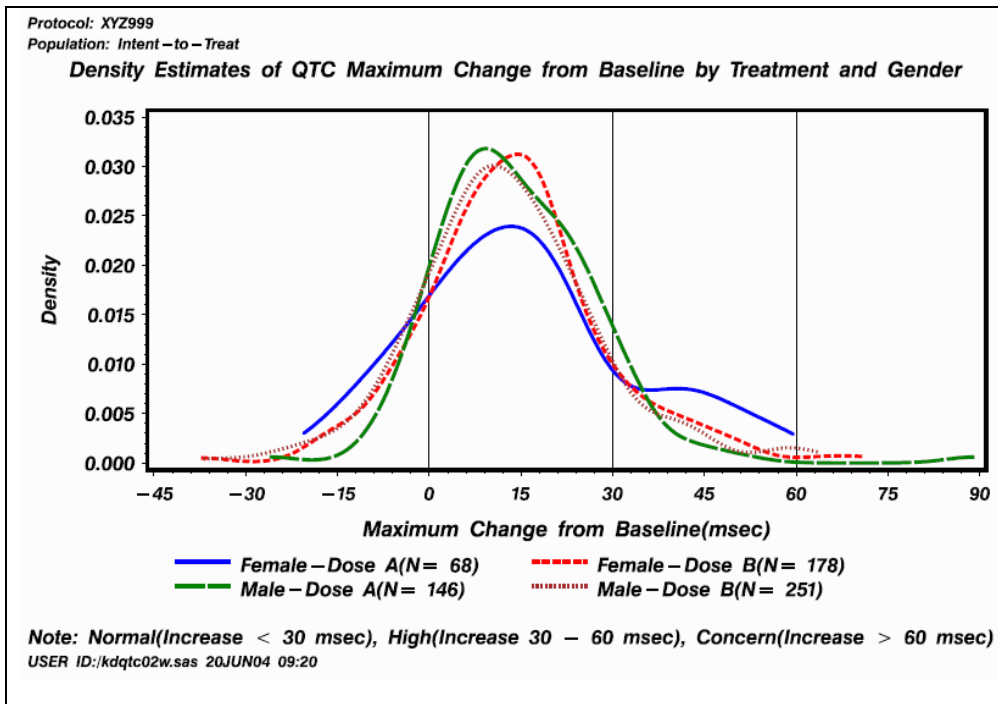


Figure 4. Sample Display of Kernel Density Estimates

CUMULATIVE DENSITY FUNCTION

The Cumulative Density Function (CDF) is a variation of the density function in which the vertical axis counts not only the data points in a single class interval, but also counts the data points for that class interval - plus all data points in all intervals in which the x-axis value is smaller than the value of the current interval. In addition, the variant that the y-axis produced can be presented by a percentage.

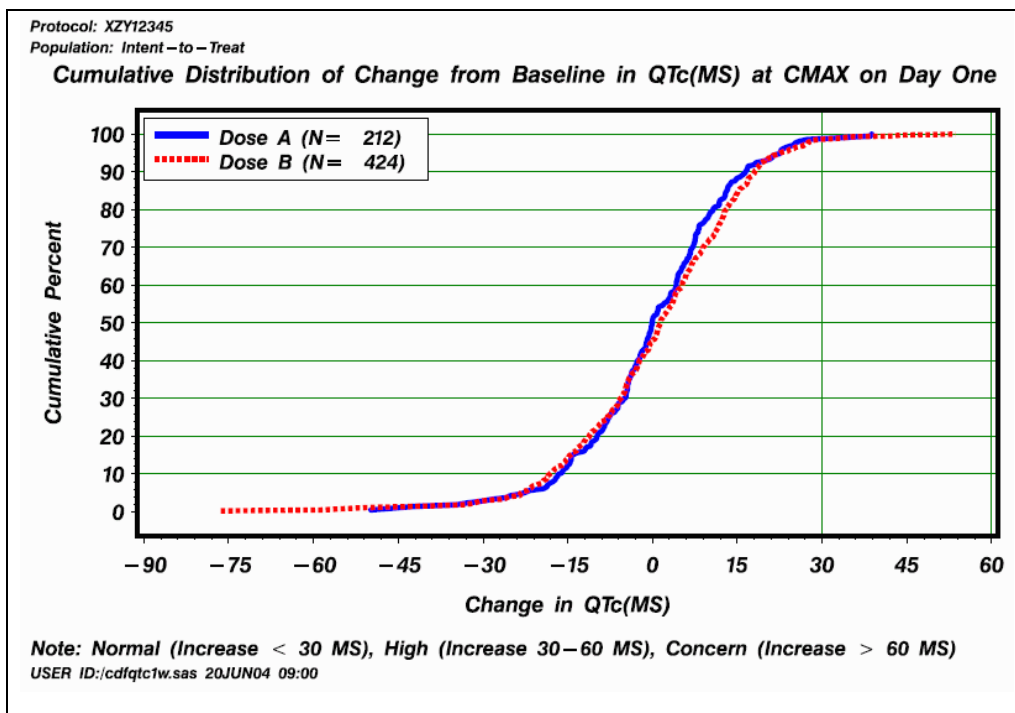


Figure 5. Sample of CDF Display

OVERLAYING CDF AND DENSITY ESTIMATES ON THE SAME GRAPH

You can overlay the CDFs ontop of density estimates to model the probability density functions or make the comparison of different groups.

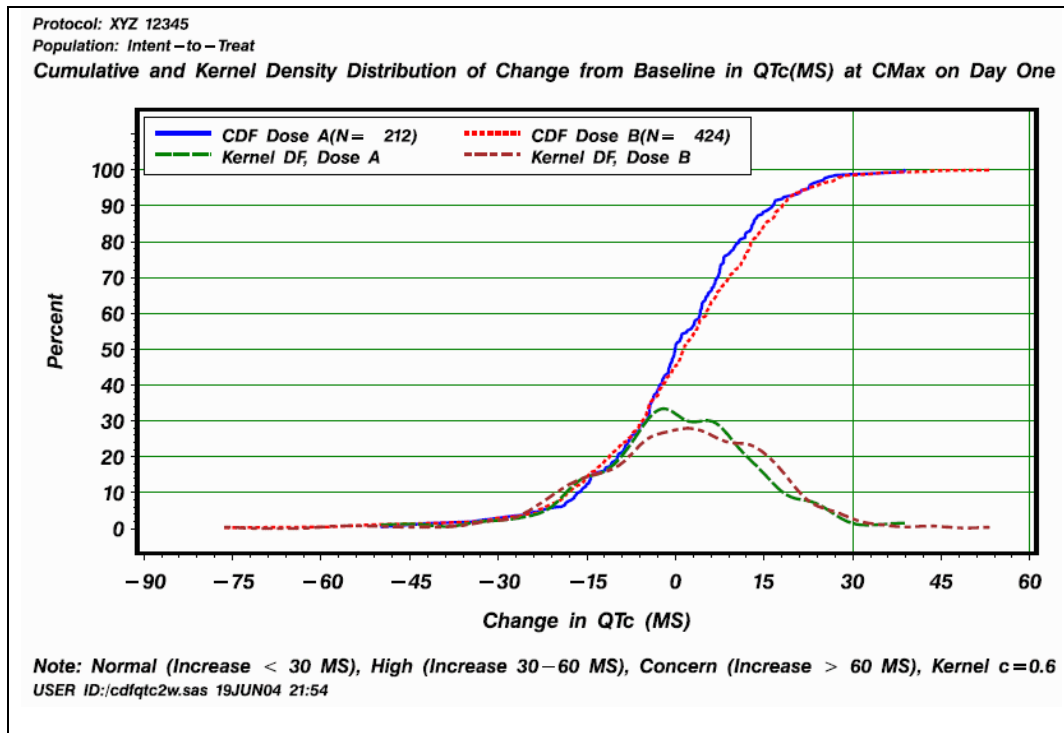


Figure 6. Sample Display of Overlaying CDF and Density Estimates

```
/* prepare ECG data set */
proc sort;by rand;
proc summary data=qtcf;
  var count;
  by rand;
output out=pl(drop=_type_ _freq_) sum=scount;
proc sort data=qtcf;by rand;
data pl sb;
  merge qtcf pl;
  by rand;
  pct=count/scount;
  if substr(rand,1,1)='P' then output pl;
  else output sb;
run;
proc sort data=pl;by base_chg;
proc sort data=sb;by base_chg;
data pl(keep=base_chg plcdf scount rand);
  set pl;
  plcdf=(pct * _n_) * 100;
  call symput('pcount',put(scount,4.));
data sb(keep=base_chg sbcdf scount rand);
  set sb;
  sbcdf=(pct * _n_) * 100;
  call symput('scount',put(scount,4.));
run;
data final;
  set pl sb;
```



```

proc sort;by rand;
proc kde data=final out=ol bwm=0.6;
by rand;
var base_chg;
proc print data=ol;run;
data ol;
set ol;
if substr(rand,1,1)='P' then pden=density*1000;
else if substr(rand,1,1)='S' then sden=density*1000;
keep rand base_chg pden sden;
data final;
set final ol;
label
  plcdf = "CDF Dose A(N= &pcount)"
  sbcdf = "CDF Dose B(N= &scount)"
  pden = 'Kernel DF, Dose A'
  sden = 'Kernel DF, Dose B'
;
proc sort; by base_chg;
goptions reset=all device=ps600c gsfname=psname rotate=landscape gsfmode=replace
  ftext=swiss vorigin=1.2in horigin=1in vsize=6.2in hsize=9.2in ;
filename psname 'temp.ps';
title1 j=1 h=10pt 'Protocol: XYZ 12345';
title2 j=1 h=10pt 'Population: Intent-to-Treat';
title3 j=c h=15pt 'Cumulative and Kernel Density Distribution of Change from
Baseline in QTc(MS) at CMax on Day One';
footnote1 j=1 h=13pt 'Note: Normal (Increase < 30 MS), High (Increase 30-60 MS),
Concern (Increase > 60 MS), Kernel c=0.6';
footnote2 j=1 h=10pt "USER ID:/cdfqtc02.sas &sysdate &systemtime";
legend1 position=(left top inside)
  frame
  across=2
  mode=protect
  label=none
  value=(h=11pt);

symbol1 i=j c=blue l=1 width=16 ;
symbol2 i=j c=red l=2 width=16;
symbol3 i=j c=green l=4 width=16 ;
symbol4 i=j c=brown l=25 width=16;
axis1 label = ( h=13pt angle=90 'Percent')
  width = 17
  order = ( 0 to 110 by 10 )
  value = ( h=13pt tick=12 ' ' ) offset=(15pt);
axis2 label = ( h=13pt 'Change in QTc (MS)')
  width = 17
  order = ( -90 to 60 by 15)
  value = ( h=13pt );
proc gplot data=final;
plot plcdf*base_chg=1 sbcdf*base_chg= 2 pden*base_chg=3 sden*base_chg=4
  / overlay
  haxis= axis2
  vaxis= axis1
  href=30 60
  vref=10 20 30 40 50 60 70 80 90 100
  chref=g
  cvref=g
  legend = legend1
;
run;
quit;

```

```
x ps2pdf temp.ps cdfqtc02.pdf;run;quit;
x rm temp.ps;run;
quit;
run;
quit;
```

BOX PLOT

The SAS system provides the procedure BOXPLOT and GPLOT with SYMBOL statements to produce box plots. This section discusses the BOXPLOT procedure. The BOXPLOT procedure creates groups of side-by-side box-and-whisker plots from the variable set. A box-and-whisker plot displays the mean, median, quartiles, minimum, and maximum observations for a group. The plot elements and statistics shown from this procedure are as follows:

| | |
|---|---|
| * | The length of the box represents the interquartile range (the distance between the 25 th and the 75 th percentiles) |
| * | The symbol in the box interior represents the mean |
| * | The horizontal line in the box interior represents the median |
| * | The vertical lines extending from the box indicating the minimum and maximum values or the specified interquartile range. |

The procedure provides the following features to customize the box plots:

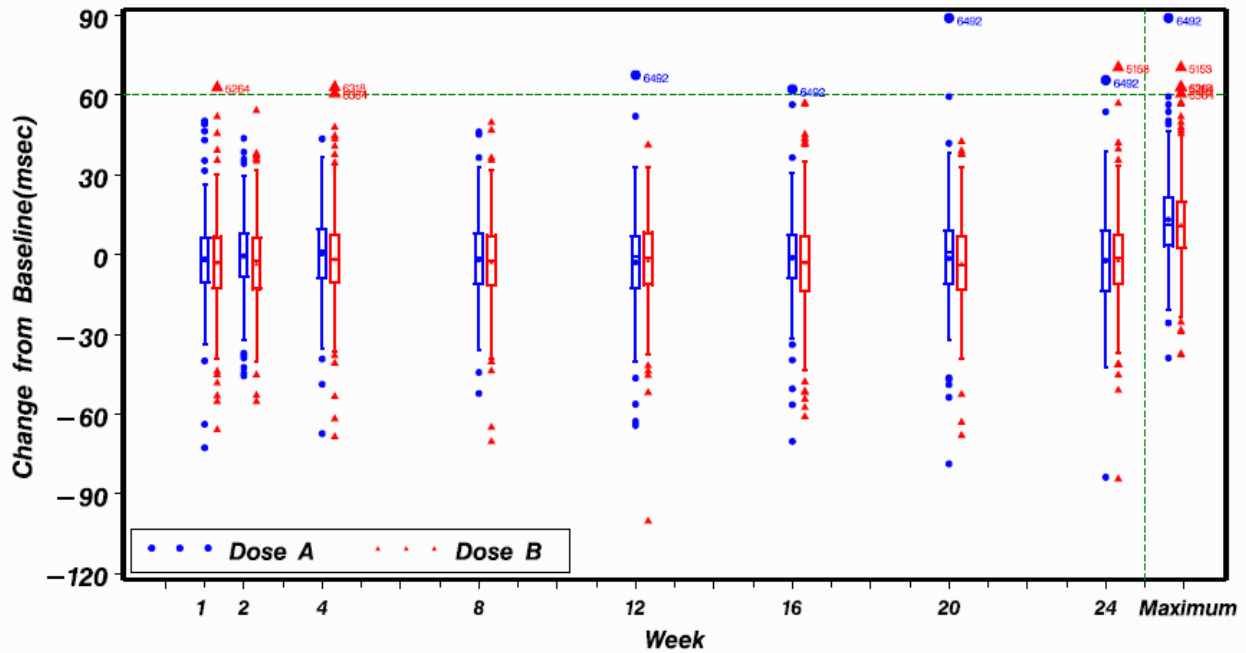
| | |
|---|---|
| * | Specify one of 5 methods for calculating quantile statistics |
| * | Select the style of the box-and-whisker plots |
| * | Display vertical and horizontal reference lines |
| * | Control the layout and appearance of the plot |
| * | Add symbols to reveal groups in data |
| * | Size of plot relates to sample size of the group |
| * | Display of outliers with outliers' ID |
| * | Connection of key statistics for longitude data display |
| * | Use block variable for grouping purpose with block legends added to graph |
| * | Display data from annotate data set for graphical enhancement |
| * | Accept SAS/GRAPH statements for graphical enhancement |

Figure 7 is a sample box plot display from the BOXPLOT procedure. The data used for this procedure are hypothetical clinical trials with the corrected QT data collected at different study time points. Changes from baseline measures are computed for each on-therapy visit. This sample box plots graph provides the following features to assist in the interpretation of the clinical data: 1) treatment groups at each visit are side-by-side for easier comparison, 2) the x-axis is a real time scale indicating the planned visit in week unit, 3) using different colors and symbols for different treatment groups, 4) when the outliers become safety and clinical concerns, the clinical concern level is displayed as reference lines (at 30 and 60) and subjects are plotted with measures above the concern level and their IDs, and 5) the quantity of subjects in each visit are displayed in the footnote area aligned with x-axis time scale to provide additional study visit information.

The option SCHEMATICID in the BOXPLOT procedure can print the subject ID next to the outlier, but when the outliers are very close to each other, it leads to overprinting, as seen in week 4 visit below.

Protocol: XYZ 12345 Study 999
 Population: Intent-to-Treat

QTc Change from Baseline by Time and Treatment



| Subjects At Visit | |
|-------------------|-------------------------------------|
| Dose A | 209 206 199 191 184 176 169 164 212 |
| Dose B | 405 384 362 337 315 311 299 293 413 |

USER ID: /bxqtc08w.sas 26JUL04 19:19

Figure 7. Sample Box Plots from BOXPLOT Procedure

MULTIVARIATE SCATTER PLOT

The LOESS procedure fits a nonparametric method for estimating local regression surfaces. It supports the use of multidimensional data, multiple dependent variables, and both direct and interpolated fitting using kd trees. Additionally, the LOESS procedure provides the normal or symmetric error distributions for statistical inference. It performs robust fitting for data sets having outliers.

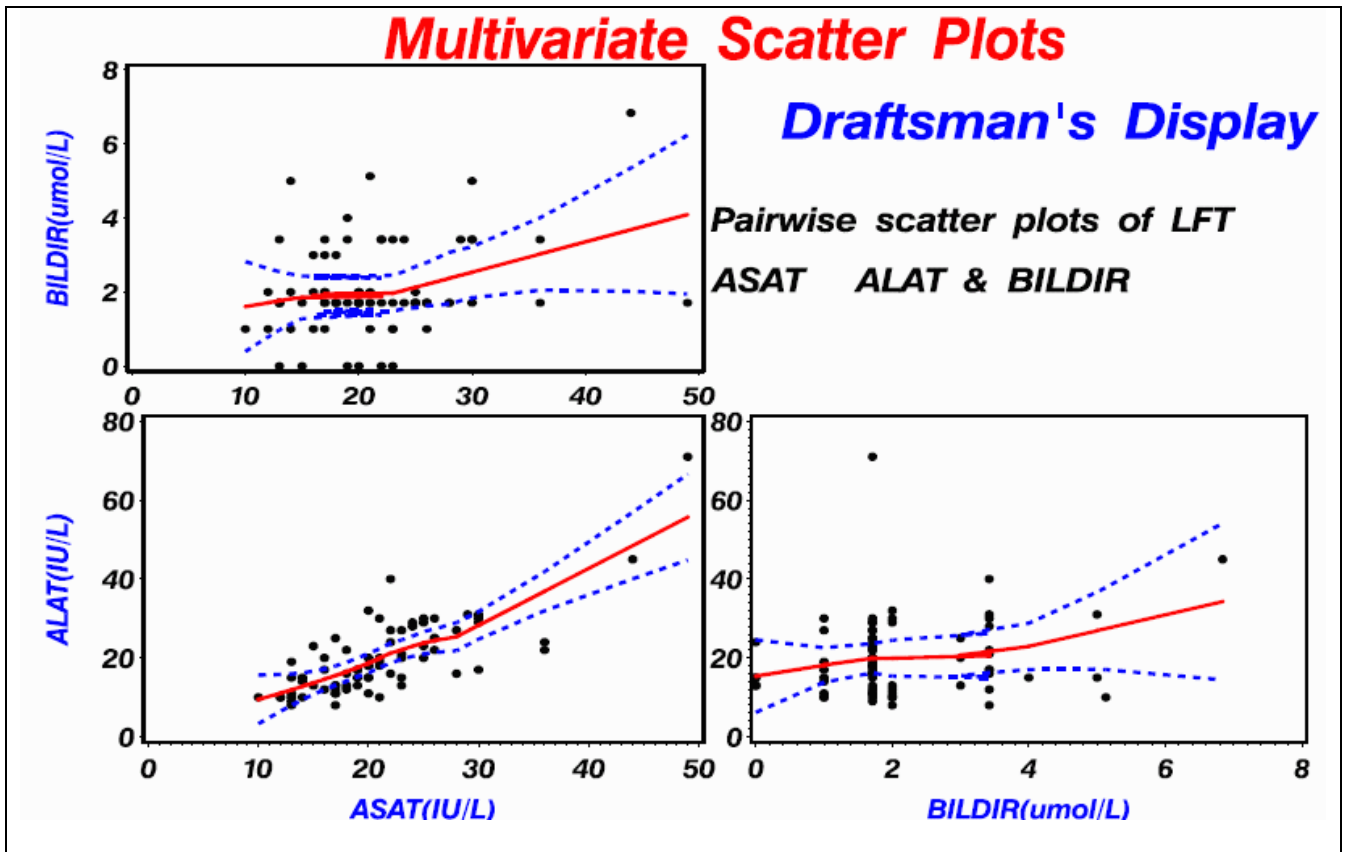


Figure 8. Sample Multivariate Scatter Plots

```

/* prepare LFT data set */

%macro setnp(y=,x=,vaxis=, haxis=);
proc loess data=final;
ods output outputstatistics =outstay;
model &y = &x /smooth=0.8 alpha=.01 all;
run;
proc sort data=outsta;by pred;
%macro val1;
order = ( 0 20 40 60 80) width=20
value = ( c=black h=3
tick=1 '0'
tick=2 '20'
tick=3 '40'
tick=4 '60'
tick=5 '80');

%mend;
%macro val2;
order = ( 0 2 4 6 8) width=20
value = ( c=black h=3
tick=1 '0'
tick=2 '2'
tick=3 '4'
tick=4 '6'
tick=5 '8');

%mend;
%macro val3;

```

```

order = ( 0 10 20 30 40 50) width=20
value = ( c=black h=3
          tick=1 '0'
          tick=2 '10'
          tick=3 '20'
          tick=4 '30'
          tick=5 '40'
          tick=6 '50');

%mend;
goptions reset=all device=ps600c gsfname=psname rotate=landscape gsfmode=replace
ftext=swissbi vorigin=1.2in horigin=1in vsize=6.2in hsize=9.2in;
filename psname 'temp.ps';
symbol1 i=none c=black v=dot h=2;
symbol2 i=j c=red l=1 width=30;
symbol3 i=j c=blue width=30 l=2;
symbol4 i=j c=blue width=30 l=2;
symbol5 i=j c=red width=24 l=5;
axis1 label = ( c= blue h=3 angle=90 'ALAT(IU/L)') %val1;
axis5 label = ( c= blue h=3 angle=90 ' ') %val1;
axis3 label = ( c=blue h=3 angle=90 'BILDIR(umol/L)') %val2;
axis4 label = ( c=blue h=3 'BILDIR(umol/L)') %val2;
axis2 label = ( c=blue h=3 'ASAT(IU/L)') %val3;
axis6 label = ( c=red h=3 ' ') %val3;

proc gplot data=outsta;
plot (depvar pred lowercl uppercl)* &x
      / overlay
      haxis= &haxis
      vaxis= &vaxis;

run;
quit;
%mend;
%setnp(y=bildir, x=asat,haxis=axis6,vaxis=axis3);
run;
%setnp(y=alat, x=asat,haxis=axis2,vaxis=axis1);
run;
goptions reset=all device=ps600c gsfname=psname rotate=landscape gsfmode=replace
ftext=swissbi vorigin=1.2in horigin=1in vsize=6.2in hsize=9.2in;
filename psname 'temp.ps';
proc gslide;
note move=(20,78)pct h=6 c=blue "Draftsman%str('s Display";
note move=(10,55)pct h=4 'Pairwise scatter plots of LFT';
note move=(10,40)pct h=4 'ASAT, ALAT & BILDIR'; run;
%setnp(y=alat, x=bildir,haxis=axis4,vaxis=axis5);

title;
proc gslide;
note move=(27,93)pct h=3.2 c=red 'Multivariate Scatter Plots';
note move=(0,2)pct h=1.7 c=red 'Red' c=black 'solid lines are Loess fitted
curves,'
c=blue 'blue' c=black ' dot lines are 99% confidence bands';
%macro setxy;
%let x1=0;
%let y1=54;
%let x2=47;
%let y2=100;
1/ ulx=&x1 uly=94 urx=&y1 ury=94
llx=&x1 lly=48 lrx=&y1 lry=48
2/ ulx=&x1 uly=54 urx=&y1 ury=54
llx=&x1 lly=5 lrx=&y1 lry=5

```

```

3/ ulx=&x2 uly=94  urx=&y2 ury=94
   llx=&x2 lly=48  lrx=&y2 lry=48
4/ ulx=&x2 uly=54  urx=&y2 ury=54
   llx=&x2 lly=5   lrx=&y2 lry=5
5/ ulx=&x1 uly=100 urx=100 ury=100
   llx=&x1 lly=0   lrx=100 lry=0
%mend;
%macro histn;
   1:gplot
   2:gplot1
   3:gslide
   4:gplot2
   5:gslide1
%mend;
filename psname 'temp.ps';
proc greplay igout=gseg nofs tc=cat1x1;
  tdef conv1x1
  %setxy;
  template conv1x1;
  treplay
    %histn;; quit;
x ps2pdf temp.ps nplab03.pdf;run;quit;
x rm temp.ps;run;
quit; run;

```

ODS GRAPHICS IN PROCEDURE KDE

The ODS Statistical Graphics (or ODS Graphics for short) is a new experimental feature in SAS 9.1. It provides commonly used statistical graphics, such as scatter plots, histograms, box plots, contour plots, and 3-D plots from several SAS products procedures automatically.

ODS Graphics can not produce just graphs without other ODS output tables and listings. In other words, ODS Graphics are part of ODS output and use at least one ODS destination to obtain graphics. ODS Graphics uses Java technology and is independent of SAS/GRAPH. ODS Graphics do not support the following statements: 1) any SAS/GRAPH statements, such as GOPTIONS, SYMBOL, PATTERN, 2) the GTITLE or GFOOTNOTE options available with the ODS destinations HTML, RTF, and MACKUP, 3) the ODS USEGOPT statement.

The ODS Graphics created from the KDE procedure are listed in Table 1.

| Statement | ODS Graph Name | Plot Description | PLOT=Option |
|-----------|--------------------|---|----------------|
| UNIVAR | Density | Univariate kernel density estimate curve | DENSITY |
| | HistDensity | Univariate histogram overlaid with kernel density estimate curve | HISTDENSITY |
| BIVAR | Histogram | Univariate histogram of data | HISTOGRAM |
| | BivariateHistogram | Bivariate histogram of data | HISTOGRAM |
| | Contour | Contour plot of bivariate kernel density estimate | CONTOUR |
| | ContourScatter | Contour plot of bivariate kernel density estimate overlaid with scatter plot | CONTOURSCATTER |
| | HistSurface | Bivariate histogram overlaid with surface plot of bivariate kernel density estimate | HISTSURFACE |
| | Scatterplot | Scatter plot of data | SCATTER |
| | SurfacePlot | Surface plot of bivariate kernel density estimate | SURFACE |

Table 1 ODS Graphics Produced by PROC KDE

The following example illustrates the creation of ODS Graphics from the KDE procedure. The clinical liver function test values of Alanine Aminotransferase (ALAT) and Asparatate Aminotransferase (ASAT), both with unit of (U/L), are used for the illustration purpose.

The following SAS code is used for producing default univariate ODS Graphics in PROC KDE.

```
ods html style=journal gpath="c:\tip2" file="b001ac.htm";

ods graphics on /imagefmt=jpeg imagename = "b001ac";

proc kde data=final ;
  univar asat /plots = density histdensity histogram; run;

ods html close;
ods graphics off;
```

Three default univariate ODS Graphics output are shown in Figure 9.

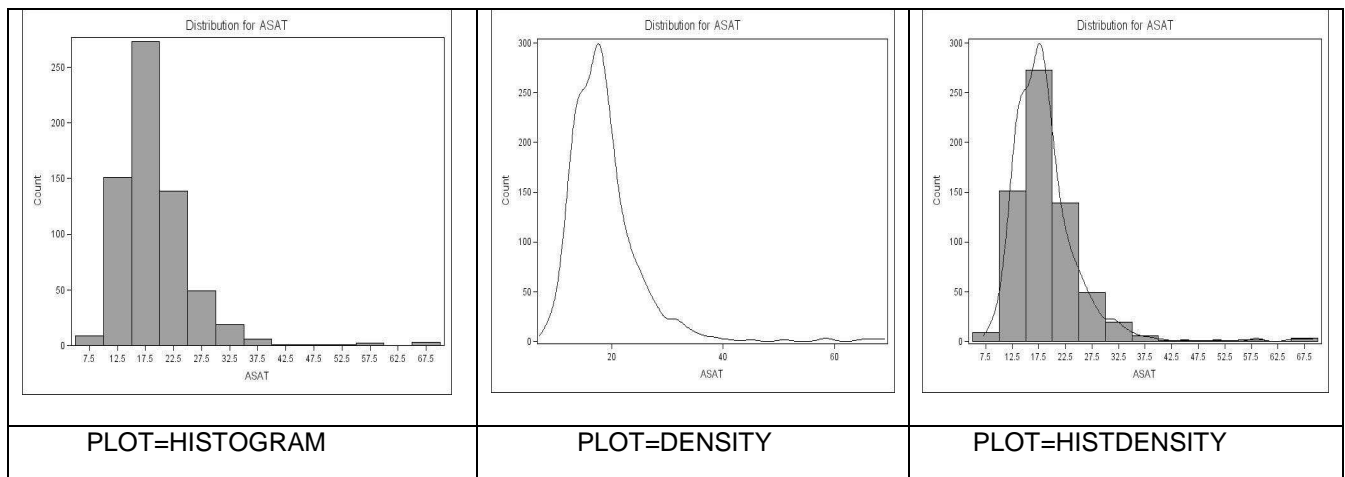


Figure 9. Default Univariate ODS Graphics from PROC KDE

The procedure KDE produces six ODS Graphics from BIVAR statement. These six ODS Graphics are: HISTOGRAM, CONTOUR, CONTOURSCATTER, HISTSURFACE, SCATTER, and SURFACE. You can use BIVAR statement below for all six bivariate displays.

```
bivar x y / plots=all;
```

The following sample SAS code is for bivariate graphics.

```
ods html style=journal gpath="c:\tip2" file="b001aa.htm";
ods graphics on /imagefmt=jpeg imagename = "b001aa";

proc kde data=final ;
  bivar asat alat/plots = all;
run;
ods html close;
ods graphics off;
```

The default ODS Graphics output from this job is shown in Figure 10.

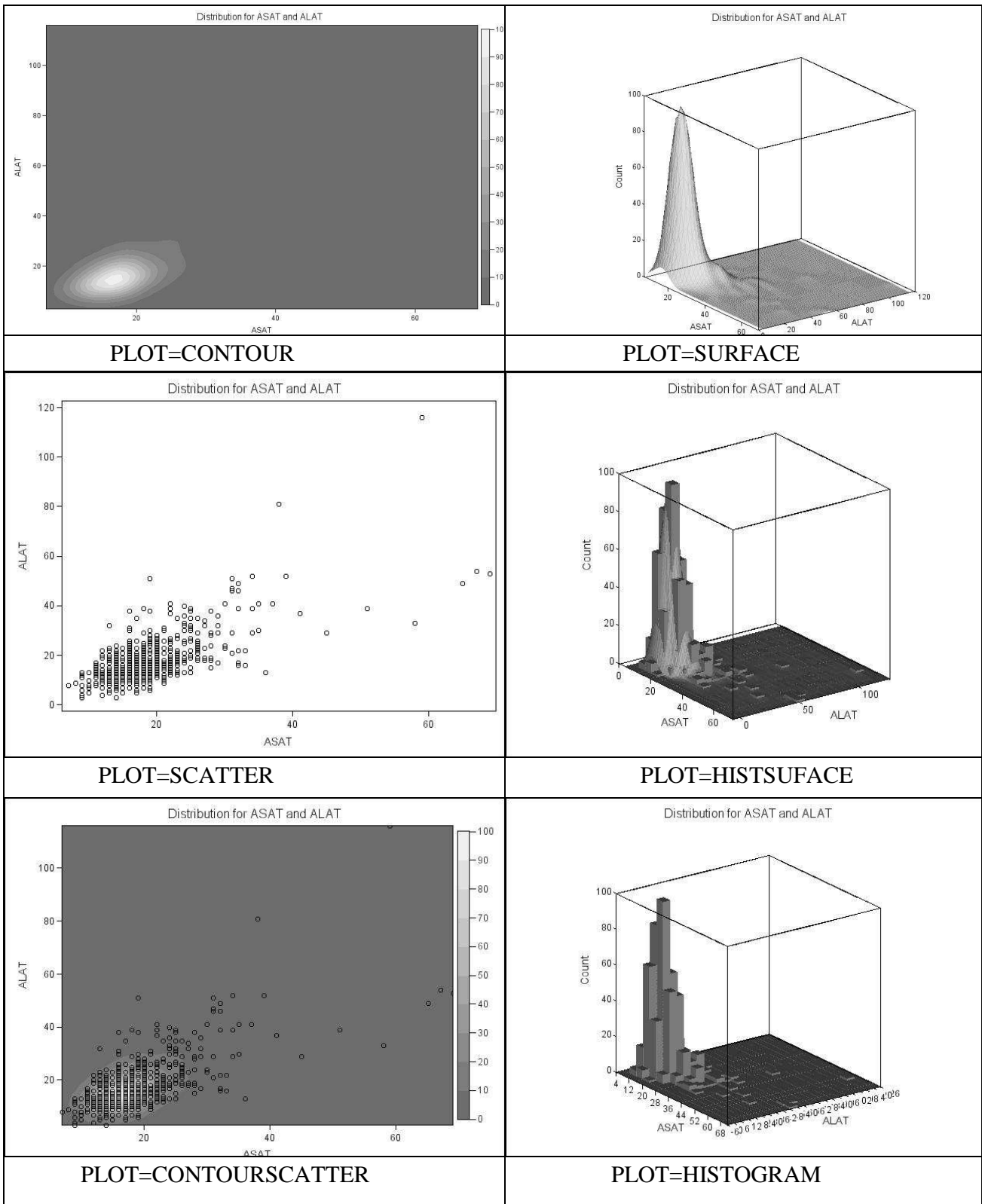


Figure 10. Default Bivariate ODS Graphics from PROC KDE

CUSTOMIZING ODS GRAPHICS

In SAS 9.1, the ODS Graphics are controlled by a new graph definition language called ODS Graphics Template Language (GTL). ODS Graphics are part of procedure ODS output and are governed by the standard ODS statements. Furthermore, graph appearance, like ODS output tables, are controlled by the current ODS style.

The names of the templates a procedure uses are predefined, the users can not change a template's name or add new templates to a procedure's output. The default SAS supplied templates are located in SASHELP.TEMPLMST. You can change the default template and keep the same template name, but store the modified template to SASUSER path that is searched before the default path.

The modification of the style template statements are as follows:

```
'gconramp3cend' = cxFF0055  
'gconramp3cneutral' = cxA7A7A7  
'gconramp3cstart' = cx66FFFF  
'gramp3cend' = cxFF0055  
'gramp3cneutral' = cxA7A7A7  
'gramp3cstart' = cx66FFFF
```

The customized output is shown in Figure 11.

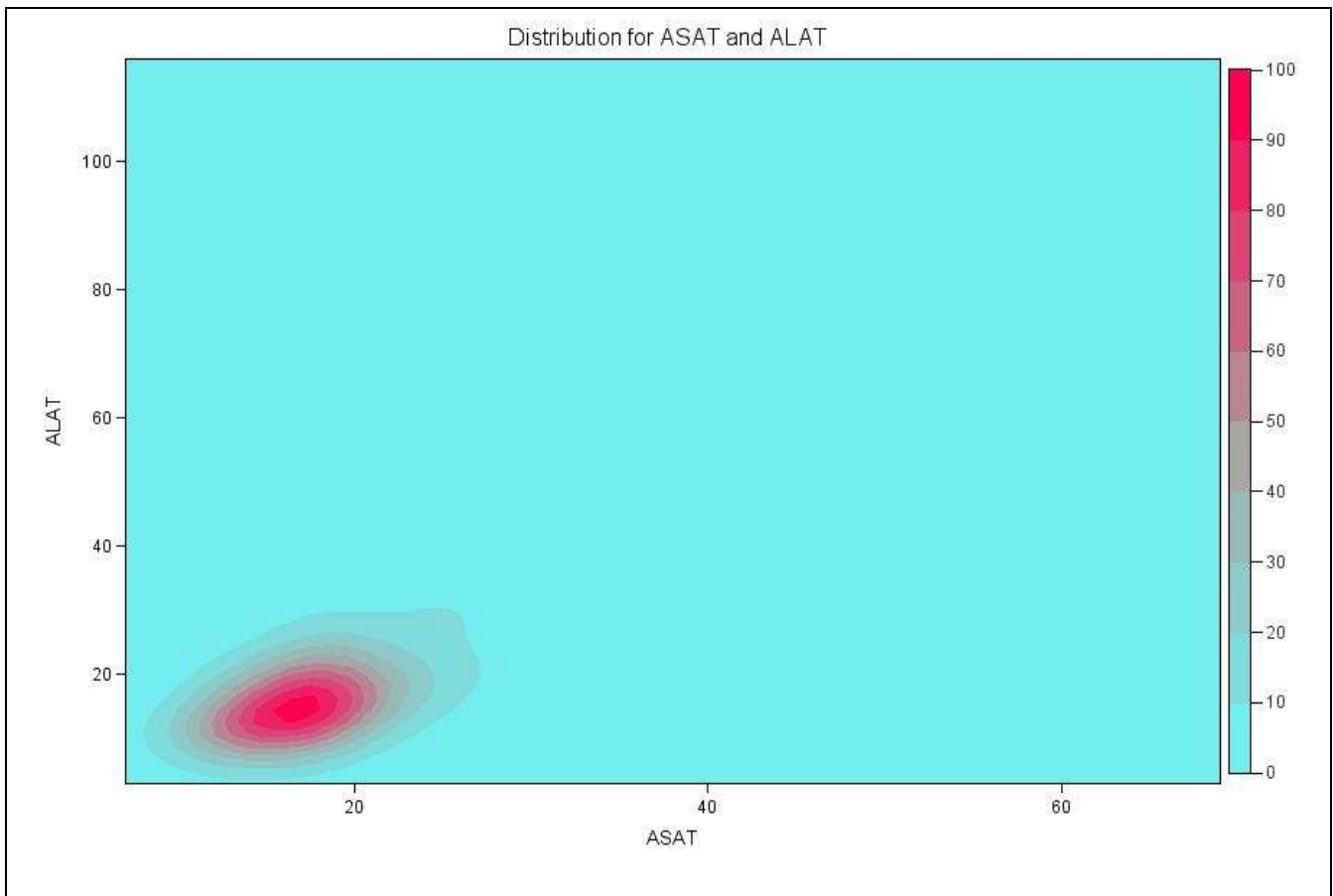


Figure 11. Customized ODS Graphics of Contour from PROC KDE

CONCLUSIONS

Graphics output is an important means to convey the information. The nonparametric methods provide a good starting point for clinical data analyses. The procedures of UNIVARIATE, BOXPLOT, KDE, LOESS perform the nonparametric estimates and save them as SAS data sets or ODS tables for subsequent graphic plotting tasks. In SAS 9.1, the experimental ODS Graphics provide a quick and easy way to produce multiple graphs automatically as part of ODS output. If customized ODS Graphics are needed, then the modifications of default graphs involve changes of SAS supplied Graphical Templates.

This paper provides some sample SAS code for nonparametric graphical displays. Sample displays shown in this paper not only help with clinical study results interpretation, but also aid in the QC tasks of checking the data, tables, and clinical reports.

REFERENCES

- [1] Barnes, G.R., P. B. Cerrito,(2000): *The Visualization of Continuous Data Using PROC KDE and PROC CAPABILITY*, Proceedings of the 26th SAS Users Group International Conference, Paper 176-26.
- [2] Cerrito, P.B., G.R. Barnes, G.R.,(2001): *The Use of Kernel Density Estimators to Monitor Protocol Compliance*, Proceedings of the 25th SAS Users Group International Conference, Paper 273-25.
- [3] Cowmeadow, M.(2004) *Box Plots and Radar Plots for Pharmaceutical Studies: Some of the Better Ways to get a Clear Picture of Study Results*. Proceedings of Annual Conference of the Pharmaceutical Industry SAS Users Group, PharmaSUG 2004, paper SP09
- [4] Cohen, R.A(1999).; *An Introduction to PROC LOESS for Local Regression*, Proceedings of the 24th SAS Users Group International Conference, Paper 273.
- [5] <http://support.sas.com/rnd/base/topics/statgraph/v91StatGraphStyles.htm>
- [6] SAS Institute Inc.(2004): *SAS/STAT 9.1 User's Guide*, SAS Institute Inc., Cary NC, USA.
- [7] SAS Institute Inc.(2004): *SAS 9.1 Output Delivery System: User's Guide*, SAS Institute Inc., Cary NC, USA.
- [8] SAS Institute Inc.(2003): *SAS OnlineDOC, Version 8.2*. Cary NC, SAS Institute Inc.
- [9] SAS Institute Inc.(2003): *The Analyst Application, Second Edition*, SAS Institute Inc., Cary NC, USA.
- [10] SAS Institute Inc(1999).: *The Complete Guide to the SAS Output Delivery System, Version 8*, SAS Institute Inc., Cary NC, USA.
- [11] SAS Institute Inc.(1993): *SAS/GRAPH Software: Example, Version 6, First Edition*. Cary NC, SAS Institute Inc.
- [12] Watt, P.(2004); *Multiple-Plot Displays: Simplified with Macros*, SAS Publishing, Books by Users press.
- [13] Yeh, S.T.(2004): "*Tips to Enhance Your SAS Statistical Graphics Output*", Proceedings of the 29th SAS USERS International Conference. Paper 169-29

TRADEMARK CITATION

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 registered trademarks or trademarks of their respective companies.

AUTHOR CONTACT INFORMATION

Shi-Tao Yeh, Ph. D.
(610)787-3856 (W)
E-mail: shi-tao_yeh-1@gsk.com