

Analysis of variance and regression

November 13, 2007

SAS graphics

- Scatter plots
- Histograms
- Probability plots
- Box plots
- Improving the plots
- Designing pages

Lene Theil Skovgaard,
Dept. of Biostatistics,
Institute of Public Health,
University of Copenhagen

e-mail: L.T.Skovgaard@biostat.ku.dk

http://staff.pubhealth.ku.dk/~lts/regression07_2

Example: Obesity and blood pressure

```
data sasuser.bp;
input sex_no obese bp;
```

```
/* coding of a new variable */
if sex_no=2 then sex='female';
if sex_no=1 then sex='male';
drop sex_no;
datalines;
1 1.31 130
1 1.31 148
1 1.19 146
.
.
2 2.20 136
2 1.64 136
2 1.73 208
;
run;
```

```
proc print data=sasuser.bp;
var sex obese bp;
run;
```

Obs	sex	obese	bp
1	male	1.31	130
2	male	1.31	148
3	male	1.19	146
.	.	.	.
.	.	.	.
101	female	1.64	136
102	female	1.73	208

After running this program, the data is located in the **permanent** data set **sasuser.bp** and contains the following variables

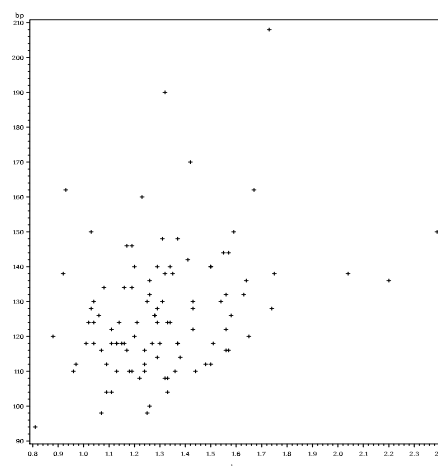
- **SEX**: character variable
- **OBESE**: obesity, defined as weight, divided by ideal weight
- **BP**: systolic blood pressure

In total

- 3 variables
- 102 observations

Scatter plot in raw form:

```
proc gplot;
plot bp*obese;
run;
```



This can be improved a lot....

Plotting facilities in SAS (apart from scatter plots)

- Histograms:

```
proc univariate plot normal gout=plotud data=sasuser.bp;
  var lbp; class sex;
  histogram / cfill=gray
             endpoints=1.9 to 2.3 by 0.05 normal;
  inset mean std skewness / header='descriptive';
run;
```

- Probability plots:

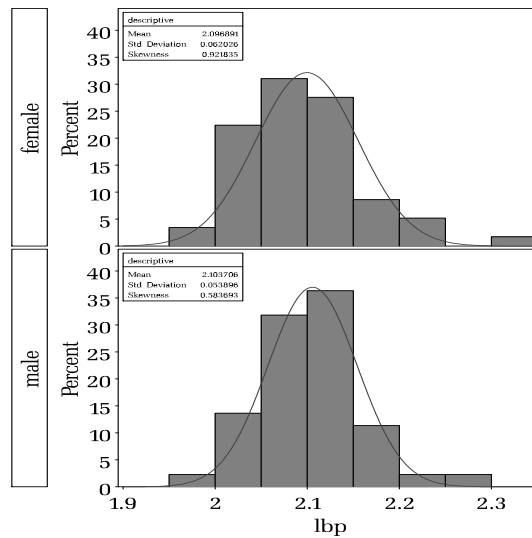
```
proc univariate plot normal gout=plotud data=sasuser.bp;
  var lbp; class sex;
  probplot / height=3 normal(mu=EST sigma=EST l=33);
run;
```

- Box plots

```
proc sort data=sasuser.bp; by sex; run;
proc boxplot gout=plotud data=sasuser.bp;
  plot lbp*sex / height=3 boxstyle=schematic;
run;
```

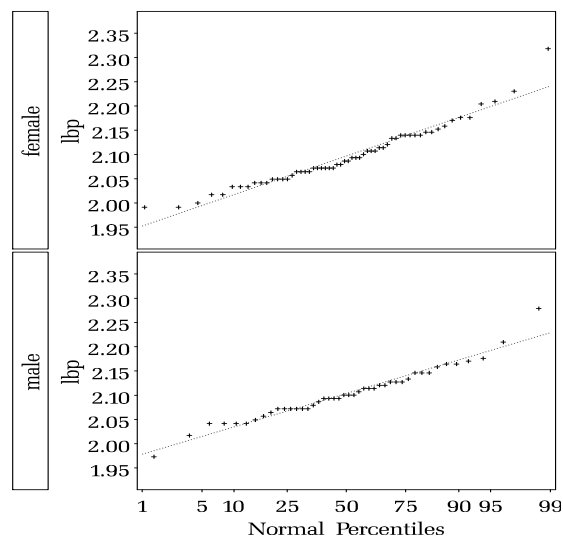
SAS graphics, November 2007

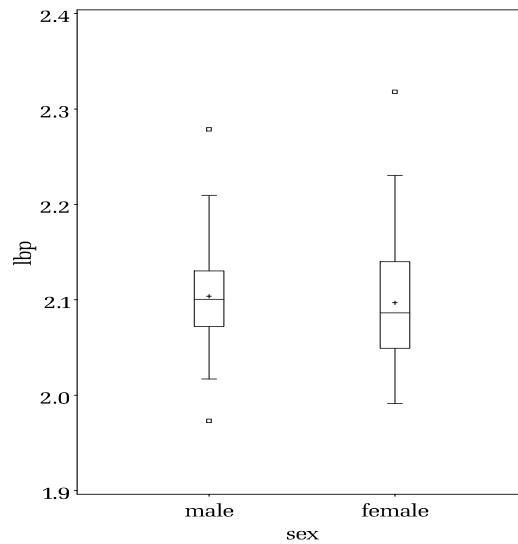
5



SAS graphics, November 2007

6

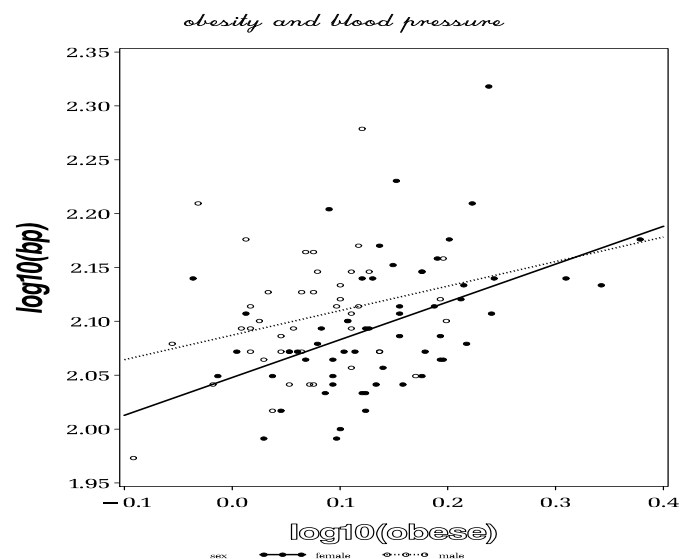




Scatter plots

Facilities in gplot:

```
proc gplot gout=plotud data=sasuser.bp;
  plot lbp*lobese=sex
  / haxis=axis1 vaxis=axis2 frame;
axis1 value=(H=2)
      minor=NONE
      label=(H=3 F=swissbe 'log10(obese)');
axis2 order=1.95 to 2.35 by 0.05
      length=12 cm
      value=(H=2)
      minor=NONE
      label=(A=90 R=0 H=3 F=swissbi 'log10(bp)');
symbol1 v=dot i=r1 c=BLACK l=1 w=2;
symbol2 v=circle i=r1 c=BLACK l=33 w=2;
title1 F=cscript h=3 'obesity and blood pressure';
run;
```



Symbol statements

One symbol statement for each group
(each value of **sex**)

Options:

- **v=circle**: plotting symbol: circle/dot/star
- **h=2**: the size of the plotting symbol
(default 1)
- **i=none**: interpolation method: none/join/rl/rlcli95
- **c=black**: colour of points: black/red/blue
- **l=1**: line type,
1:solid, 2-46: different dashings
- **r=3**:
number of repetitions of this symbol statement

Plotting symbols

'v= ' in symbol statements

VALUE=	Plot Symbol	VALUE=	Plot Symbol	VALUE=	Plot Symbol
PLUS	+	- (underscore)	⊖	+ (plus)	⊕
X	X	* (double quote)	⌘	> (greater than)	⌢
STAR	*	# (pound sign)	♥	.	⋮
SQUARE	□	\$ (dollar sign)	◇	< (less than)	ℏ
DIAMOND	◇	% (percent)	⊞	,	⋈
TRIANGLE	△	& (ampersand)	⊞	/ (slash)	Ψ
HASH	#	' (single quote)	⌘	? (question mark)	ℙ
Y	Y	= (equals)	☆	((left parenthesis)	℄
Z	Z	- (hyphen)	⊖	) (right parenthesis)	℄
PAW	⋈	@ (at)	⊞	:	*
POINT	⋈	* (asterisk)	⊞		
DOT	●				
CIRCLE	○				

Note: The special symbols in this table are listed in default order.
Note: Only the values in column one are specified by name. The values in columns two and three are specified only by character. The names of the characters are included for clarity only.

Interpolation methods

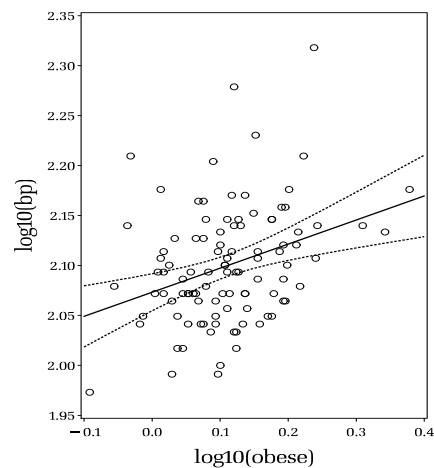
'i= ' (or 'interpol= ')
in symbol statements

Table 18.2 Selected Interpolation Methods		
If your data have . . .	Then you might choose one of these methods . . .	And specify INTERPOL=
one Y value for each X value	join	JOIN
	fitting a regression line	R<L C Q><options>
	needle	NEEDLE
	spline	SPLINE<options>
	spline with Lagrange interpolation	L<1 3 5><options>
	spline with user-defined smoothing	SM<0 . . . 99><options>
one or more Y values for each X value	fitting a regression line	R<L C Q><options>
	spline	SPLINE<options>
	spline with Lagrange interpolation	L<1 3 5><options>
	spline with user-defined smoothing	SM<0 . . . 99><options>
	box plots	BOX<options>
several Y values for each X value	high-low or high-low-close	HILO<L><options>
	standard deviation	STD<1 2 3><options>

Note: If you do not specify an interpolation method, the GPLOT procedure simply marks the data points with the plot symbol. This is equivalent to specifying INTERPOL=NONE.

Confidence interval for regression line

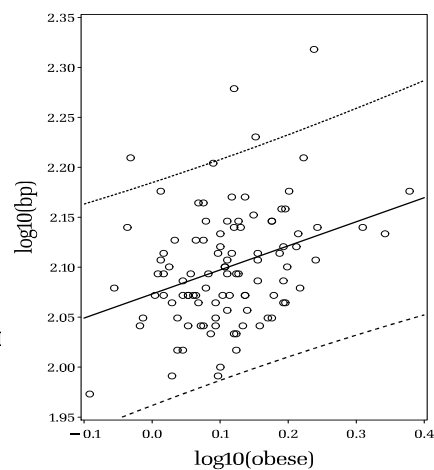
```
proc gplot;
  plot lbp*lobese
  / haxis=axis1 vaxis=axis2 frame;
axis1 value=(H=2) minor=NONE
      label=(H=3 'log10(obese)');
axis2 order=1.95 to 2.35 by 0.05
      value=(H=2) minor=NONE
      label=(A=90 R=0 H=3 'log10(bp)');
symbol1 v=circle i=rlclm95
        c=BLACK h=2 l=1 w=2;
run;
```



Prediction limits:
(normal region
for single observations)

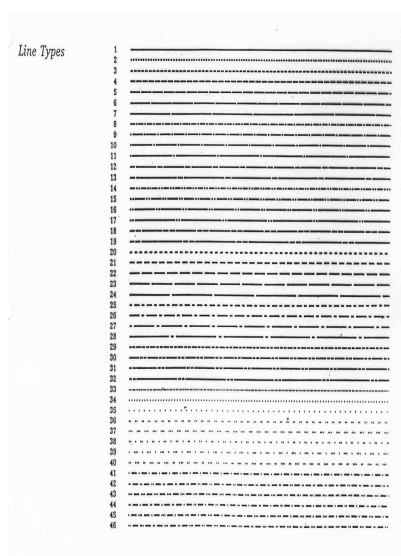
The only change
from the previous graph:

```
symbol1 v=circle i=rlcli95 c=BLACK ;
```



Line types

'l=' in symbol statements



Fonts in SAS

'f=' in symbol, axis or title

Type Style	Font Name	Type Sample	Uniform Font
Brush	BRUSH	<i>A B C a b c 1 2 3</i>	
Century			
Bold	CENTB	A B C a b c 1 2 3	CENTBU
Bold Empty	CENTBE	A B C a b c 1 2 3	
Bold Italic	CENTBI	<i>A B C a b c 1 2 3</i>	CENTBIU
Bold Italic Empty	CENTBIE	<i>A B C a b c 1 2 3</i>	
Expanded	CENTX	A B C a b c 1 2 3	CENTXU
Expanded Empty	CENTXE	A B C a b c 1 2 3	
Expanded Italic	CENTXI	<i>A B C a b c 1 2 3</i>	CENTXIU
Expanded Italic Empty	CENTXIE	<i>A B C a b c 1 2 3</i>	
German	GERMAN	A B C a b c 1 2 3	GERMANU
German Italic	GITALIC	<i>A B C a b c 1 2 3</i>	GITALICU
Hershey			
Sans Serif	SIMPLEX	A B C a b c 1 2 3	SIMPLEXU
Sans Serif Bold	DUPLEX	A B C a b c 1 2 3	DUPLEXU
Serif	COMPLEX	A B C a b c 1 2 3	COMPLEXU
Serif Bold	TRIPLEX	A B C a b c 1 2 3	TRIPLEXU
Serif Bold Italic	TITALIC	<i>A B C a b c 1 2 3</i>	TITALICU
Serif Italic	ITALIC	<i>A B C a b c 1 2 3</i>	ITALICU
Old English	OLDENG	A B C a b c 1 2 3	OLDENGU
Script	SCRIPT	<i>A B C a b c 1 2 3</i>	
Cacript	* CSCRIPT	<i>A B C a b c 1 2 3</i>	
Simulate	SIMULATE	A B C a b c 1 2 3	SIMULATE
Swiss	SWISS	A B C a b c 1 2 3	SWISSU
Empty	SWISSE	A B C a b c 1 2 3	
Bold	SWISSB	A B C a b c 1 2 3	SWISSBU
Bold Empty	* SWISSBE	A B C a b c 1 2 3	
Bold Italic	* SWISSBI	<i>A B C a b c 1 2 3</i>	SWISSBIU

(continued)

AXIS specifications

- **length=12cm**: the length of the axis
should not be used for the screen
- **value=(h=2)**: the size of the digits on the axis
- **minor=9**: number of tickmarks between the numbers, may be set to none
- **label=(A=90 R=0 h=2 'text')** specifies the axis text, the size of this, and its direction
 - **A=90**: The whole text has to be rotated 90 degrees counterclockwise, so that it fits the Y axis
 - **R=0** this may make *the letters* slant
- **order=(0 to 10 by 1)** specifies the desired numbers on the axis

More plots on the same sheet

```

goptions display;
proc greplay tc = tmpcat nofs ;

  tdef toxtoc des = '2 gange 2'
    1/llx= 0 lly= 55 ulx= 0 uly= 100
      urx= 45 ury= 100 lrx= 45 lry= 55
    2/llx= 55 lly= 55 ulx= 55 uly= 100
      urx= 100 ury= 100 lrx= 100 lry= 55
    3/llx= 0 lly= 0 ulx= 0 uly= 45
      urx= 45 ury= 45 lrx= 45 lry= 0
    4/llx= 55 lly= 0 ulx= 55 uly= 45
      urx= 100 ury= 45 lrx= 100 lry= 0 ;

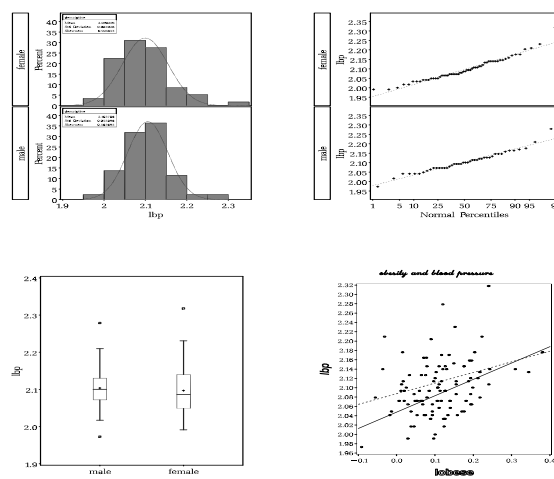
  list tc ;
  quit ;
run ;

```

More plots on the same sheet, cont.

```
proc greplay igout=plotud tc=tmpcat gout=navn nofs ;
  list igout ;
  list tc ;
  template toxt0 ;
  treplay 1:1
          2:2
          3:3
          4:4;

  quit ;
run ;
```



Graphical goptions

(may be put anywhere in the program)

- **display/nodisplay** may suppress a graph and save it instead (catalogue must be specified in the plot)
- **hsize=/vsize=** specifies the size of the plots
- **device=:** name on device - driver
- **rotate:** rotate the plot 90 degrees
- **hby= :** the height of the title **SEX='male'**, when the plot is made **by sex**;
- **reset=:** cancels previous goptions (most useful: **reset=all**)