General SAS Skills and Knowledge:

- SAS Windows (Program, Log, Output, Results, Explorer)
- SAS Menus (Open files, Submit files, View windows)
- Clearing SAS Windows (Log and Output)
- Creating and using a SAS library
- Working with SAS datasets in a library
- Working with SAS datasets in the work directory
- Viewing the contents of a SAS dataset using PROC CONTENTS

```
proc contents data=bio.whas500;
run;

proc contents data=bio.whas500 varnum;
ods select position;
run;
```

- Viewing the contents of a SAS dataset using PROC PRINT

```
proc print data=bio.whas500;
run;

proc print data=bio.whas500 (obs=10);
var hr gender cvd year;
run;
```

- Sorting a dataset using PROC SORT

```
proc sort data=bio.whas500;
by hr;
run;
```

- Creating new datasets, new variables, and labeling variables using a DATA step

```
data library.new;
set library.old;
var4 = var1/100;
var5 = var2*1000;
if 0 <= var4 <= 1.6 then var6 = 0;
if var4 > 1.6 then var6 = 1;
label       var1="Height (cm)"
            var2="Weight (kg)"
            var3="Gender"
            var4="Height (m)"
            var5="Weight (g)"
            var6="Height Over 1.6 m?";
run;
```

- Create and clear custom titles

```
/* Create your own title */
title 'My Title';

/* Clear all Default Titles */
title;
```

- Using ODS RTF and ODS PDF to export output to common formats
  - Basic commands:
    (If this doesn't work specify the file name and open the file manually)

```
ods pdf;
/* SAS code for which output is requested */
ods pdf close;

ods rtf;
/* SAS code for which output is requested */
ods rtf close;
```

  - More advanced examples:

```
ods pdf notoc file="C:\MySAS\output.pdf";
/* SAS code for which output is requested */
ods pdf close;

ods rtf bodytitle file="C:\MySAS\output.rtf";
/* SAS code for which output is requested */
ods rtf close;
```

- Using ODS GRAPHICS to obtain additional graphs from common analyses (this may or may not be needed in SAS 9.3 and beyond – try and see)

  Note: Throughout we have included the ods graphics on/off command when they produce useful output in SAS 9.2. Most readers should now be using SAS version 9.3 or higher which, by default, provide ods graphics without the need for these commands and, in that case, they can usually be ignored.

```
ods graphics on;
/* SAS Procedures */
ods graphics off;
```

Advanced SAS Skills:

- Common SAS options:

```
Options nodate nonumber;
```

- Using PROC FORMAT and the FORMAT statement in a DATA step to associate a user defined format with a variable in a dataset

```
proc format;
value YesNoFmt 1='Yes'
               0='No';
value Sex01Ft  0='Male'
               1='Female';
value yr       1='1997'
               2='1999'
               3='2001';
run;

data WHAS500_formatted;
set  whas500_unformatted;
format Gender Sex0Fmt.
       cvd afb sho chf av3 YesNoFmt.
       year yr.;
run;
```

- Using ODS TRACE/SELECT/EXCLUDE to obtain only needed output

```
ods trace on;
/* SAS Procedures */
ods trace off;
```

- Adding text to ODS Output

```
/* Add Text to RTF output
   For PDF output change the rtf below to pdf */
   ods escapechar='^';
   ods rtf text='^S={just=center font=("Times Roman",22PT,Bold)}
         Descriptive Statistics using SAS';
```

- Adjusting graph sizes for ODS graphics

```
ods graphics / width=5in height=4in;
```

- Using the INSET statement in PROG SGPLOT and other advanced graphics topics
- Using a WHERE statement in a PROC step to analyze a subset of the data

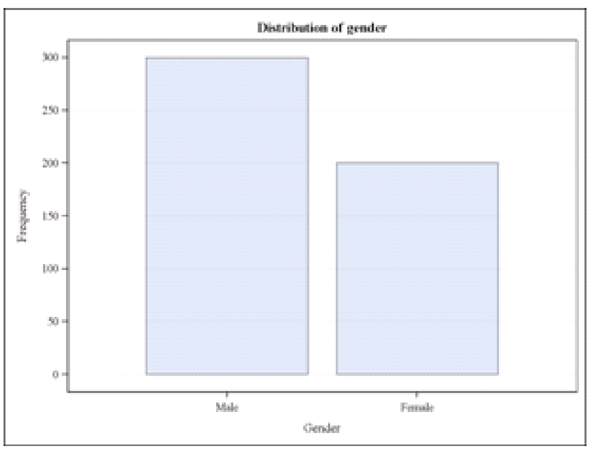
```
where 40 <= diasbp <= 120;
```

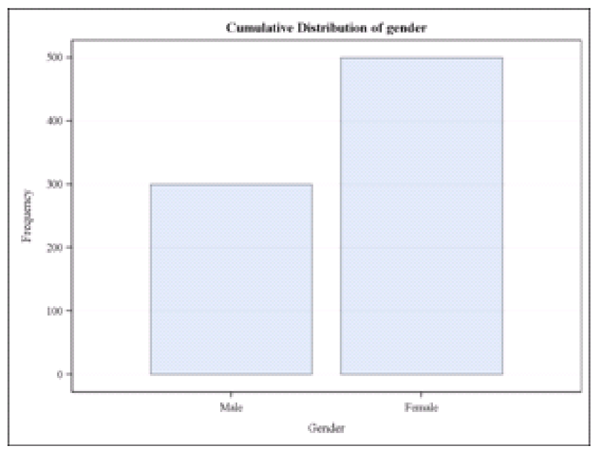
Descriptive Methods – One Categorical Variable

- Frequency distributions and bar charts using PROC FREQ

```
ods graphics on;
proc freq  data=bio.whas500;
tables gender;
run;
ods graphics off;
```

Gender				
gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	300	60.00	300	60.00
Female	200	40.00	500	100.00





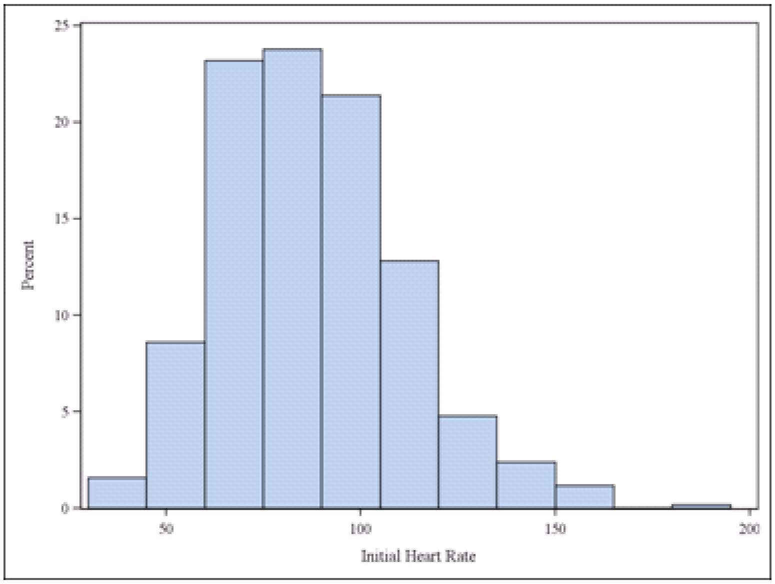
Multiple variables can be used in a given tables statement as in the example below:

```
proc freq  data=bio.whas500;
tables afb sho chf year;  /* List variables separated by spaces */
run;
```

Descriptive Methods – One Quantitative Variable

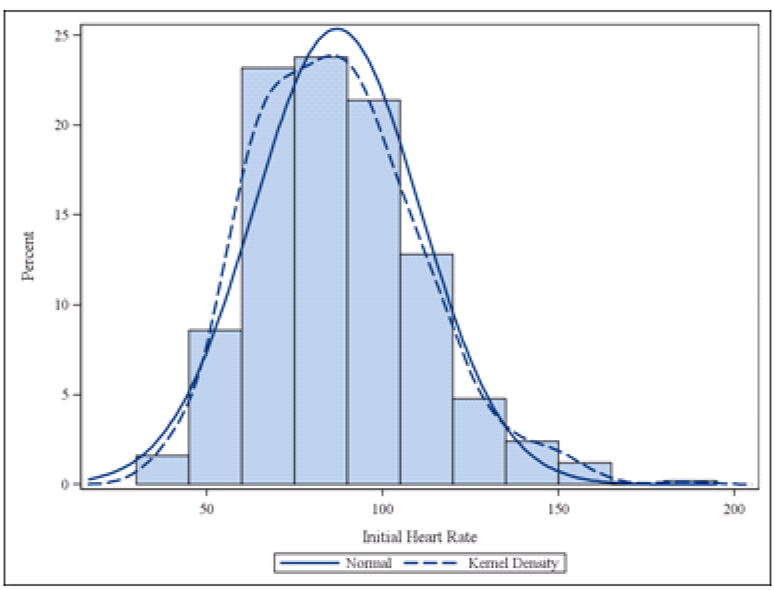
- Histograms using PROC SGPLOT

```
proc sgplot data=bio.whas500;
histogram hr; /* can only have one variable in HISTOGRAM statement */
run;
```



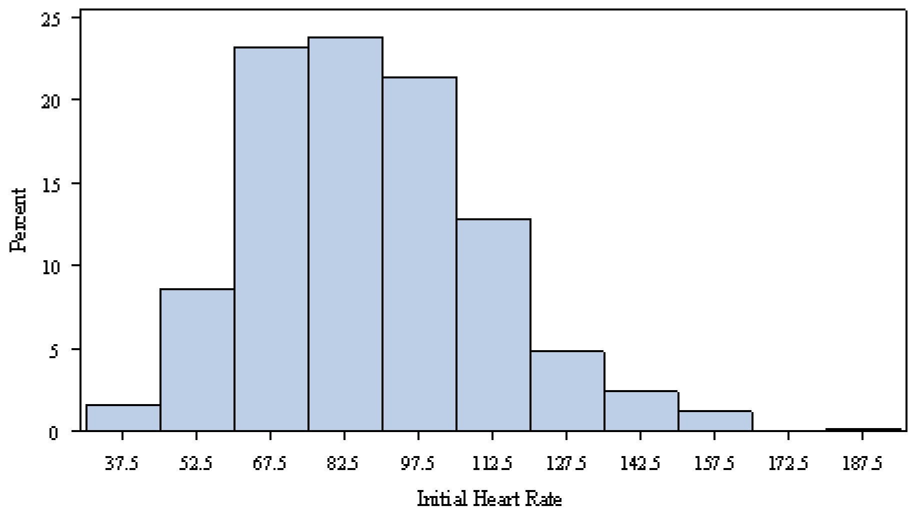
Normal or non-parametric (kernel) density estimates can be added using the DENSITY statement

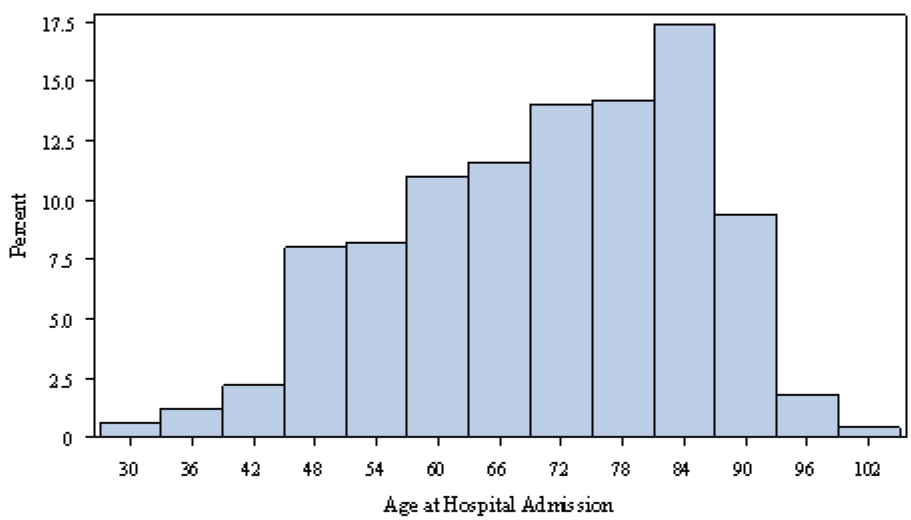
```
proc sgplot  data=bio.whas500;
histogram hr;
density hr / type=normal;
density hr / type=kernel;
run;
```



- Histograms using PROC UNIVARIATE (density curves can also be added but generally SGPLOT is preferred for this)

```
proc univariate data=bio.whas500 noprint; /* NOPRINT = No text output*/
var hr age;
histogram;
run;
```

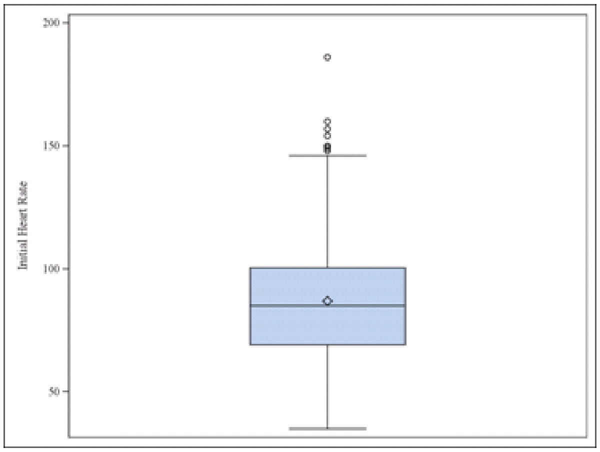


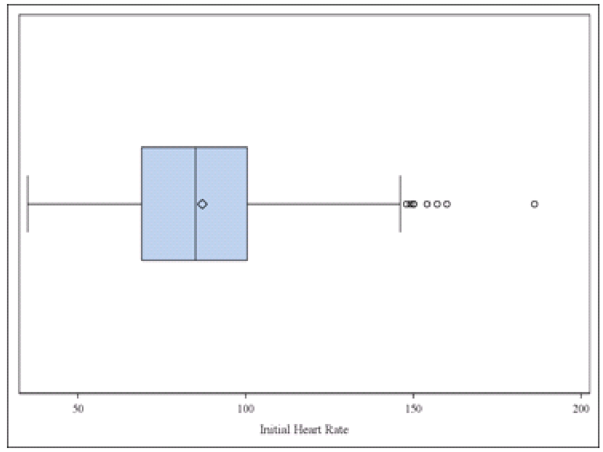


- Boxplots using PROC SGPLOT

```
proc sgplot data=bio.whas500;
vbox hr;     /* VBOX for Vertical Boxplots */
run;

proc sgplot data=bio.whas500;
hbox hr;     /* HBOX for Horizontal Boxplots */
run;
```





- Summary statistics using PROC MEANS

```
/* Default Summary Statistics */
proc means data=bio.whas500;
var hr;      /* can have more than one variable in the VAR statement */
run;
```

Analysis Variable : hr Initial Heart Rate				
N	Mean	Std Dev	Minimum	Maximum
500	87.0180000	23.5862311	35.0000000	186.0000000

```
/* Requesting different Summary Statistics */
proc means data=bio.whas500 min q1 median q3 max fw=7 maxdec=2;
var hr;
run;
```

Analysis Variable : hr Initial Heart Rate				
Minimum	Lower Quartile	Median	Upper Quartile	Maximum
35.00	69.00	85.00	100.50	186.00

- Summary statistics using PROC UNIVARIATE (lengthy output)

```
proc univariate data=bio.whas500 all;   /* ALL option requests all
UNIVARIATE output possible */
var hr;      /* can have more than one variable in the VAR statement */
run;
```

The UNIVARIATE Procedure
Variable: hr (Initial Heart Rate)

Moments			
N	500	**Sum Weights**	500
Mean	87.018	**Sum Observations**	43509
Std Deviation	23.5862311	**Variance**	556.310297
Skewness	0.56676662	**Kurtosis**	0.47176453
Uncorrected SS	4063665	**Corrected SS**	277598.838
Coeff Variation	27.1050025	**Std Error Mean**	1.05480832

Basic Statistical Measures			
Location		**Variability**	
Mean	87.0180	**Std Deviation**	23.58623
Median	85.0000	**Variance**	556.31030
Mode	100.0000	**Range**	151.00000
		Interquartile Range	31.50000

Modes	
Mode	**Count**
100	16

Basic Confidence Limits Assuming Normality			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	87.01800	84.94559	89.09041
Std Deviation	23.58623	22.20935	25.14650
Variance	556.31030	493.25519	632.34644

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	82.49651	**Pr > |t|**	<.0001
Sign	**M**	250	**Pr >= |M|**	<.0001
Signed Rank	**S**	62625	**Pr >= |S|**	<.0001

Location Counts: Mu0=0.00	
Count	**Value**
Num Obs > Mu0	500
Num Obs ^= Mu0	500
Num Obs < Mu0	0

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.980406	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.049799	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.278942	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	1.971014	**Pr > A-Sq**	<0.0050

Trimmed Means								
Percent Trimmed in Tail	**Number Trimmed in Tail**	**Trimmed Mean**	**Std Error Trimmed Mean**	**95% Confidence Limits**		**DF**	**t for H0: Mu0=0.00**	**Pr > |t|**
25.00	125	85.20000	1.139391	82.95593	87.44407	249	74.77676	<.0001

Winsorized Means								
Percent Winsorized in Tail	**Number Winsorized in Tail**	**Winsorized Mean**	**Std Error Winsorized Mean**	**95% Confidence Limits**		**DF**	**t for H0: Mu0=0.00**	**Pr > |t|**
25.00	125	84.85000	1.140535	82.60367	87.09633	249	74.39492	<.0001

Robust Measures of Scale		
Measure	**Value**	**Estimate of Sigma**
Interquartile Range	31.50000	23.35098
Gini's Mean Difference	26.37960	23.37831
MAD	16.00000	23.72160
Sn	23.85200	23.85200
Qn	22.21900	22.05141

Quantiles (Definition 5)								
						Order Statistics		
Quantile	**Estimate**	**95% Confidence Limits Assuming Normality**		**95% Confidence Limits Distribution Free**		**LCL Rank**	**UCL Rank**	**Coverage**
100% Max	186.0							
99%	150.0	138.1125	146.1037	146	186	491	500	96.23
95%	128.5	122.7901	129.1550	123	139	466	486	95.89
90%	117.0	114.5712	120.1696	114	121	437	464	95.63
75% Q3	100.5	100.6978	105.2915	99	105	357	395	95.01
50% Median	85.0	84.9456	89.0904	83	88	229	273	95.08
25% Q1	69.0	68.7445	73.3382	67	72	106	144	95.01
10%	59.0	53.8664	59.4648	57	61	37	64	95.63
5%	54.0	44.8810	51.2459	47	56	15	35	95.89
1%	42.0	27.9323	35.9235	35	45	1	10	96.23
0% Min	35.0							

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
35	1	150	496
36	3	154	497
36	2	157	498
38	4	160	499
42	7	186	500

Extreme Values					
Lowest			Highest		
Order	Value	Freq	Order	Value	Freq
1	35	1	101	150	3
2	36	2	102	154	1
3	38	1	103	157	1
4	42	3	104	160	1
5	44	1	105	186	1

Frequency Counts			
		Percents	
Value	Count	Cell	Cum
35	1	0.2	0.2
36	2	0.4	0.6
38	1	0.2	0.8
42	3	0.6	1.4
44	1	0.2	1.6
All Observed Data Values			
148	1	0.2	98.4
149	1	0.2	98.6
150	3	0.6	99.2
154	1	0.2	99.4
157	1	0.2	99.6
160	1	0.2	99.8
186	1	0.2	100.0

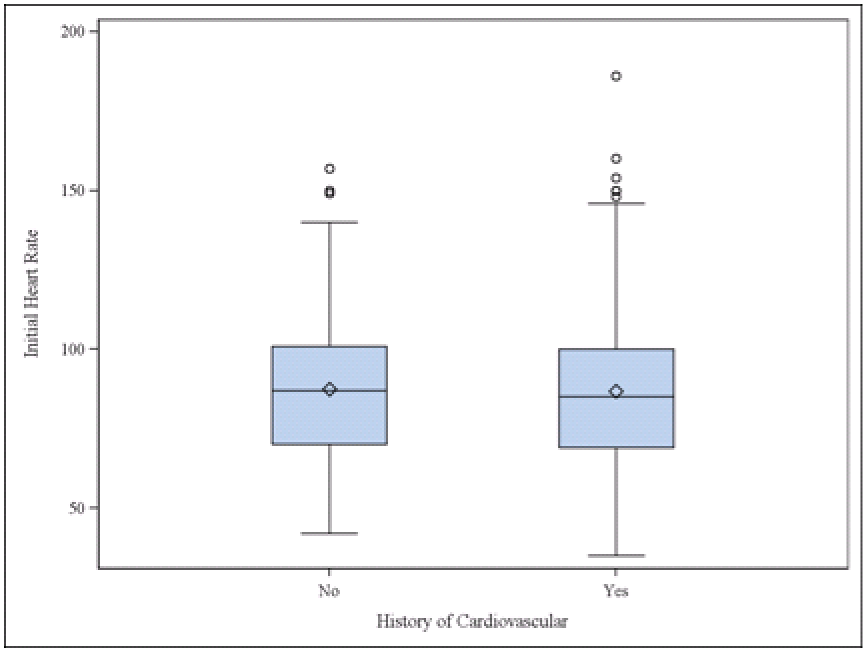
```
                Histogram                          #             Boxplot
185+*                                              1                0
   .
   .*                                              1                0
155+***                                            5                0
   .****                                           7                0
   .*****                                         10                |
125+**********                                    19                |
   .*********************                         41                |
   .*****************************                 57             +-----+
 95+*************************************         73             |     |
   .********************************************  87             *--+--*
   .***********************************           70             |     |
 65+***************************************       78             +-----+
   .******************                            35                |
   .******                                        12                |
 35+**                                             4                |
    ----+----+----+----+----+----+----+----+----
    * may represent up to 2 counts


                       Normal Probability Plot
        185+                                                  *
           |
           |                                                  *
        155+                                               ****
           |                                            ****+++
           |                                          ***+++
        125+                                       ****+
           |                                   *****
           |                               *****
         95+                           *****
           |                      ******
           |                  +****
         65+            *******
           |       ******+
           | ******+++
         35+**++++
            +----+----+----+----+----+----+----+----+----+----+
                 -2        -1         0        +1        +2
```

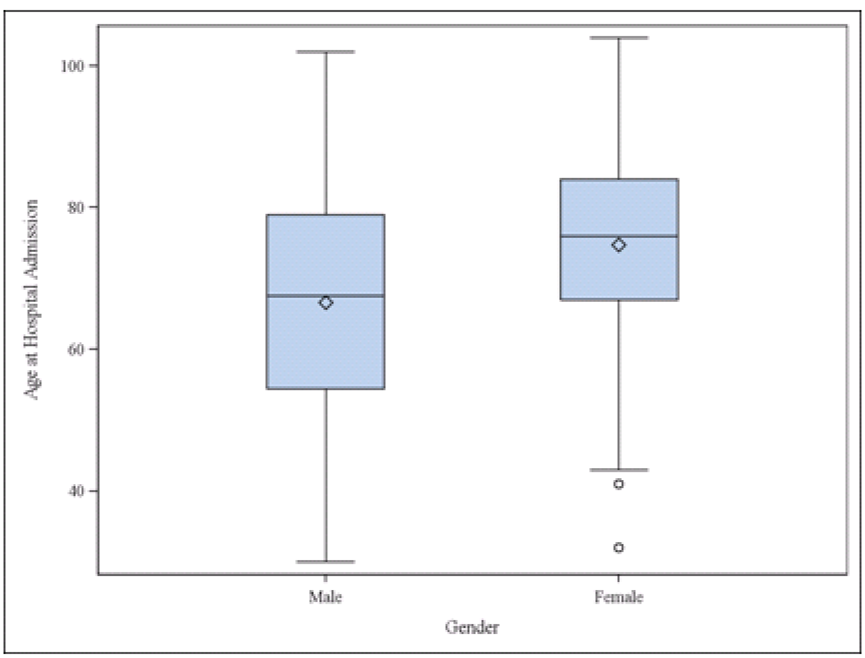
Descriptive Methods – One Quantitative and One Categorical Variable

- Boxplots of the quantitative variable by the categorical variable using PROC SGPLOT (similar results can be obtained using HBOX)

```
proc sgplot data=bio.whas500;
vbox hr / category=cvd; /* can only have 1 variable in VBOX statement*/
run;
```




```
proc sgplot data=bio.whas500;
vbox age / category=gender;
run;
```



- Summaries of the quantitative variable by the categorical variable using PROC MEANS

```
proc means data=bio.whas500;
class cvd;
var hr; /* can have more than one variable in the VAR statement */
run;
```

Analysis Variable : hr Initial Heart Rate						
History of Cardiovascular	**N Obs**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
No	125	125	87.4400000	24.0390811	42.0000000	157.0000000
Yes	375	375	86.8773333	23.4641812	35.0000000	186.0000000

```
proc means data=bio.whas500;
class gender;
var age;
run;
```

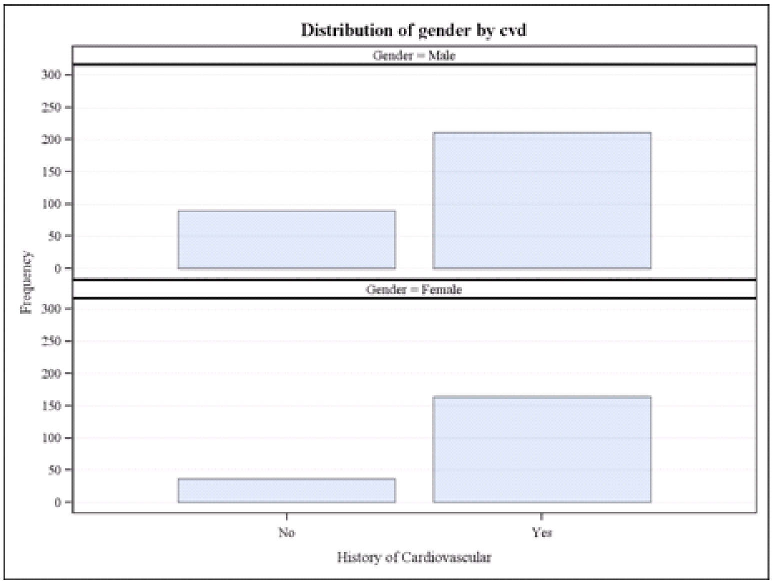
Analysis Variable : age Age at Hospital Admission						
Gender	**N Obs**	**N**	**Mean**	**Std Dev**	**Minimum**	**Maximum**
Male	300	300	66.5966667	14.9427219	30.0000000	102.0000000
Female	200	200	74.7200000	12.3007374	32.0000000	104.0000000

Descriptive Methods – Two Categorical Variables

- Two-Way Tables (Contingency Tables) using PROC FREQ

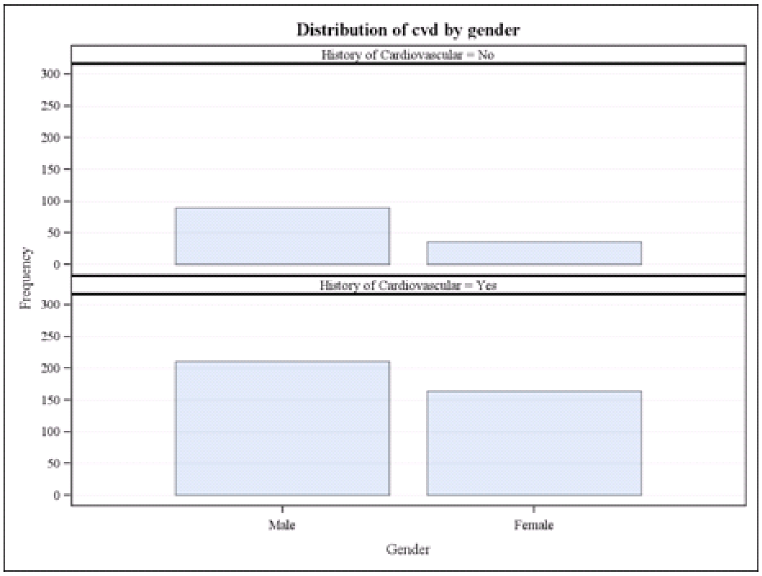
```
ods graphics on;
proc freq data=bio.whas500;
tables gender*cvd;
run;
ods graphics off;
```

Table of gender by cvd			
gender(Gender)	cvd(History of Cardiovascular)		
Frequency Percent Row Pct Col Pct	No	Yes	Total
Male	89 17.80 29.67 71.20	211 42.20 70.33 56.27	300 60.00
Female	36 7.20 18.00 28.80	164 32.80 82.00 43.73	200 40.00
Total	125 25.00	375 75.00	500 100.00




```
ods graphics on;
proc freq data=bio.whas500;
tables cvd*gender;
run;
ods graphics off;
```

Table of cvd by gender			
cvd(History of Cardiovascular)	gender(Gender)		
Frequency Percent Row Pct Col Pct	Male	Female	Total
No	89 17.80 71.20 29.67	36 7.20 28.80 18.00	125 25.00
Yes	211 42.20 56.27 70.33	164 32.80 43.73 82.00	375 75.00
Total	300 60.00	200 40.00	500 100.00



Any combination of variables may be used in PROC FREQ with parentheses containing groups of variables to be treated the same way

```
proc freq data=bio.whas500;
tables gender (cvd chf)*gender;
run;
```

Gender				
gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Male	300	60.00	300	60.00
Female	200	40.00	500	100.00

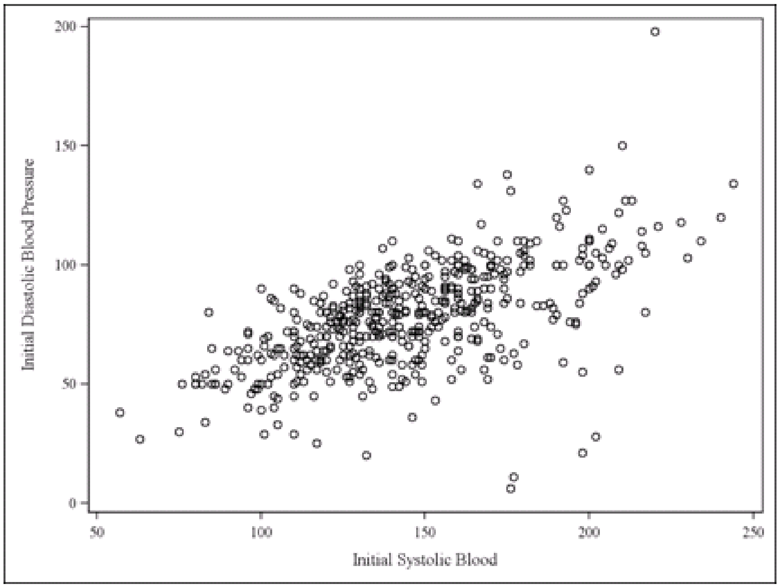
Table of cvd by gender			
cvd(History of Cardiovascular)	gender(Gender)		
Frequency Percent Row Pct Col Pct	Male	Female	Total
No	89 17.80 71.20 29.67	36 7.20 28.80 18.00	125 25.00
Yes	211 42.20 56.27 70.33	164 32.80 43.73 82.00	375 75.00
Total	300 60.00	200 40.00	500 100.00

Table of chf by gender			
chf(Congestive Heart Complications)	gender(Gender)		
Frequency Percent Row Pct Col Pct	Male	Female	Total
No	225 45.00 65.22 75.00	120 24.00 34.78 60.00	345 69.00
Yes	75 15.00 48.39 25.00	80 16.00 51.61 40.00	155 31.00
Total	300 60.00	200 40.00	500 100.00

Descriptive Methods – Two Quantitative Variables

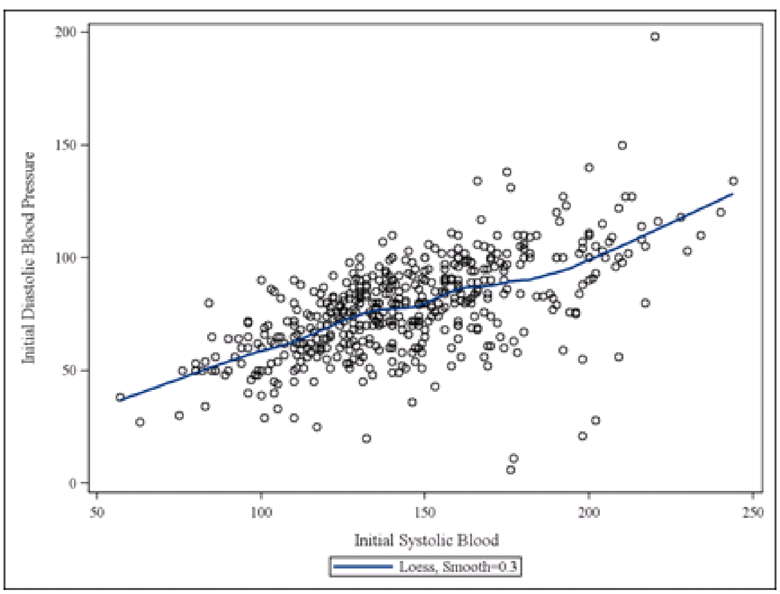
- Scatterplots using PROC SGPLOT

```
proc sgplot data=bio.whas500;
scatter y=diasbp x=sysbp;    /* Only one plot per SGPLOT
                                Can overlay plots */
run;
```



- Scatterplots with LOESS smoothed trend line using PROC SGPLOT

```
proc sgplot data=bio.whas500;
loess y=diasbp x=sysbp / smooth=0.3;  /* smooth ranges from
                                         0=no smoothing to 1=linear */
run;
```



- Calculate Pearson's correlation coefficient (only meaningful for approximately linear trend)

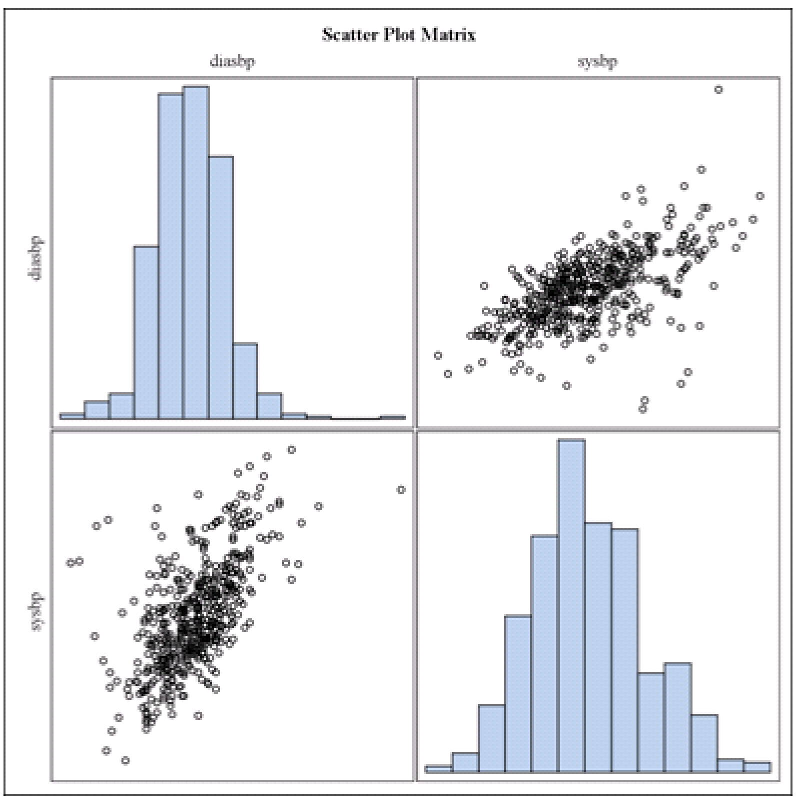
```
ods graphics on;
proc corr data=bio.whas500 plots=matrix(histogram);
var diasbp sysbp;
run;
ods graphics off;
```

The CORR Procedure

2 Variables:	diasbp sysbp

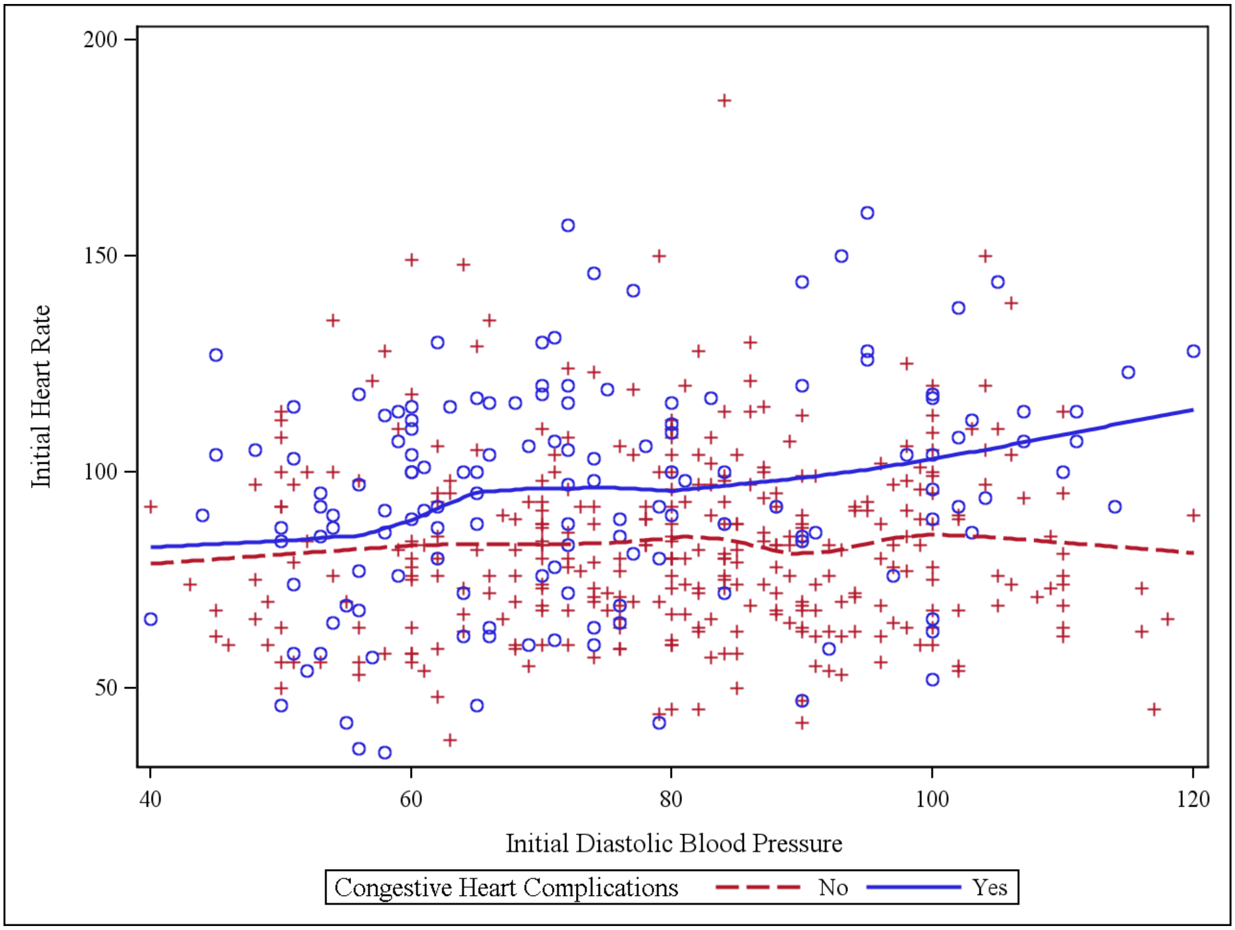
Simple Statistics							
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**	**Label**
diasbp	500	78.26600	21.54529	39133	6.00000	198.00000	Initial Diastolic Blood Pressure
sysbp	500	144.70400	32.29486	72352	57.00000	244.00000	Initial Systolic Blood

Pearson Correlation Coefficients, N = 500 Prob > |r| under H0: Rho=0		
	diasbp	**sysbp**
diasbp Initial Diastolic Blood Pressure	1.00000	0.61092 <.0001
sysbp Initial Systolic Blood	0.61092 <.0001	1.00000



- Adding information about a categorical variable to a scatterplot

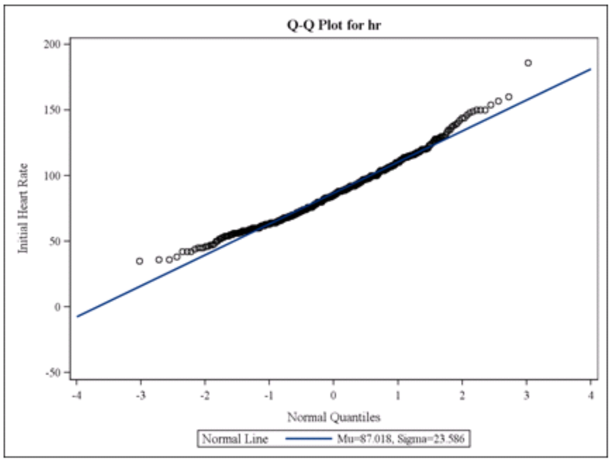
```
proc sgplot data=bio.whas500;
where 40 <= diasbp <= 120;
loess y=hr x=diasbp / group=chf smooth=0.5;
run;
```

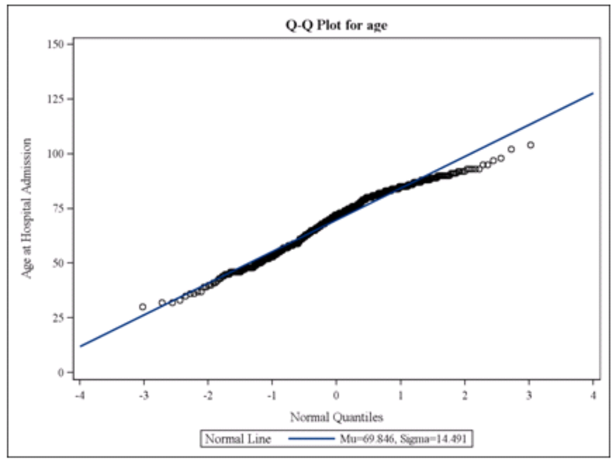


Inferential Statistics – Investigating Normality

- Normal probability plots using PROC UNIVARIATE

```
ods graphics on;
proc univariate data=bio.whas500 noprint;
var hr age;
qqplot / normal(mu=est sigma=est);
run;
ods graphics off;
```





- Common tests of normality using PROC UNIVARIATE

```
proc univariate data=bio.whas500 normaltest;
var hr age;
ods select testsfornormality;
run;
```

The UNIVARIATE Procedure
Variable: hr (Initial Heart Rate)

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.980406	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.049799	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.278942	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	1.971014	**Pr > A-Sq**	<0.0050

The UNIVARIATE Procedure
Variable: age (Age at Hospital Admission)

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.973506	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.082251	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.732381	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	4.44345	**Pr > A-Sq**	<0.0050

Inferential Statistics – One Sample – Binomial Proportion

- Tests and Intervals for one binomial proportion using PROC FREQ. The following conducts a test of the null hypothesis that the proportion of congestive heart complications in the population is 25%. Confidence intervals are also provided via this output.

```
proc freq data=bio.whas500;
tables chf / binomial(level="Yes" p=0.25);
run;
```

Congestive Heart Complications				
chf	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	345	69.00	345	69.00
Yes	155	31.00	500	100.00

Binomial Proportion for chf = Yes	
Proportion	0.3100
ASE	0.0207
95% Lower Conf Limit	0.2695
95% Upper Conf Limit	0.3505
Exact Conf Limits	
95% Lower Conf Limit	0.2697
95% Upper Conf Limit	0.3526

Test of H0: Proportion = 0.25	
ASE under H0	0.0194
Z	3.0984
One-sided Pr > Z	0.0010
Two-sided Pr > |Z|	0.0019

Sample Size = 500

- Power calculations for one binomial proportion using PROC POWER

```
proc power;
onesamplefreq
    test=z
    method=normal
    proportion = .30
    nullproportion = .25
    ntotal = 100
    alpha = 0.05
    sides = 1
    power=.;
run;
```

The POWER Procedure
Z Test for Binomial Proportion

Fixed Scenario Elements	
Method	Normal approximation
Number of Sides	1
Null Proportion	0.25
Alpha	0.05
Binomial Proportion	0.3
Total Sample Size	100
Variance Estimate	Null Variance

Computed Power
Power
0.322

- Sample size calculations for one binomial proportion using PROC POWER

```
proc power;
onesamplefreq
    test=z
    method=normal
    proportion = .30
    nullproportion = .25
    ntotal = .
    alpha = 0.05
    sides = 1
    power=.90;
run;
```

The POWER Procedure
Z Test for Binomial Proportion

Fixed Scenario Elements	
Method	Normal approximation
Number of Sides	1
Null Proportion	0.25
Alpha	0.05
Binomial Proportion	0.3
Nominal Power	0.9
Variance Estimate	Null Variance

Computed N Total	
Actual Power	N Total
0.900	676

Inferential Statistics – One Sample – Quantitative Variable

- T-tests and confidence intervals for the mean of a normally distributed population using PROC TTEST (confidence intervals for the standard deviation of the population are also reported)

```
ods graphics on;
proc ttest data=bio.whas500 h0=80;
var hr;
run;
ods graphics off;
```

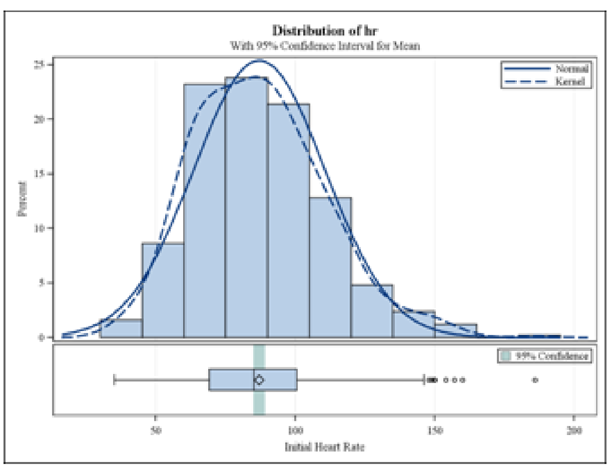
The TTEST Procedure

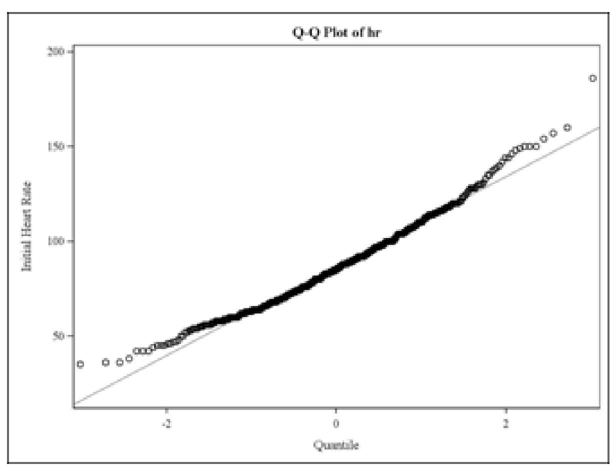
Variable: hr (Initial Heart Rate)

N	Mean	Std Dev	Std Err	Minimum	Maximum
500	87.0180	23.5862	1.0548	35.0000	186.0

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
87.0180	84.9456	89.0904	23.5862	22.2093	25.1465

DF	t Value	Pr > |t|
499	6.65	<.0001





- T-tests and confidence intervals for the mean of a normally distributed population using PROC UNIVARIATE (confidence intervals for the standard deviation and variance of the population are also reported)

```
proc univariate data=bio.whas500 mu0=80 cibasic;
var hr;
ods select testsforlocation basicintervals;
run;
```

The UNIVARIATE Procedure
Variable: hr (Initial Heart Rate)

Basic Confidence Limits Assuming Normality			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	87.01800	84.94559	89.09041
Std Deviation	23.58623	22.20935	25.14650
Variance	556.31030	493.25519	632.34644

Tests for Location: Mu0=80				
Test	**Statistic**		**p Value**	
Student's t	**t**	6.653342	**Pr > |t|**	<.0001
Sign	**M**	43.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	17269.5	**Pr >= |S|**	<.0001

- Power calculations for one sample t-tests using PROC POWER

```
proc power;
onesamplemeans
    test=t
    dist=normal
    mean = 30
    nullmean = 28
    ntotal = 100
    alpha = 0.05
    sides = 1
    std = 6
    power=.;
run;
```

The POWER Procedure
One-sample t Test for Mean

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Number of Sides	1
Null Mean	28
Alpha	0.05
Mean	30
Standard Deviation	6
Total Sample Size	100

Computed Power
Power
0.952

- Sample size calculations for one sample t-tests using PROC POWER

```
proc power;
onesamplemeans
    test=t
    dist=normal
    mean = 30
    nullmean = 28
    ntotal = .
    alpha = 0.05
    sides = 1
    std = 6
    power=.99;
run;
```

The POWER Procedure
One-sample t Test for Mean

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Number of Sides	1
Null Mean	28
Alpha	0.05
Mean	30
Standard Deviation	6
Nominal Power	0.99

Computed N Total	
Actual Power	**N Total**
0.990	144

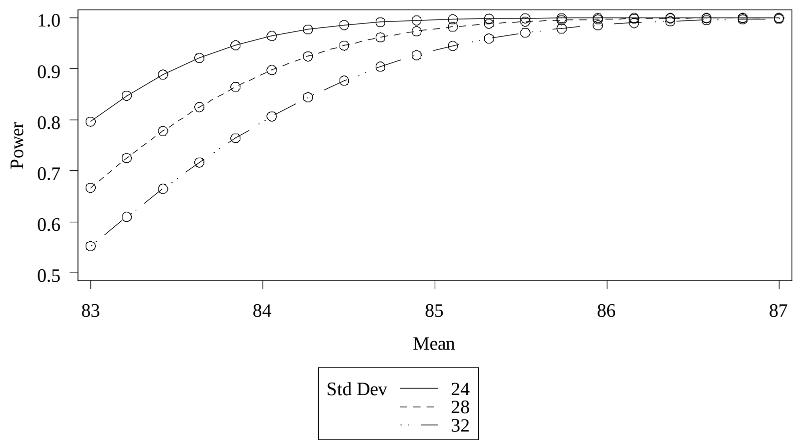
- Sign test and Signed Rank test for the **median** of a population using PROC UNIVARIATE (same code and output as above).

```
proc univariate data=bio.whas500 mu0=80 cibasic;
var hr;
ods select testsforlocation basicintervals;
run;
```

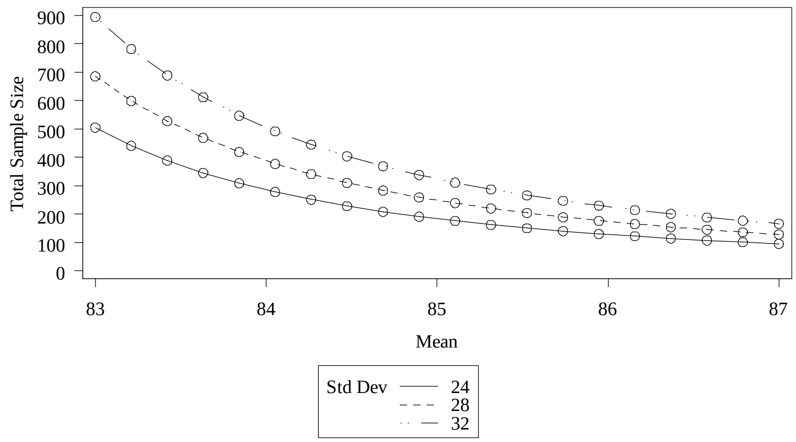
Tests for Location: Mu0=80				
Test	**Statistic**		**p Value**	
Student's t	**t**	6.653342	**Pr > |t|**	<.0001
Sign	**M**	43.5	**Pr >= |M|**	<.0001
Signed Rank	**S**	17269.5	**Pr >= |S|**	<.0001

- Power graphs using PROC POWER (can be requested for any design, not just onesamplemeans)

```
proc power;
onesamplemeans  test=T
                dist=normal
                power=.
                ntotal=500
                nullmean=80
                mean=83 85 87
                std= 24 28 32;
plot x=effect;
run;
```




```
proc power;
onesamplemeans  test=T
                dist=normal
                power=0.8
                ntotal=.
                nullmean=80
                mean=83 85 87
                std= 24 28 32;
plot x=effect;
run;
```



Inferential Statistics – Paired Samples – Quantitative Variable

- For the following we use the data below, entered into SAS directly using a DATA step

```
data example;
input jan apr;
diff=jan-apr;
datalines;
139 104
122 113
126 100
64 88
78 61
run;
```

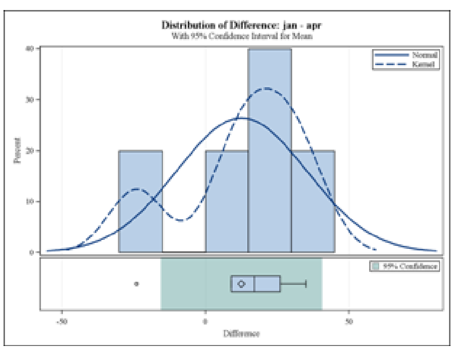
- Paired t-tests using PROC TTEST and the PAIRED statement

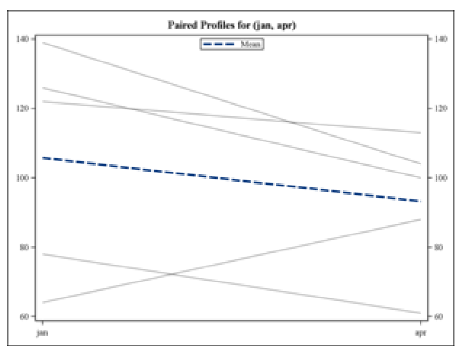
```
ods graphics on;
proc ttest data=example;
paired jan*apr;
run;
ods graphics off;
```

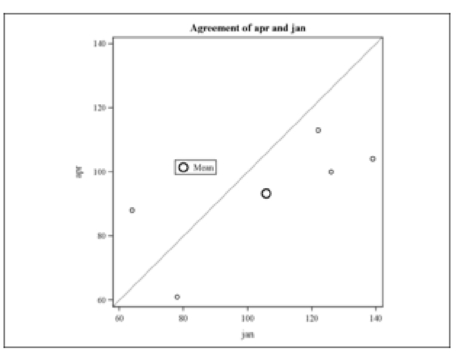
N	Mean	Std Dev	Std Err	Minimum	Maximum
5	12.6000	22.6561	10.1321	-24.0000	35.0000

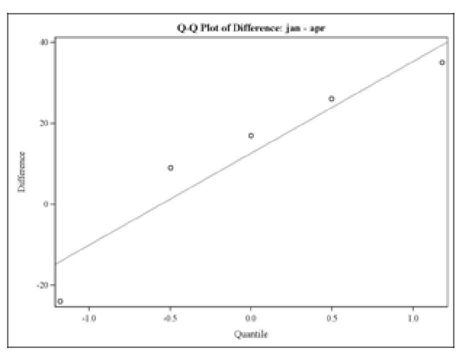
Mean	95% CL Mean		Std Dev	95% CL Std Dev	
12.6000	-15.5313	40.7313	22.6561	13.5740	65.1036

DF	t Value	Pr > |t|
4	1.24	0.2816









- Paired t-tests using PROC UNIVARIATE (this code also conducts the sign test and signed rank test for paired samples). I have requested tests for normality as well.

```
proc univariate data=example cibasic normaltest;
var diff;
ods select basicintervals testsforlocation testsfornormality;
run;
```

The UNIVARIATE Procedure
Variable: diff

Basic Confidence Limits Assuming Normality			
Parameter	**Estimate**	**95% Confidence Limits**	
Mean	12.60000	-15.53129	40.73129
Std Deviation	22.65612	13.57404	65.10364
Variance	513.30000	184.25443	4238

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1.243569	**Pr > $\lvert t \rvert$**	0.2816
Sign	**M**	1.5	**Pr >= $\lvert M \rvert$**	0.3750
Signed Rank	**S**	4.5	**Pr >= $\lvert S \rvert$**	0.3125

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.913865	**Pr < W**	0.4912
Kolmogorov-Smirnov	**D**	0.236875	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.047823	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.302348	**Pr > A-Sq**	>0.2500

Inferential Statistics – Two Independent Samples – Quantitative Variable

- Comparing means of two independent samples from normal populations using PROC TTEST

```
ods graphics on;
proc ttest data=bio.whas500;
class gender;
var hr sysbp bmi;
run;
ods graphics off;
```

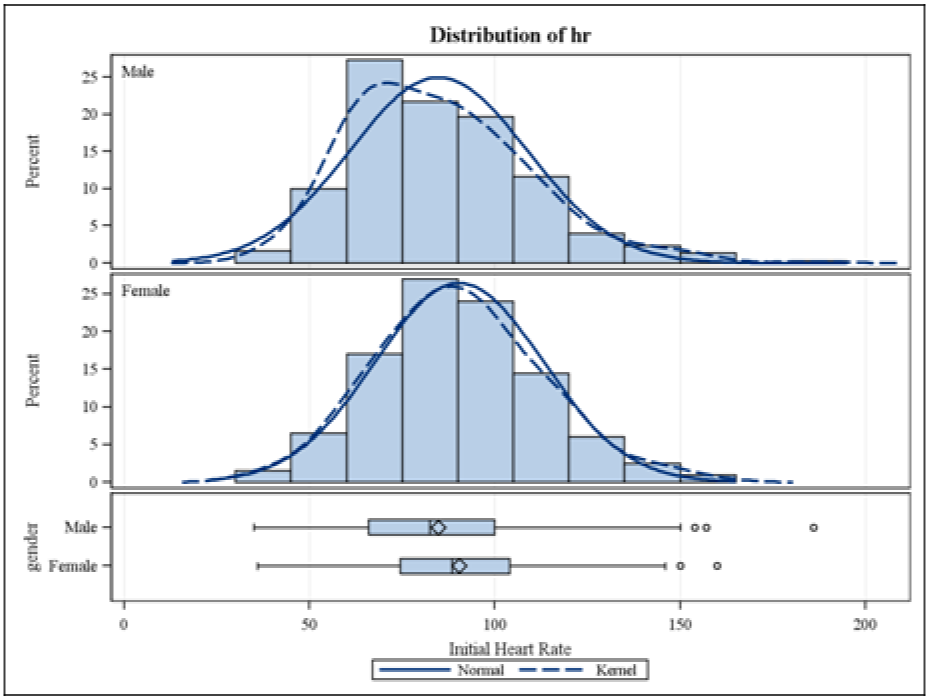
Variable: hr (Initial Heart Rate)

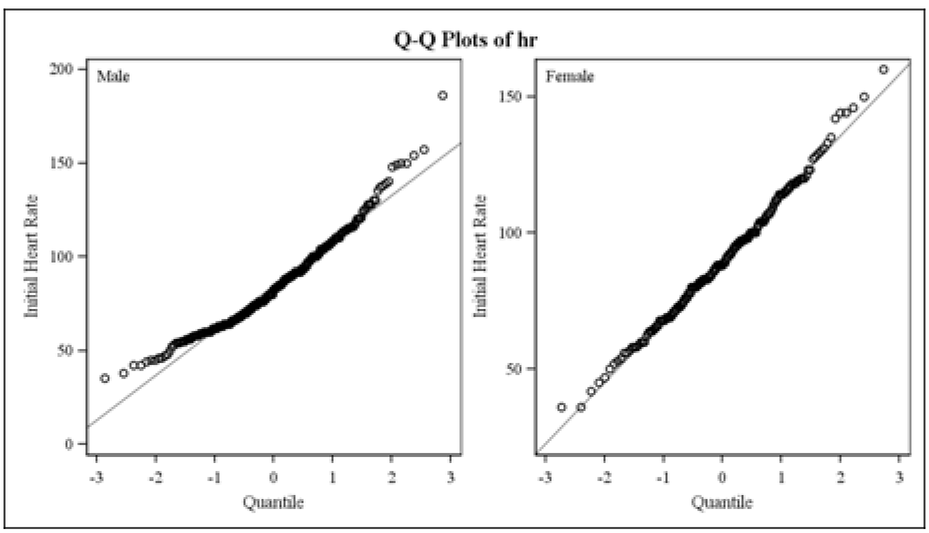
gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
Male	300	84.7867	23.9829	1.3847	35.0000	186.0
Female	200	90.3650	22.6272	1.6000	36.0000	160.0
Diff (1-2)		-5.5783	23.4506	2.1407		

gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Male		84.7867	82.0618	87.5116	23.9829	22.2051	26.0725
Female		90.3650	87.2099	93.5201	22.6272	20.6058	25.0919
Diff (1-2)	Pooled	-5.5783	-9.7843	-1.3723	23.4506	22.0803	25.0035
Diff (1-2)	Satterthwaite	-5.5783	-9.7369	-1.4198			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	498	-2.61	0.0094
Satterthwaite	Unequal	443.23	-2.64	0.0087

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	299	199	1.12	0.3755


Distribution of hr
Male
Female
Percent
gender
Initial Heart Rate
Normal
Kernel


Q-Q Plots of hr
Male
Female
Initial Heart Rate
Quantile

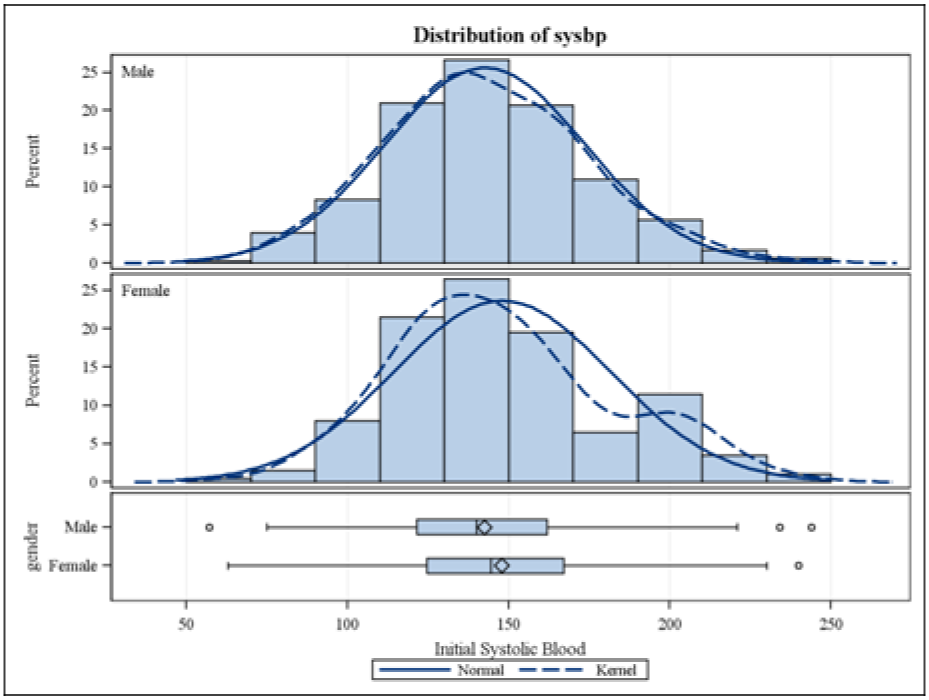
Variable: sysbp (Initial Systolic Blood)

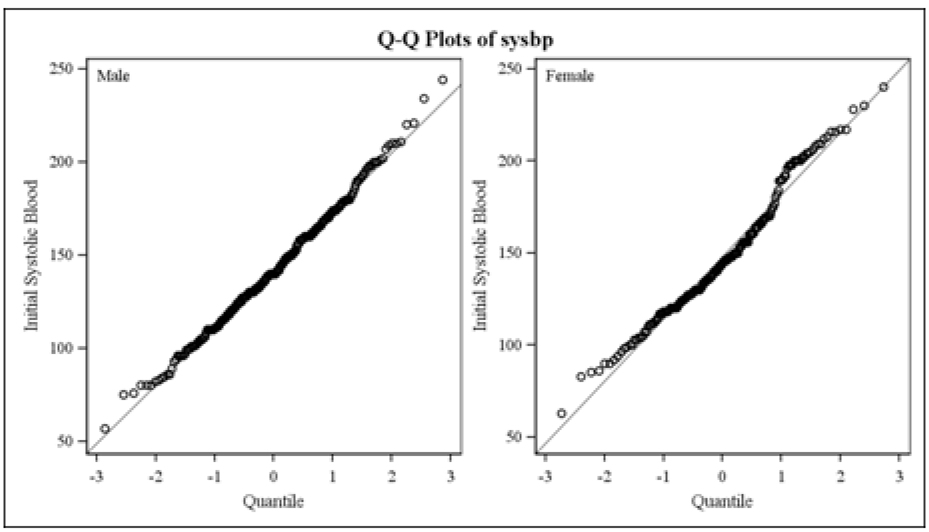
gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
Male	300	142.6	31.1602	1.7990	57.0000	244.0
Female	200	147.9	33.7491	2.3864	63.0000	240.0
Diff (1-2)		-5.3683	32.2197	2.9412		

gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Male		142.6	139.0	146.1	31.1602	28.8503	33.8752
Female		147.9	143.2	152.6	33.7491	30.7341	37.4253
Diff (1-2)	Pooled	-5.3683	-11.1471	0.4104	32.2197	30.3370	34.3534
Diff (1-2)	Satterthwaite	-5.3683	-11.2435	0.5068			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	498	-1.83	0.0686
Satterthwaite	Unequal	402.86	-1.80	0.0732

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	199	299	1.17	0.2123


Distribution of sysbp
Male
Female
Percent
gender
Initial Systolic Blood
Normal
Kernel


Q-Q Plots of sysbp
Male
Female
Initial Systolic Blood
Quantile

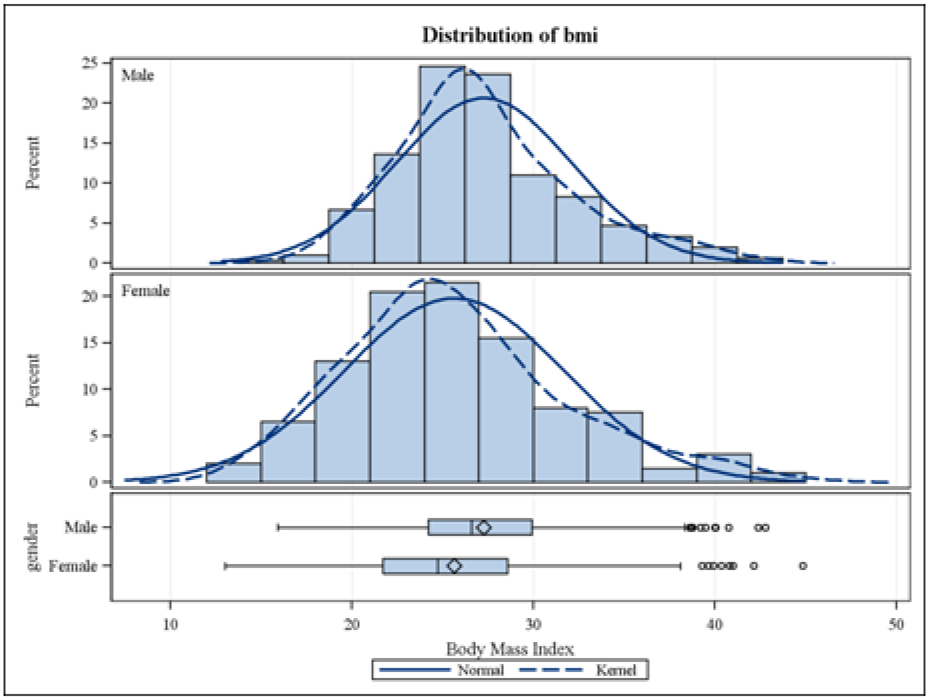
Variable: bmi (Body Mass Index)

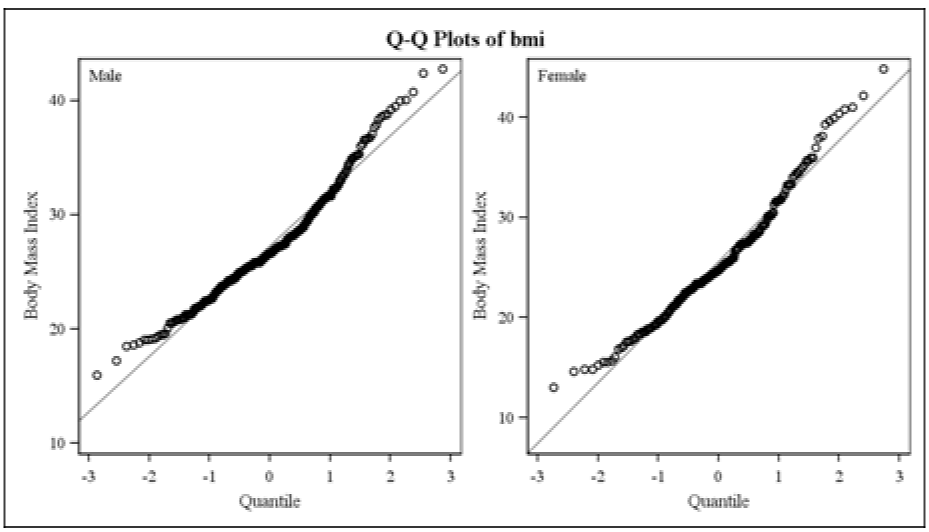
gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
Male	300	27.2689	4.8284	0.2788	15.9270	42.7659
Female	200	25.6311	6.0520	0.4279	13.0455	44.8389
Diff (1-2)		1.6378	5.3510	0.4885		

gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Male		27.2689	26.7203	27.8175	4.8284	4.4705	5.2491
Female		25.6311	24.7872	26.4750	6.0520	5.5114	6.7113
Diff (1-2)	Pooled	1.6378	0.6781	2.5975	5.3510	5.0384	5.7054
Diff (1-2)	Satterthwaite	1.6378	0.6334	2.6422			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	498	3.35	0.0009
Satterthwaite	Unequal	360.51	3.21	0.0015

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	199	299	1.57	0.0004





- Comparing two distributions with the Wilcoxon Rank Sum test using PROC NPAR1WAY

```
proc npar1way data=bio.whas500 wilcoxon;
class gender;
var hr sysbp bmi;
run;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable hr Classified by Variable gender					
gender	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Male	300	70286.50	75150.0	1582.48016	234.288333
Female	200	54963.50	50100.0	1582.48016	274.817500
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	54963.5000
Normal Approximation	
Z	3.0730
One-Sided Pr > Z	0.0011
Two-Sided Pr > |Z|	0.0021
t Approximation	
One-Sided Pr > Z	0.0011
Two-Sided Pr > |Z|	0.0022
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	9.4454
DF	1
Pr > Chi-Square	0.0021

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable sysbp Classified by Variable gender					
gender	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Male	300	73063.0	75150.0	1582.53600	243.543333
Female	200	52187.0	50100.0	1582.53600	260.935000
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	52187.0000
Normal Approximation	
Z	1.3185
One-Sided Pr > Z	0.0937
Two-Sided Pr > |Z|	0.1874
t Approximation	
One-Sided Pr > Z	0.0940
Two-Sided Pr > |Z|	0.1880
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	1.7392
DF	1
Pr > Chi-Square	0.1872

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable bmi Classified by Variable gender					
gender	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Male	300	81163.0	75150.0	1582.71329	270.543333
Female	200	44087.0	50100.0	1582.71329	220.435000
Average scores were used for ties.					

Wilcoxon Two-Sample Test	
Statistic	44087.0000
Normal Approximation	
Z	-3.7989
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	0.0001
t Approximation	
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	0.0002
Z includes a continuity correction of 0.5.	

Kruskal-Wallis Test	
Chi-Square	14.4337
DF	1
Pr > Chi-Square	0.0001

- Obtaining the exact p-value for the Wilcoxon Rank Sum test using PROC NPAR1WAY (only the 2nd table – of three – is provided below, the rest remains the same)

```
data smallsamp;
set bio.whas500;
where sho=1; /* keep only 22 obs. with cardiogenic shock */
run;

proc npar1way data=smallsamp wilcoxon;
class gender;
var hr;
exact wilcoxon;
run;
```

Wilcoxon Two-Sample Test	
Statistic (S)	114.0000
Normal Approximation	
Z	0.6680
One-Sided Pr > Z	0.2521
Two-Sided Pr > |Z|	0.5042
t Approximation	
One-Sided Pr > Z	0.2557
Two-Sided Pr > |Z|	0.5114
Exact Test	
One-Sided Pr >= S	0.2510
Two-Sided Pr >= |S - Mean|	0.5009
Z includes a continuity correction of 0.5.	

Inferential Statistics – Two Categorical Variables

- Chi-square and Fisher's exact tests for association or equality of two proportions (2x2 tables)

```
proc freq data=bio.whas500;
table gender*(cvd afb) / chisq riskdiff ;
run;
```

The FREQ Procedure

Table of gender by cvd

gender(Gender)	cvd(History of Cardiovascular)		
Frequency Percent Row Pct Col Pct	No	Yes	Total
Male	89 17.80 29.67 71.20	211 42.20 70.33 56.27	300 60.00
Female	36 7.20 18.00 28.80	164 32.80 82.00 43.73	200 40.00
Total	125 25.00	375 75.00	500 100.00

Statistics for Table of gender by cvd

Statistic	DF	Value	Prob
Chi-Square	1	8.7111	0.0032
Likelihood Ratio Chi-Square	1	8.9697	0.0027
Continuity Adj. Chi-Square	1	8.1000	0.0044
Mantel-Haenszel Chi-Square	1	8.6937	0.0032
Phi Coefficient		0.1320	
Contingency Coefficient		0.1309	
Cramer's V		0.1320	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	89
Left-sided Pr <= F	0.9990
Right-sided Pr >= F	0.0020
Table Probability (P)	0.0010
Two-sided Pr <= P	0.0032

Column 1 Risk Estimates

	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.2967	0.0264	0.2450	0.3484	0.2455	0.3519
Row 2	0.1800	0.0272	0.1268	0.2332	0.1294	0.2404
Total	0.2500	0.0194	0.2120	0.2880	0.2126	0.2904
Difference	0.1167	0.0379	0.0425	0.1909		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates

	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.7033	0.0264	0.6516	0.7550	0.6481	0.7545
Row 2	0.8200	0.0272	0.7668	0.8732	0.7596	0.8706
Total	0.7500	0.0194	0.7120	0.7880	0.7096	0.7874
Difference	-0.1167	0.0379	-0.1909	-0.0425		
Difference is (Row 1 - Row 2)						

Sample Size = 500

The FREQ Procedure

Table of gender by afb			
gender(Gender)	afb(Atrial Fibrillation)		
Frequency Percent Row Pct Col Pct	No	Yes	Total
Male	257 51.40 85.67 60.90	43 8.60 14.33 55.13	300 60.00
Female	165 33.00 82.50 39.10	35 7.00 17.50 44.87	200 40.00
Total	422 84.40	78 15.60	500 100.00

Statistics for Table of gender by afb

Statistic	DF	Value	Prob
Chi-Square	1	0.9139	0.3391
Likelihood Ratio Chi-Square	1	0.9053	0.3414
Continuity Adj. Chi-Square	1	0.6893	0.4064
Mantel-Haenszel Chi-Square	1	0.9121	0.3396
Phi Coefficient		0.0428	
Contingency Coefficient		0.0427	
Cramer's V		0.0428	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	257
Left-sided Pr <= F	0.8602
Right-sided Pr >= F	0.2027
Table Probability (P)	0.0629
Two-sided Pr <= P	0.3789

Column 1 Risk Estimates						
	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.8567	0.0202	0.8170	0.8963	0.8118	0.8943
Row 2	0.8250	0.0269	0.7723	0.8777	0.7651	0.8750
Total	0.8440	0.0162	0.8122	0.8758	0.8092	0.8747
Difference	0.0317	0.0336	-0.0343	0.0976		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.1433	0.0202	0.1037	0.1830	0.1057	0.1882
Row 2	0.1750	0.0269	0.1223	0.2277	0.1250	0.2349
Total	0.1560	0.0162	0.1242	0.1878	0.1253	0.1908
Difference	-0.0317	0.0336	-0.0976	0.0343		
Difference is (Row 1 - Row 2)						

Sample Size = 500

- Chi-square and Fisher's exact tests for association (RxC tables). Fisher's exact test must be requested in this case.

```
proc freq data=bio.whas500;
table gender*year / chisq fisher ;
run;
```

The FREQ Procedure

Table of gender by year				
gender(Gender)	year(Cohort Year)			
Frequency Percent Row Pct Col Pct	1997	1999	2001	Total
Male	97 19.40 32.33 60.63	111 22.20 37.00 59.04	92 18.40 30.67 60.53	300 60.00
Female	63 12.60 31.50 39.38	77 15.40 38.50 40.96	60 12.00 30.00 39.47	200 40.00
Total	160 32.00	188 37.60	152 30.40	500 100.00

Statistics for Table of gender by year

Statistic	DF	Value	Prob
Chi-Square	2	0.1154	0.9439
Likelihood Ratio Chi-Square	2	0.1153	0.9440
Mantel-Haenszel Chi-Square	1	0.0005	0.9816
Phi Coefficient		0.0152	
Contingency Coefficient		0.0152	
Cramer's V		0.0152	

Fisher's Exact Test	
Table Probability (P)	0.0065
Pr <= P	0.9527

Sample Size = 500

- McNemar's tests for paired proportions. Consider a test for high blood pressure (Y/N) at baseline and again at 6 months.

```
data mcnemar;
input ID HBP0 HBP1 @@;
label       HBP0 = "High BP at Baseline"
            HBP1 = "High BP at 6 Mo.";
datalines;
1   0 0     2   1 1     3   1 0     4   1 1     5   1 1     6   0 0
7   0 0     8   0 0     9   0 0     10  1 1     11  0 0     12  1 1
13  1 1     14  1 1     15  0 0     16  1 1     17  0 0     18  1 1
19  0 1     20  0 0     21  0 1     22  1 1     23  1 0     24  0 0
25  0 0     26  1 0     27  0 0     28  0 0     29  0 0     30  1 0
31  0 1     32  0 0     33  1 0     34  1 0     35  1 1     36  0 0
37  1 1     38  0 0     39  0 0     40  0 0     41  1 1     42  0 0
43  0 1     44  1 1     45  0 0     46  0 0     47  0 0     48  1 1
49  0 0     50  1 0     51  0 0     52  0 0     53  0 0     54  1 1
55  1 0     56  1 1     57  0 0     58  1 1     59  0 0     60  0 1
61  0 0     62  1 1     63  1 0     64  0 0     65  1 0     66  0 0
67  1 0     68  0 0     69  1 0     70  1 1     71  1 0     72  0 0
73  0 0     74  0 1     75  0 0     76  0 0     77  0 0     78  1 1
79  1 0     80  1 0     81  1 1     82  0 0     83  1 0     84  0 0
85  0 0     86  1 0     87  1 0     88  0 0     89  1 1     90  0 0
91  0 0     92  0 0     93  1 1     94  0 0     95  1 0     96  1 0
;
run;
```

```
proc freq data=mcnemar;
tables HBP0*HBP1 / agree;
exact mcnem;
run;
```

Table of HBP0 by HBP1			
HBP0 (High BP at Baseline)	HBP1 (High BP at 6 Mo.)		
Frequency Percent Row Pct Col Pct	0	1	Total
0	46 47.92 88.46 69.70	6 6.25 11.54 20.00	52 54.17
1	20 20.83 45.45 30.30	24 25.00 54.55 80.00	44 45.83
Total	66 68.75	30 31.25	96 100.00

Statistics for Table of HBP0 by HBP1

McNemar's Test	
Statistic (S)	7.5385
DF	1
Asymptotic Pr > S	0.0060
Exact Pr >= S	0.0094

Simple Kappa Coefficient	
Kappa	0.4409
ASE	0.0888
95% Lower Conf Limit	0.2668
95% Upper Conf Limit	0.6150

Sample Size = 96

Inferential Statistics – Two Quantitative Variables

- Tests and confidence intervals for the population correlation coefficient using PROC CORR. Using the VAR statement alone computes for all possible pairs. Using the VAR and WITH statements produces only combinations with one variable from each statement.

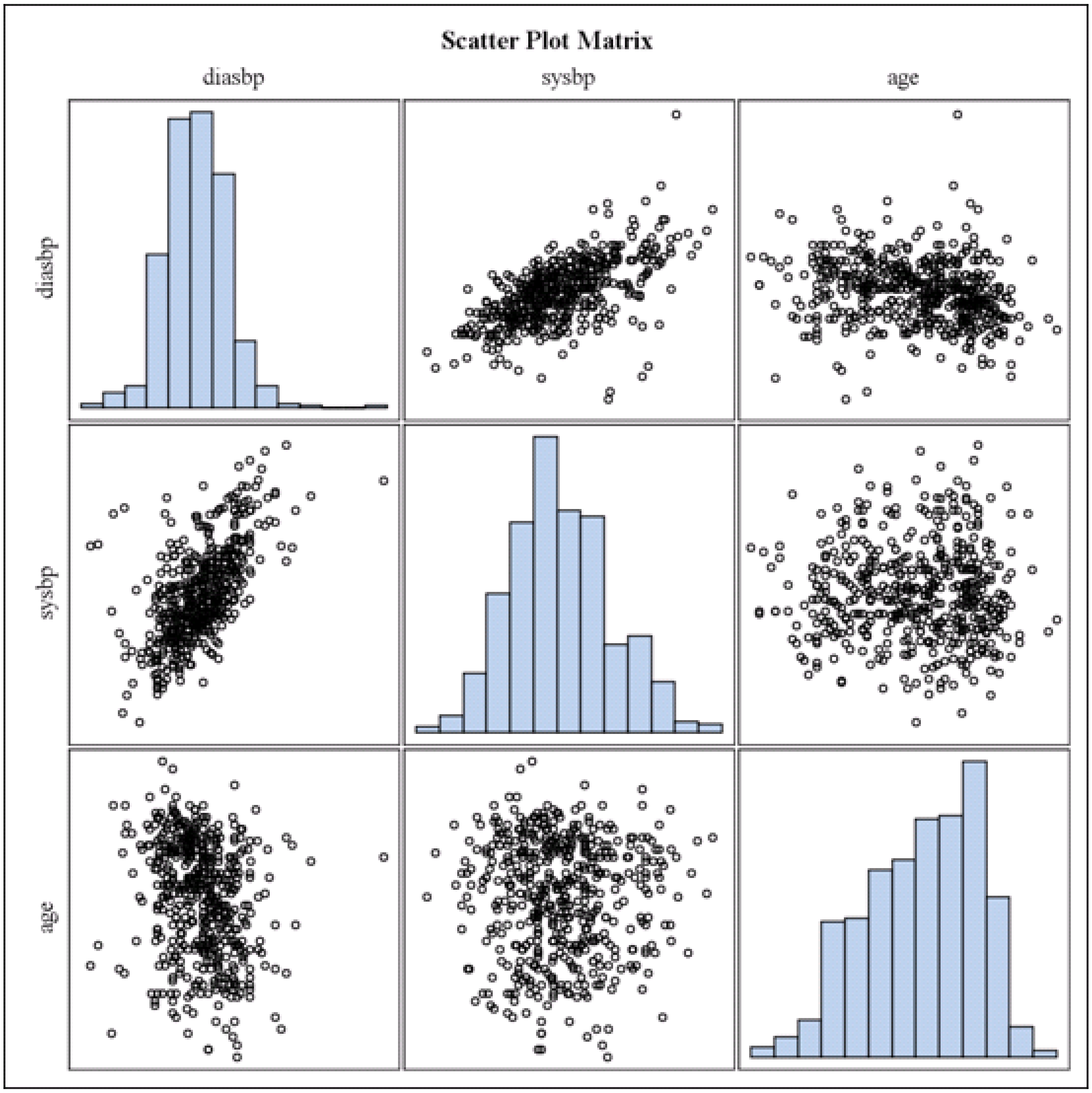
```
ods graphics on;
proc corr data=bio.whas500 plots=matrix(histogram) fisher;
var diasbp sysbp age ;
run;
ods graphics off;
```

3 Variables:	diasbp sysbp age

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
diasbp	500	78.26600	21.54529	39133	6.00000	198.00000	Initial Diastolic Blood Pressure
sysbp	500	144.70400	32.29486	72352	57.00000	244.00000	Initial Systolic Blood
age	500	69.84600	14.49146	34923	30.00000	104.00000	Age at Hospital Admission

Pearson Correlation Coefficients, N = 500 Prob > |r| under H0: Rho=0			
	diasbp	sysbp	age
diasbp Initial Diastolic Blood Pressure	1.00000	0.61092 <.0001	-0.20604 <.0001
sysbp Initial Systolic Blood	0.61092 <.0001	1.00000	-0.01560 0.7279
age Age at Hospital Admission	-0.20604 <.0001	-0.01560 0.7279	1.00000

Pearson Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
diasbp	sysbp	500	0.61092	0.71038	0.0006121	0.61053	0.552417	0.662741	<.0001
diasbp	age	500	-0.20604	-0.20904	-0.0002065	-0.20585	-0.288333	-0.120328	<.0001
sysbp	age	500	-0.01560	-0.01560	-0.0000156	-0.01558	-0.103133	0.072206	0.7280


Scatter Plot Matrix
diasbp
sysbp
age
diasbp
sysbp
age

Using VAR and WITH

```
proc corr data=bio.whas500 fisher;
var diasbp age ;
with sysbp ;
run;
```

1 With Variables:	sysbp
2 Variables:	diasbp age

Simple Statistics							
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
sysbp	500	144.70400	32.29486	72352	57.00000	244.00000	Initial Systolic Blood
diasbp	500	78.26600	21.54529	39133	6.00000	198.00000	Initial Diastolic Blood Pressure
age	500	69.84600	14.49146	34923	30.00000	104.00000	Age at Hospital Admission

Pearson Correlation Coefficients, N = 500 Prob > |r| under H0: Rho=0		
	diasbp	age
sysbp Initial Systolic Blood	0.61092 <.0001	-0.01560 0.7279

Pearson Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
diasbp	sysbp	500	0.61092	0.71038	0.0006121	0.61053	0.552417	0.662741	<.0001
age	sysbp	500	-0.01560	-0.01560	-0.0000156	-0.01558	-0.103133	0.072206	0.7280

Spearman Rank Correlation:

```
proc corr data=bio.whas500 spearman;
var diasbp sysbp  ;
run;
```

2 Variables:	diasbp sysbp

Simple Statistics							
Variable	N	Mean	Std Dev	Median	Minimum	Maximum	Label
diasbp	500	78.26600	21.54529	79.00000	6.00000	198.00000	Initial Diastolic Blood Pressure
sysbp	500	144.70400	32.29486	141.50000	57.00000	244.00000	Initial Systolic Blood

Spearman Correlation Coefficients, N = 500 Prob > |r| under H0: Rho=0		
	diasbp	sysbp
diasbp Initial Diastolic Blood Pressure	1.00000	0.61461 <.0001
sysbp Initial Systolic Blood	0.61461 <.0001	1.00000

- Simple linear regression using PROC REG

  Continuous Predictor

```
ods graphics on;
proc reg data=bio.whas500;
model  sysbp = diasbp;
run;
quit;
ods graphics off;
```

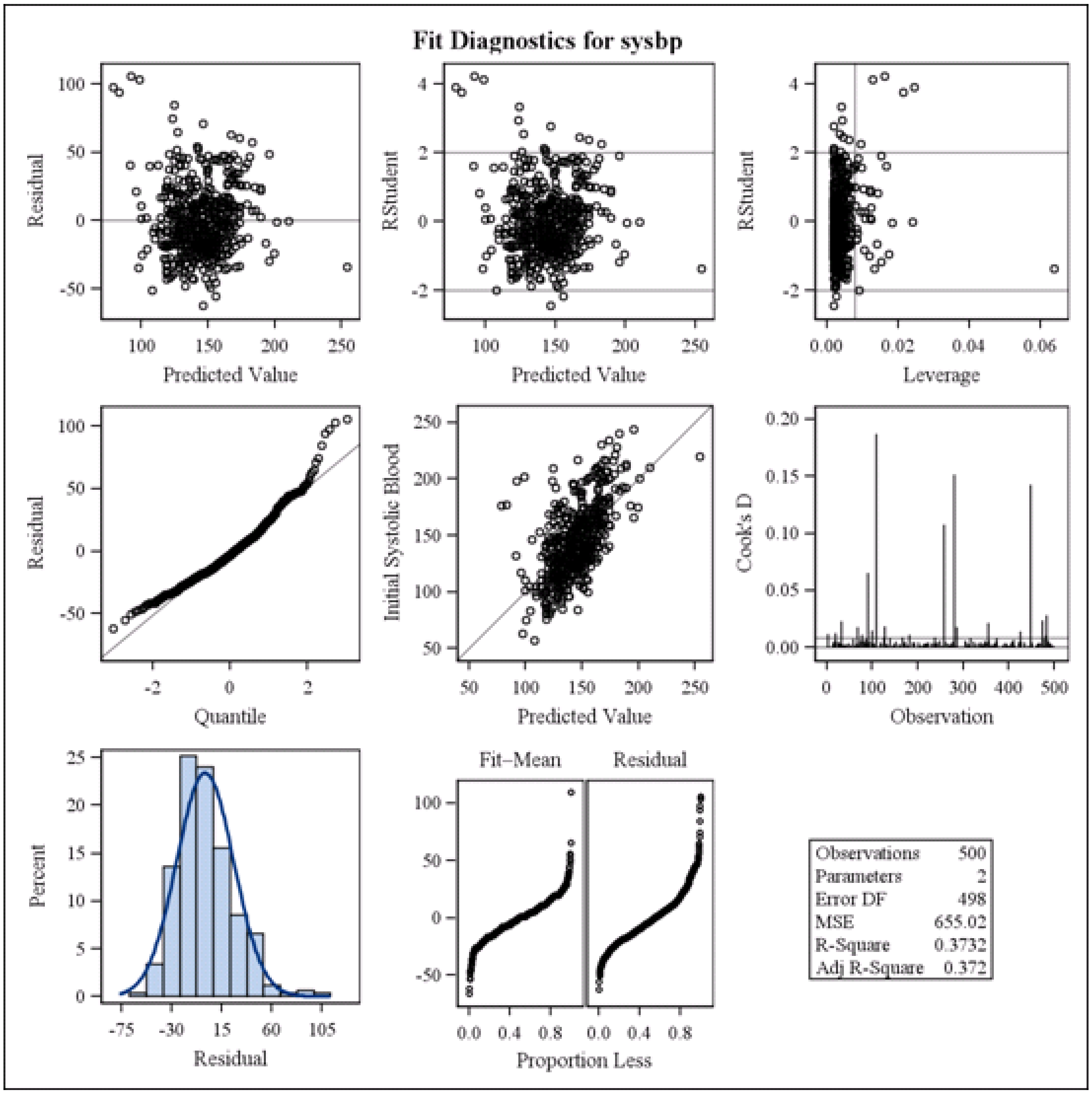
The REG Procedure
Model: MODEL1
Dependent Variable: sysbp Initial Systolic Blood

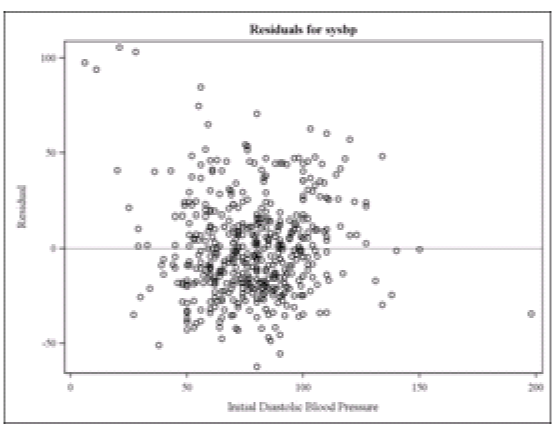
Number of Observations Read	500
Number of Observations Used	500

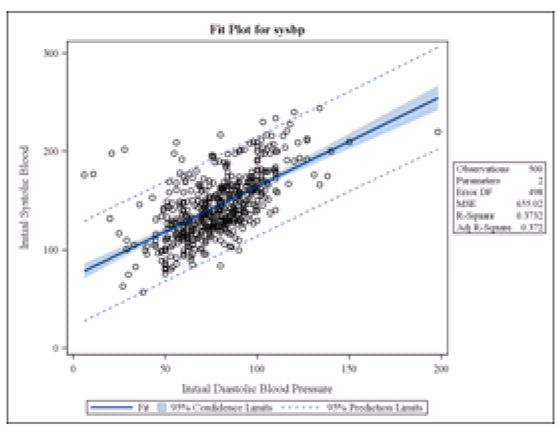
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	194236	194236	296.54	<.0001
Error	498	326200	655.01956		
Corrected Total	499	520436			

Root MSE	25.59335	**R-Square**	0.3732
Dependent Mean	144.70400	**Adj R-Sq**	0.3720
Coeff Var	17.68669		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	73.03426	4.31647	16.92	<.0001
diasbp	Initial Diastolic Blood Pressure	1	0.91572	0.05318	17.22	<.0001


Fit Diagnostics for sysbp
Residual
Predicted Value
RStudent
Leverage
Quantile
Initial Systolic Blood
Cook's D
Observation
Percent
Residual
Fit–Mean
Proportion Less
Observations 500
Parameters 2
Error DF 498
MSE 655.02
R-Square 0.3732
Adj R-Square 0.372


Residuals for sysbp
Residual
Initial Diastolic Blood Pressure


Fit Plot for sysbp
Initial Systolic Blood
Initial Diastolic Blood Pressure
Fit
95% Confidence Limits
95% Prediction Limits

Categorical Predictor

```
ods graphics on;
proc reg data=bio.whas500;
model  sysbp = cvd;
run;
quit;
ods graphics off;
```

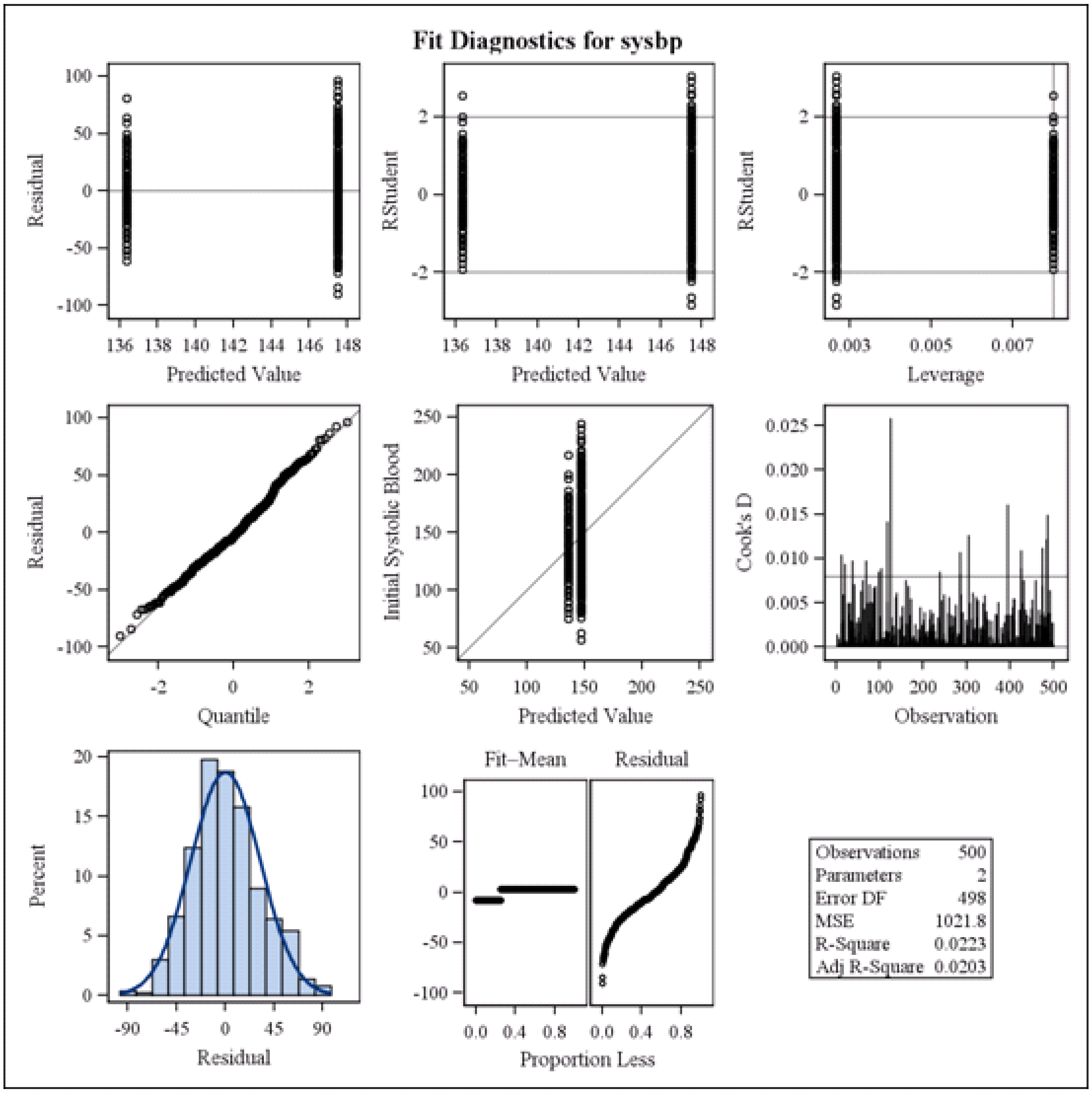
The REG Procedure
Model: MODEL1
Dependent Variable: sysbp Initial Systolic Blood

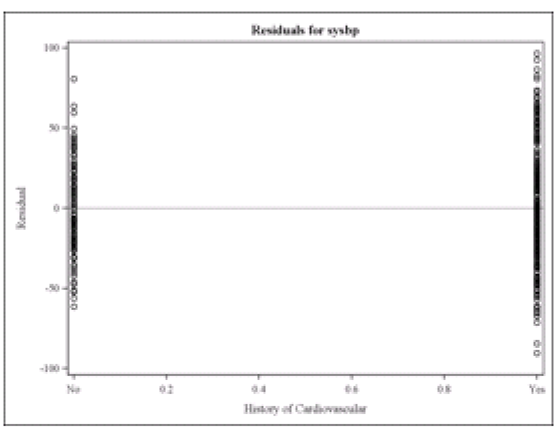
Number of Observations Read	500
Number of Observations Used	500

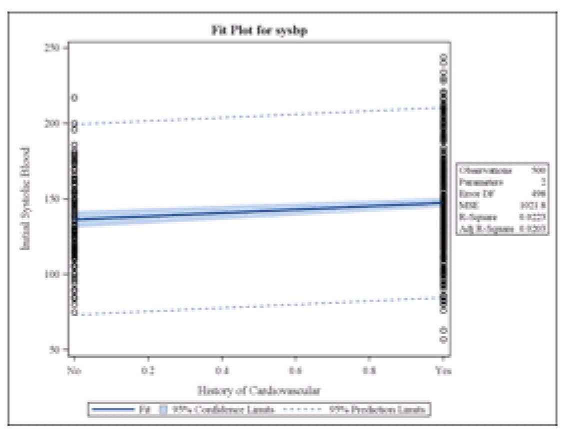
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	11604	11604	11.36	0.0008
Error	498	508832	1021.75195		
Corrected Total	499	520436			

Root MSE	31.96485	**R-Square**	0.0223
Dependent Mean	144.70400	**Adj R-Sq**	0.0203
Coeff Var	22.08982		

Parameter Estimates						
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	Intercept	1	136.36000	2.85902	47.69	<.0001
cvd	History of Cardiovascular	1	11.12533	3.30132	3.37	0.0008


Fit Diagnostics for sysbp
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Initial Systolic Blood
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 500
Parameters 2
Error DF 498
MSE 1021.8
R-Square 0.0223
Adj R-Square 0.0203


Residuals for sysbp
Residual
No
Yes
History of Cardiovascular


Fit Plot for sysbp
Initial Systolic Blood
No
Yes
History of Cardiovascular

Equivalent tests with continuous outcome and bivariate predictor

```
proc ttest data=bio.whas500;
class cvd;
var sysbp;
run;
```

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	498	-3.37	0.0008
Satterthwaite	Unequal	257.55	-3.73	0.0002

```
proc corr data=bio.whas500;
var cvd ;
with sysbp;
run;
```

Pearson Correlation Coefficients, N = 500 Prob > |r| under H0: Rho=0	
	cvd
sysbp Initial Systolic Blood	0.14932 0.0008

Inferential Statistics – Multiple Linear Regression

- Using PROC REG

```
ods graphics on;
proc reg data=bio.whas500;
model  sysbp = diasbp age gender cvd ;
run;
quit;
ods graphics off;
```

The REG Procedure
Model: MODEL1
Dependent Variable: sysbp Initial Systolic Blood

Number of Observations Read	500
Number of Observations Used	500

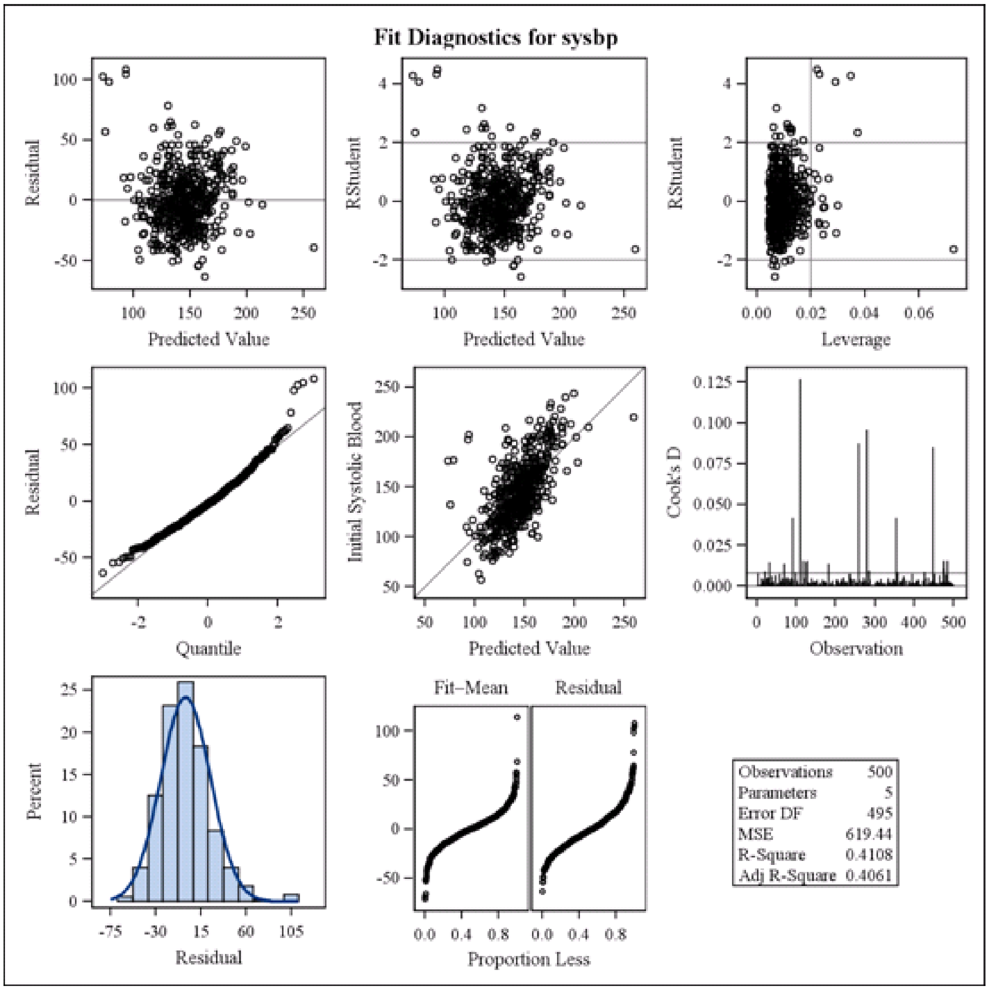
Analysis of Variance

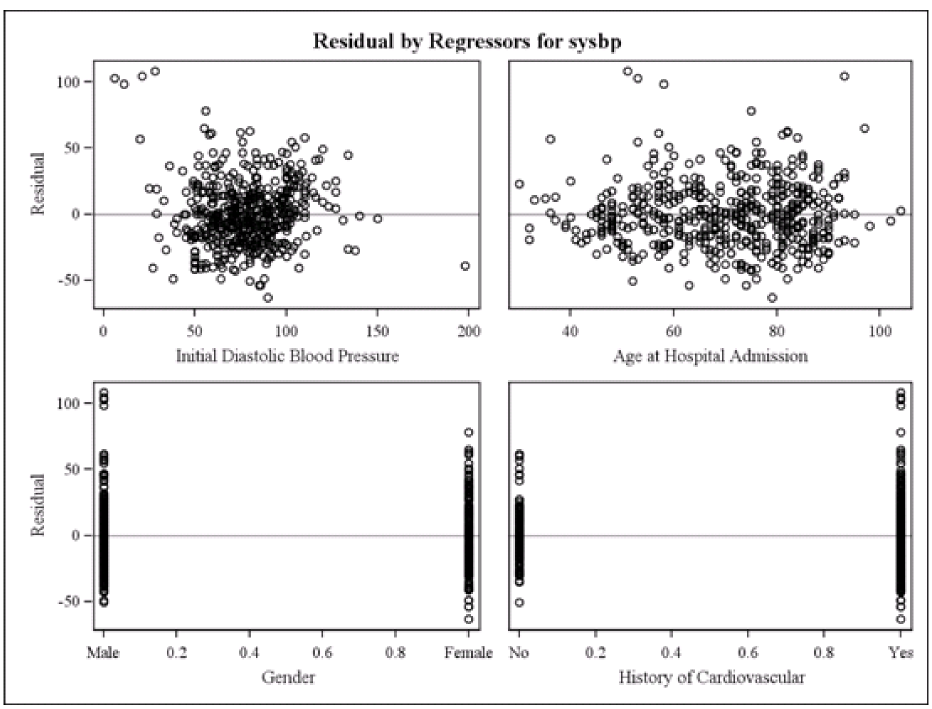
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	213813	53453	86.29	<.0001
Error	495	306623	619.44055		
Corrected Total	499	520436			

Root MSE	24.88856	**R-Square**	0.4108
Dependent Mean	144.70400	**Adj R-Sq**	0.4061
Coeff Var	17.19964		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	50.80299	7.66412	6.63	<.0001
diasbp	Initial Diastolic Blood Pressure	1	0.94773	0.05295	17.90	<.0001
age	Age at Hospital Admission	1	0.15127	0.08222	1.84	0.0664
gender	Gender	1	6.81240	2.37505	2.87	0.0043
cvd	History of Cardiovascular	1	8.58139	2.61881	3.28	0.0011


Fit Diagnostics for sysbp
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
Initial Systolic Blood
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 500
Parameters 5
Error DF 495
MSE 619.44
R-Square 0.4108
Adj R-Square 0.4061


Residual by Regressors for sysbp
Residual
Initial Diastolic Blood Pressure
Age at Hospital Admission
Residual
Male
Female
Gender
No
Yes
History of Cardiovascular

Inferential Statistics – Additional Topics

- Non-parametric confidence intervals for the median or other percentiles.

```
proc univariate data=bio.whas500 ciquantdf;
var hr ;
ods select quantiles;
run;
```

The UNIVARIATE Procedure
Variable: hr (Initial Heart Rate)

Quantiles (Definition 5)						
				Order Statistics		
Quantile	Estimate	95% Confidence Limits Distribution Free		LCL Rank	UCL Rank	Coverage
100% Max	186.0					
99%	150.0	146	186	491	500	96.23
95%	128.5	123	139	466	486	95.89
90%	117.0	114	121	437	464	95.63
75% Q3	100.5	99	105	357	395	95.01
50% Median	85.0	83	88	229	273	95.08
25% Q1	69.0	67	72	106	144	95.01
10%	59.0	57	61	37	64	95.63
5%	54.0	47	56	15	35	95.89
1%	42.0	35	45	1	10	96.23
0% Min	35.0					