Animesh Kansal

Jaahnavi Tiruthani

Medha Tiwary

Hanson Wu

Assignment 3

/\*Question 1a\*/

LIBNAME homework 'C:\Users\Hanson\Documents\Hanson\College\UTD Spring 2018\Predicive Analaysis using SAS\Datasets';

**DATA** Earthquake;

SET homework.earthquakes;

IF Year >= **2000**;

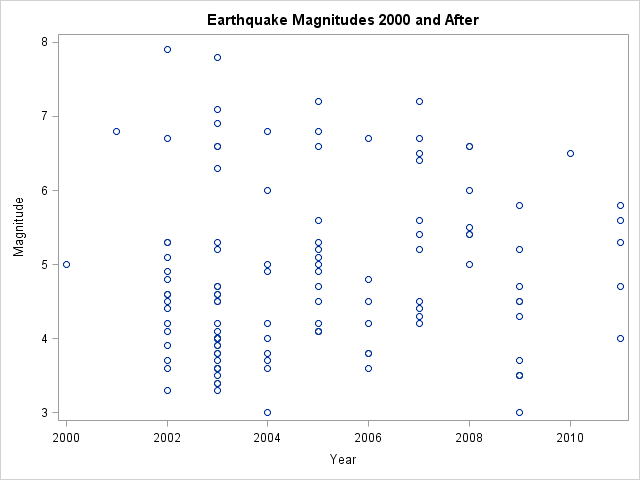
**PROC** **SGPLOT** DATA = Earthquake;

SCATTER X=Year Y=Magnitude;

TITLE 'Earthquake Magnitudes 2000 and After';

LABEL Year = 'Year' Magnitude = 'Magnitude';

**RUN**;



/\*Question 1b\*/

ODS GRAPHICS ON;

**PROC** **SORT** DATA= Earthquake;

BY Year;

**run**;

**PROC** **MEANS** DATA = Earthquake mean;

BY Year;

VAR Magnitude;

OUTPUT OUT=Means

mean = AvgMagnitude;

**run**;

**Data** Merged;

MERGE Earthquake Means;

BY Year;

**PROC** **PRINT** data=Merged;

**RUN**;

**PROC** **SGPLOT** DATA = Merged;

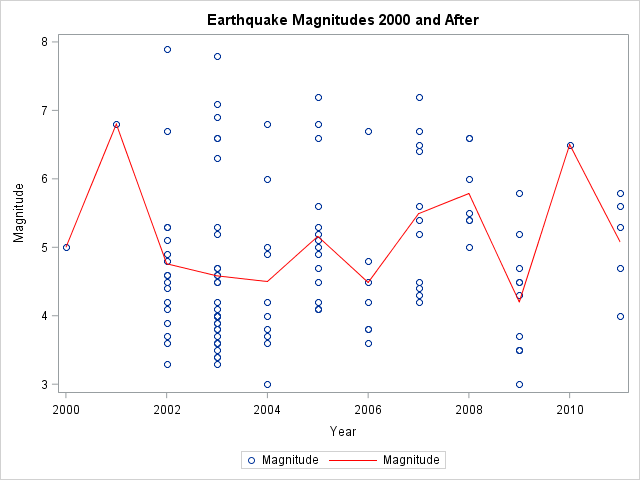
SCATTER X=Year Y=Magnitude;

TITLE 'Earthquake Magnitudes 2000 and After';

LABEL Year = 'Year' Magnitude = 'Magnitude';

SERIES X=Year Y=AvgMagnitude / LINEATTRS=(COLOR=red);

**RUN**;



/\*Question 1c\*/

**PROC** **SGPLOT** DATA = Merged;

SCATTER X=Year Y=Magnitude;

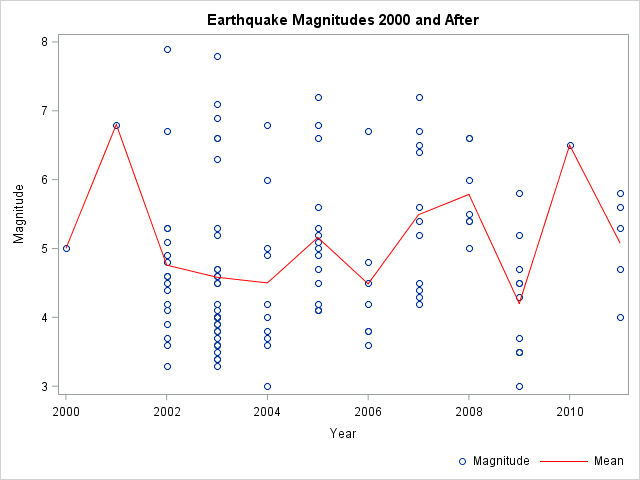
TITLE 'Earthquake Magnitudes 2000 and After';

LABEL Year = 'Year' Magnitude = 'Magnitude';

SERIES X=Year Y=AvgMagnitude / LINEATTRS=(COLOR=red) LEGENDLABEL="Mean";

KEYLEGEND / NOBORDER POSITION=BOTTOMRIGHT;

**RUN**;



/\*Question 1d\*/

**PROC** **SGPLOT** DATA = Merged;

SCATTER X=Year Y=Magnitude;

TITLE 'Earthquake Magnitudes 2000 and After';

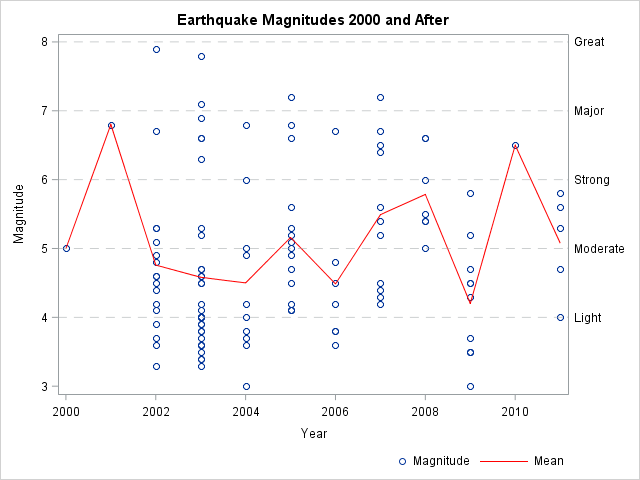
LABEL Year = 'Year' Magnitude = 'Magnitude';

SERIES X=Year Y=AvgMagnitude / lineattrs=(COLOR=red) LEGENDLABEL="Mean";

KEYLEGEND / NOBORDER POSITION=BOTTOMRIGHT;

REFLINE **4** **5** **6** **7** **8** / LABEL=('Light' 'Moderate' 'Strong' 'Major' 'Great') TRANSPARENCY=**0.5** LINEATTRS=(PATTERN=Dash);

**RUN**;



/\*Question 1e\*/

**PROC** **SGPLOT** DATA = Merged;

SCATTER X=Year Y=Magnitude;

XAXIS TYPE = discreet; /\*added this\*/

TITLE 'Earthquake Magnitudes 2000 and After';

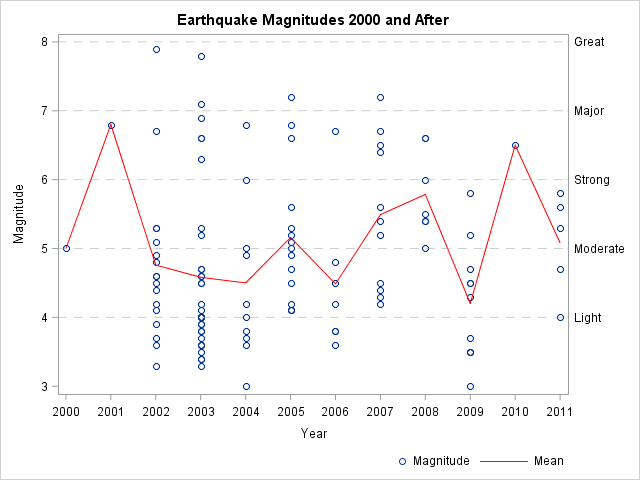
LABEL Year = 'Year' Magnitude = 'Magnitude';

SERIES X=Year Y=AvgMagnitude / lineattrs=(COLOR=red) LEGENDLABEL="Mean";

KEYLEGEND / NOBORDER POSITION=BOTTOMRIGHT;

REFLINE **4** **5** **6** **7** **8** / LABEL=('Light' 'Moderate' 'Strong' 'Major' 'Great') TRANSPARENCY=**0.5** LINEATTRS=(PATTERN=Dash);

**RUN**;



/\* Question 2\*/

/\* a) Reading the data \*/

libname Homework 'E:\Users\jxt163130\Documents\Homework';

**data** Study;

set Homework.study\_gpa;

**run**;

**proc** **sort** data=Study;

by Section;

**run**;

**proc** **print** data = Study;

**run**;

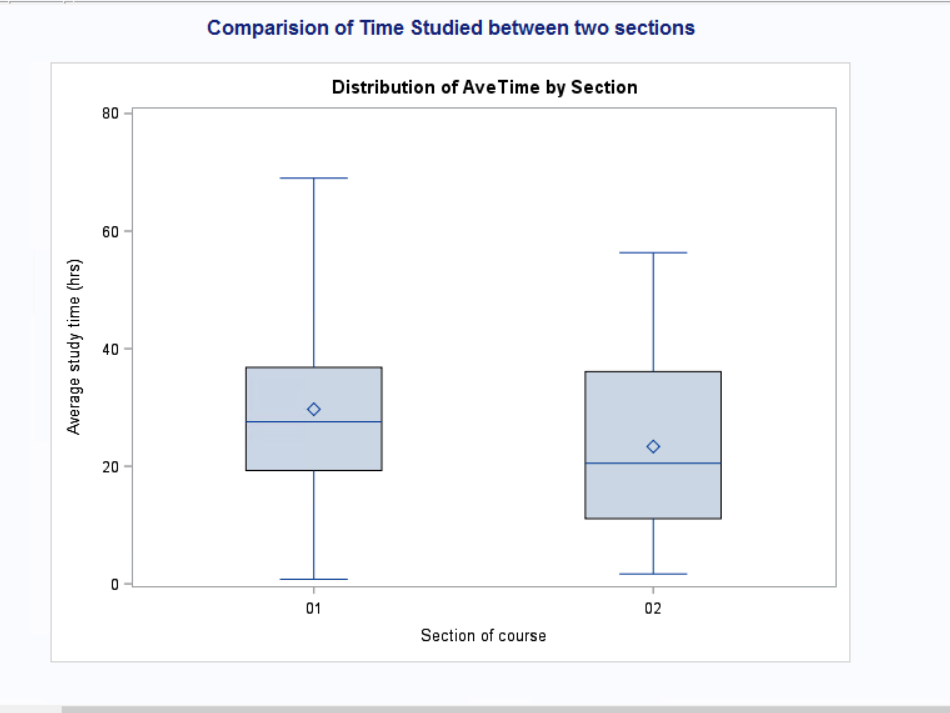
/\*Create box plots to compare the time studied between the two sections. \*/

**proc** **boxplot** data=Study;

plot AveTime\*Section;

title 'Comparison of Time Studied between two sections';

**run**;



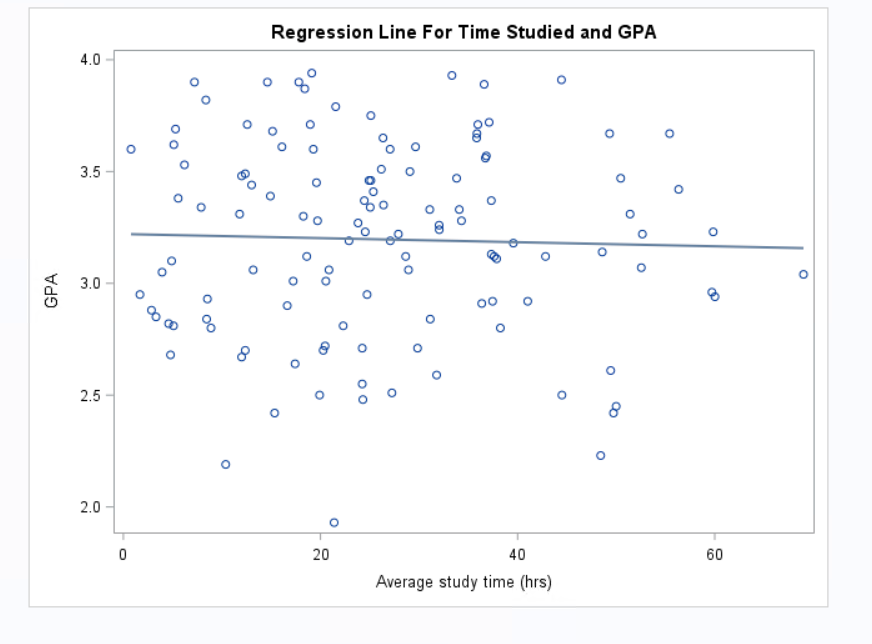
/\* b) Create a simple linear regression plot for time studied and GPA \*/

**proc** **sgplot** data=Study noautolegend;

title "Regression Line For Time Studied and GPA";

reg y=GPA x=AveTime;

**run**;



/\* c) Create a simple linear regression plot for time studied and GPA with a line for each section. Move the legend to the far right side of the plot. \*/

**proc** **sgpanel** data = Study ;

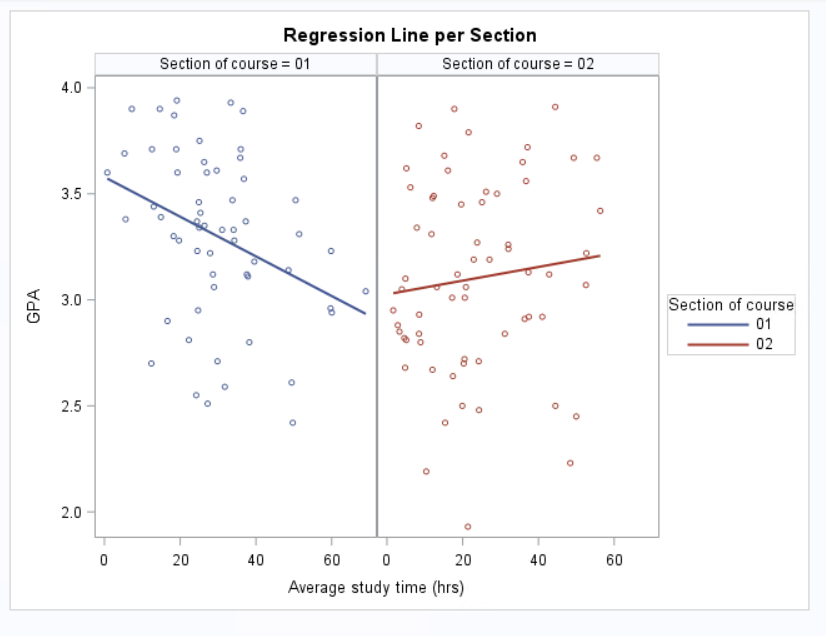
panelby Section;

reg X=AveTime Y=GPA / group= Section ;

title 'Regression Line per Section';

keylegend / BORDER position =RIGHT;

**run**;



/\* d)Add 95% confidence limits for the mean predicted values to your plot \*/

**proc** **sgpanel** data = Study ;

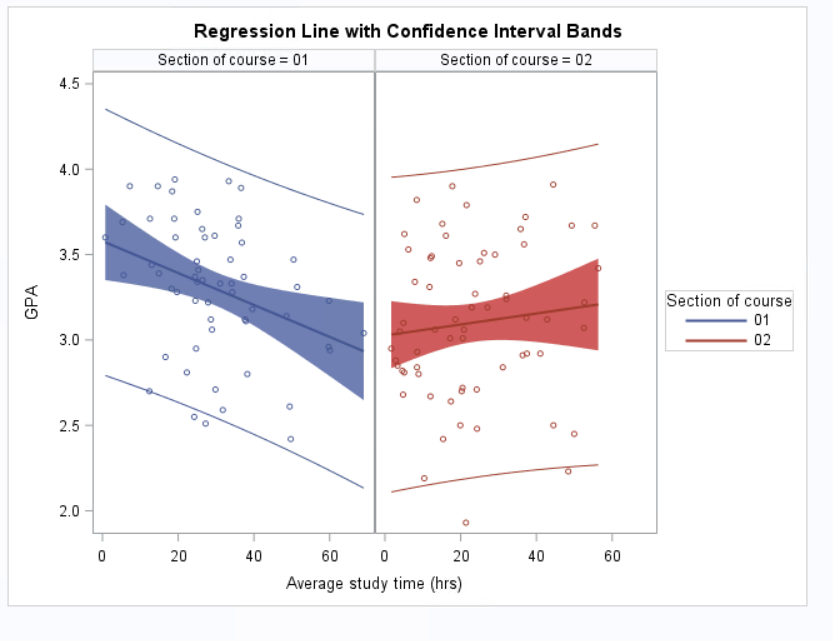
panelby Section;

reg X=AveTime Y=GPA / group= Section CLI CLM;

title 'Regression Line with Confidence Interval Bands';

keylegend / BORDER position =RIGHT;

**run**;



/\* e) comment about any potential relationships that you see between the variables included on these three plots.

In terms of GPA and study time across both sections there is a marginal inverse relation.

When the sections are considered independently, Section 01 is seen to have an acute inverse relationship between the GPA and study time. On the other hand, section 02 shows a direct relationship between the GPA and study time.

Majority of the data points fit within the 95% confidence interval. \*/

/\* Q 3.A.

For strata = 1

Ho = the true mean systolic blood pressure at baseline = 140 mm/HG

H1 =the true mean systolic blood pressure at baseline > 140 mm/HG

p value = 0.0219

reject null hypothesis \*/

**proc** **ttest** data = vite H0 = **140** SIDES = U ALPHA = **0.05**;

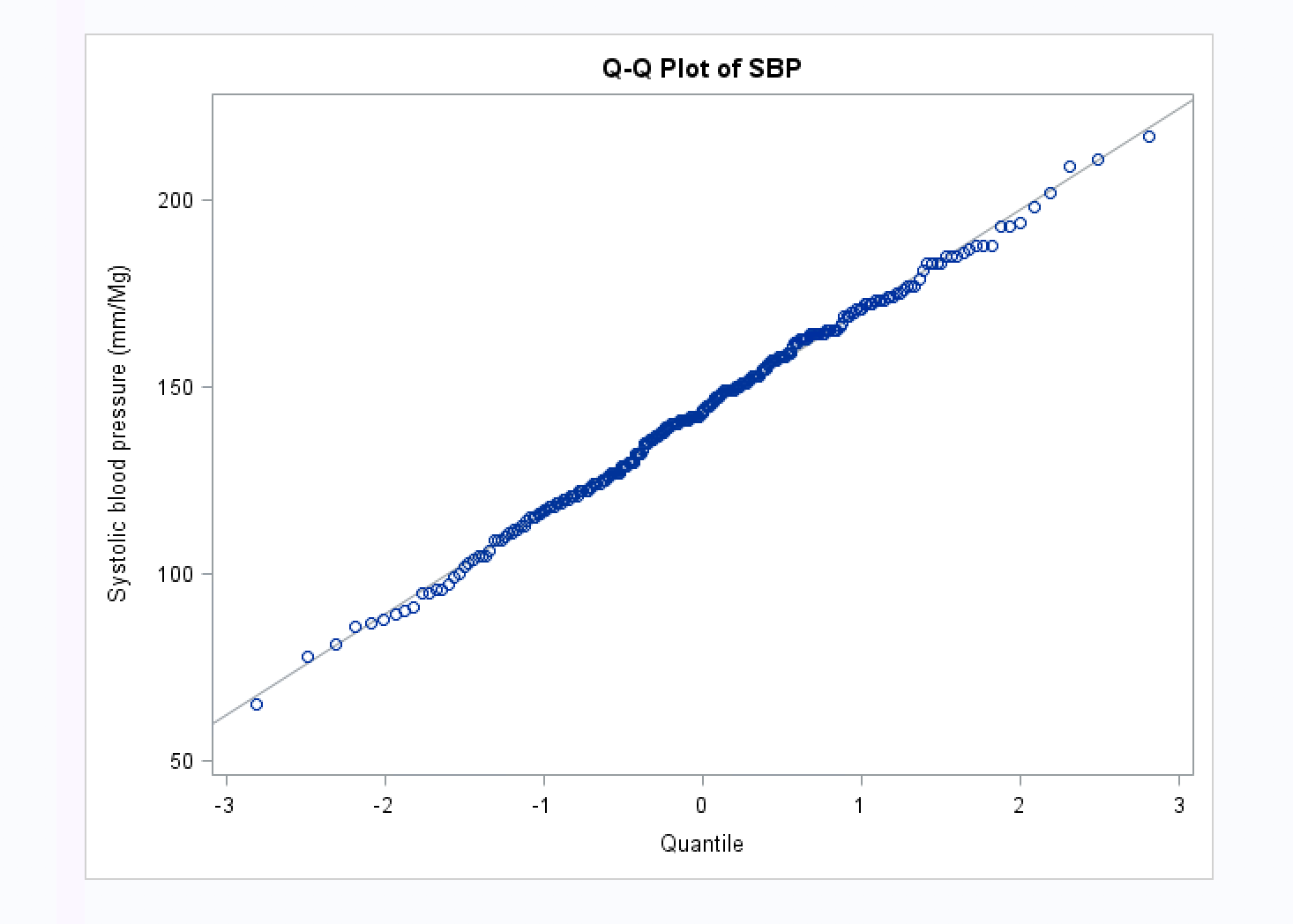
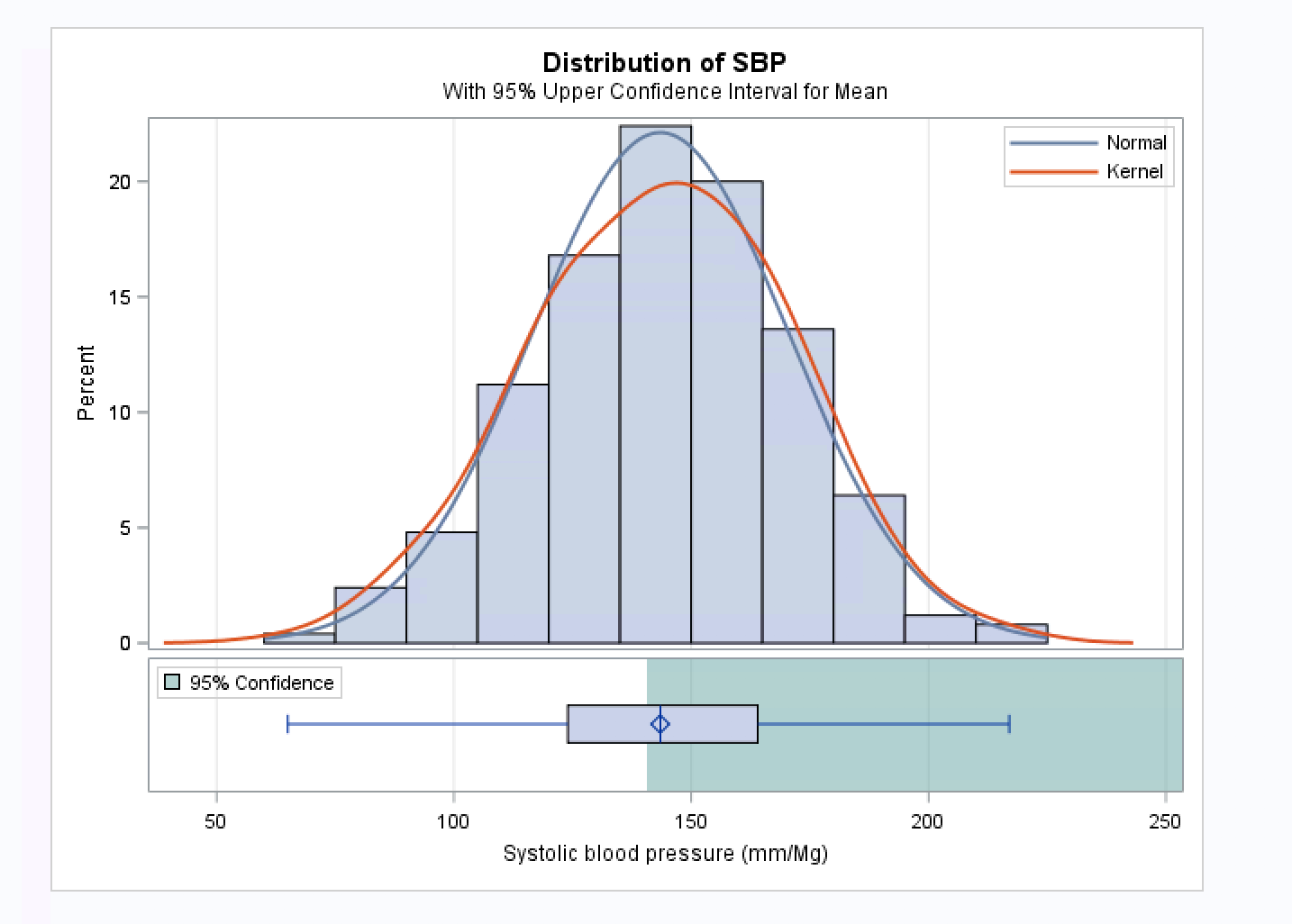
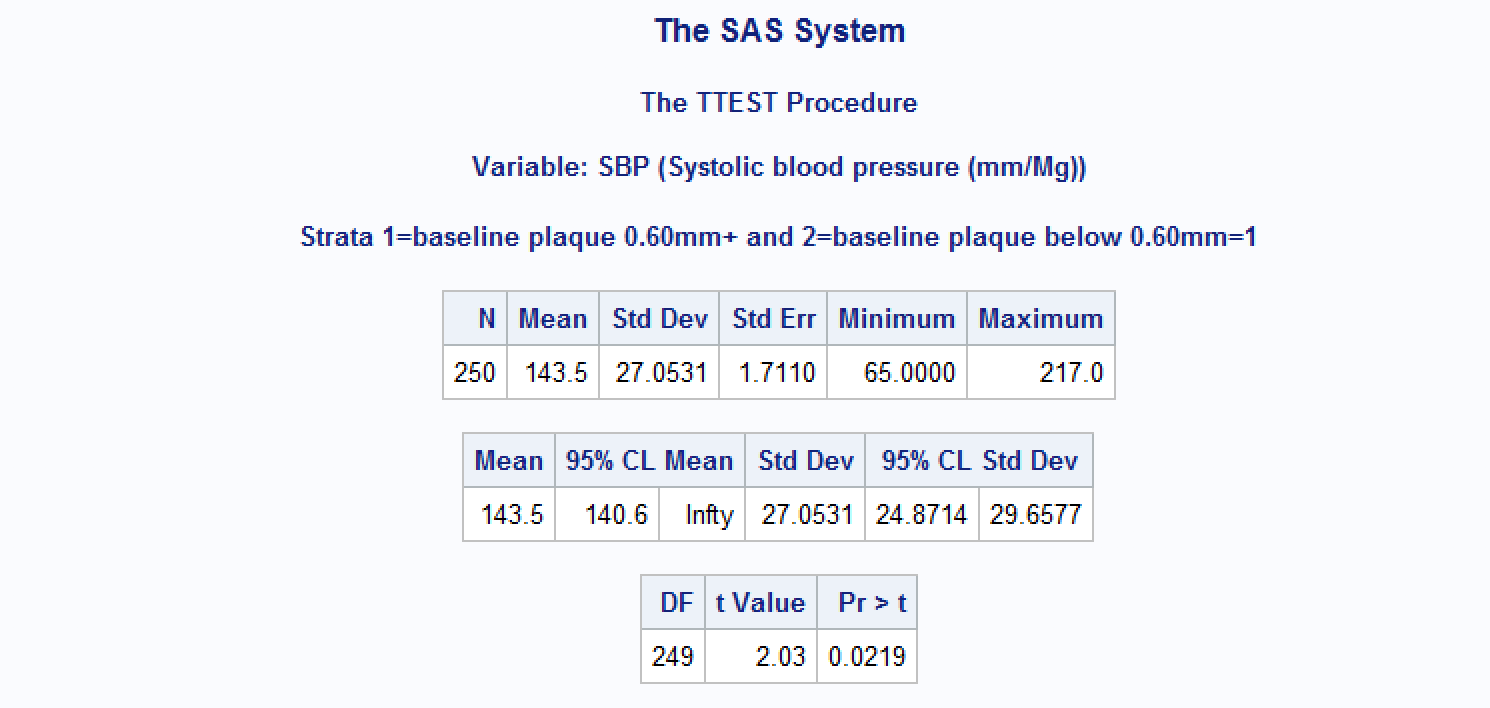
Var SBP ;

where visit = **0** ;

by Strata;

**run**;

/\*So, the true mean systolic blood pressure at baseline > 140 mm/HG for Strata = 1\*/



/\*For strata = 2

Ho = the true mean systolic blood pressure at baseline = 140 mm/HG

H1 =the true mean systolic blood pressure at baseline > 140 mm/HG

p value = 0.0350

reject null hypothesis\*/

**proc** **ttest** data = vite H0 = **140** SIDES = U ALPHA = **0.05**;

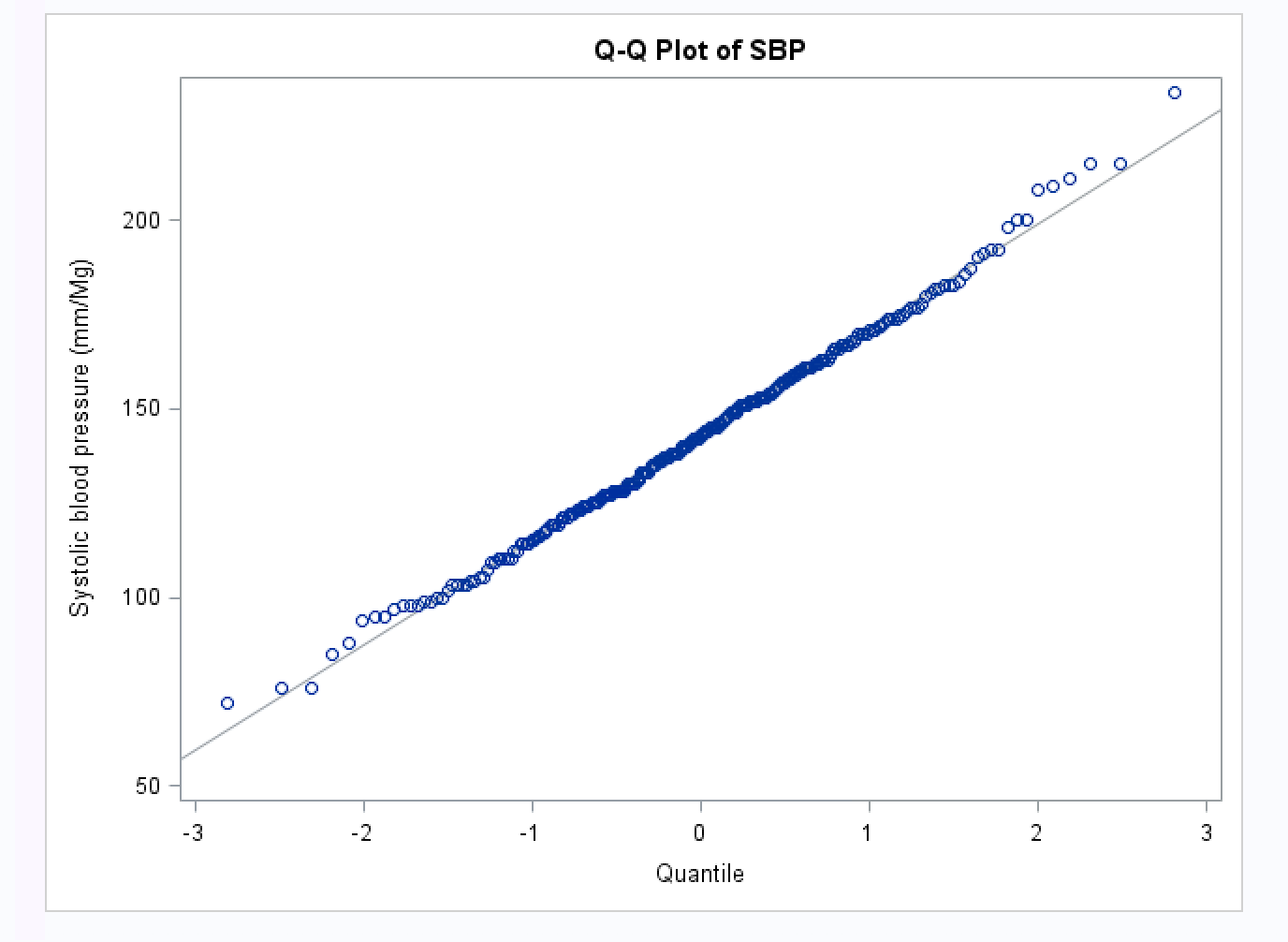
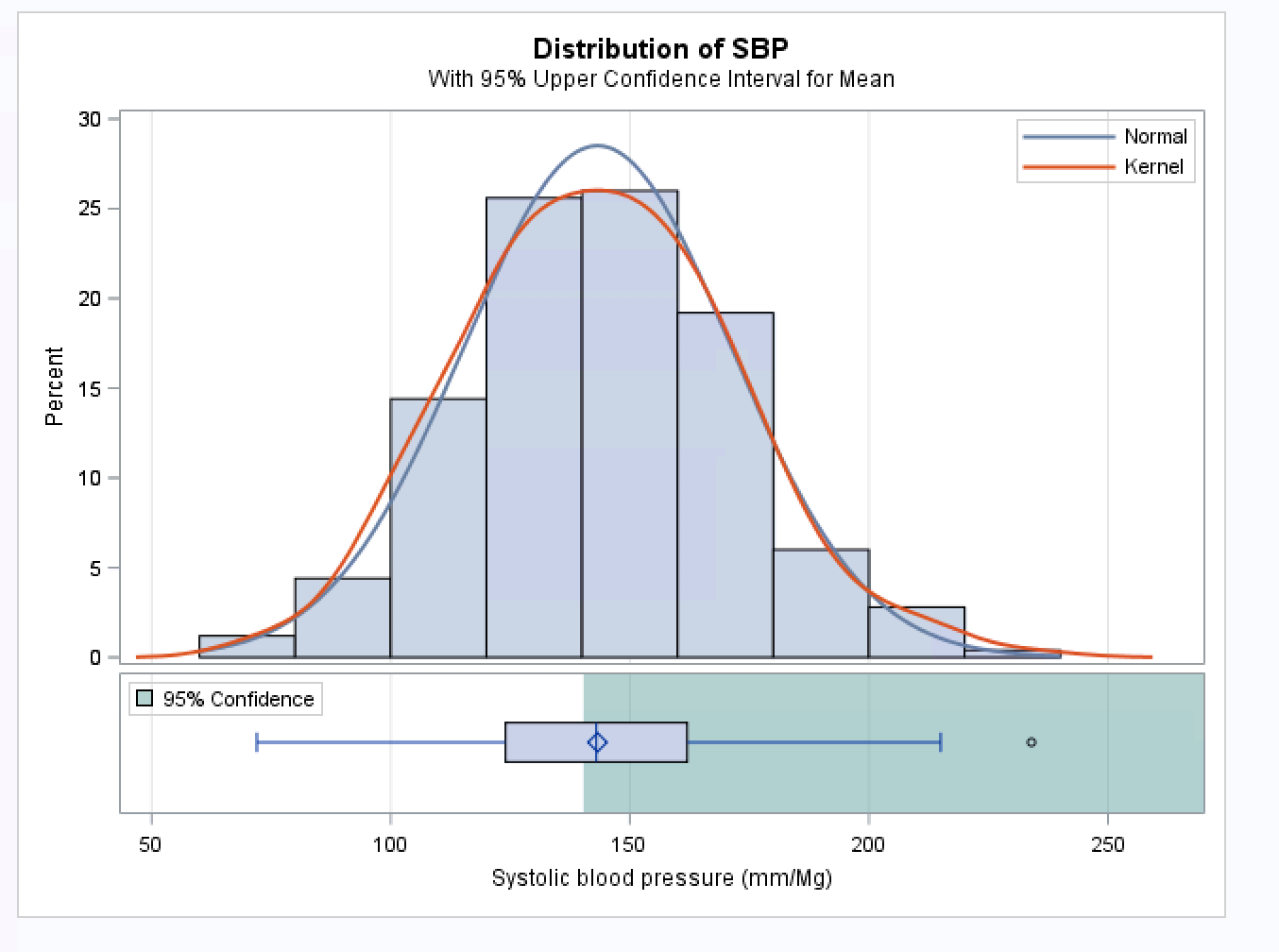
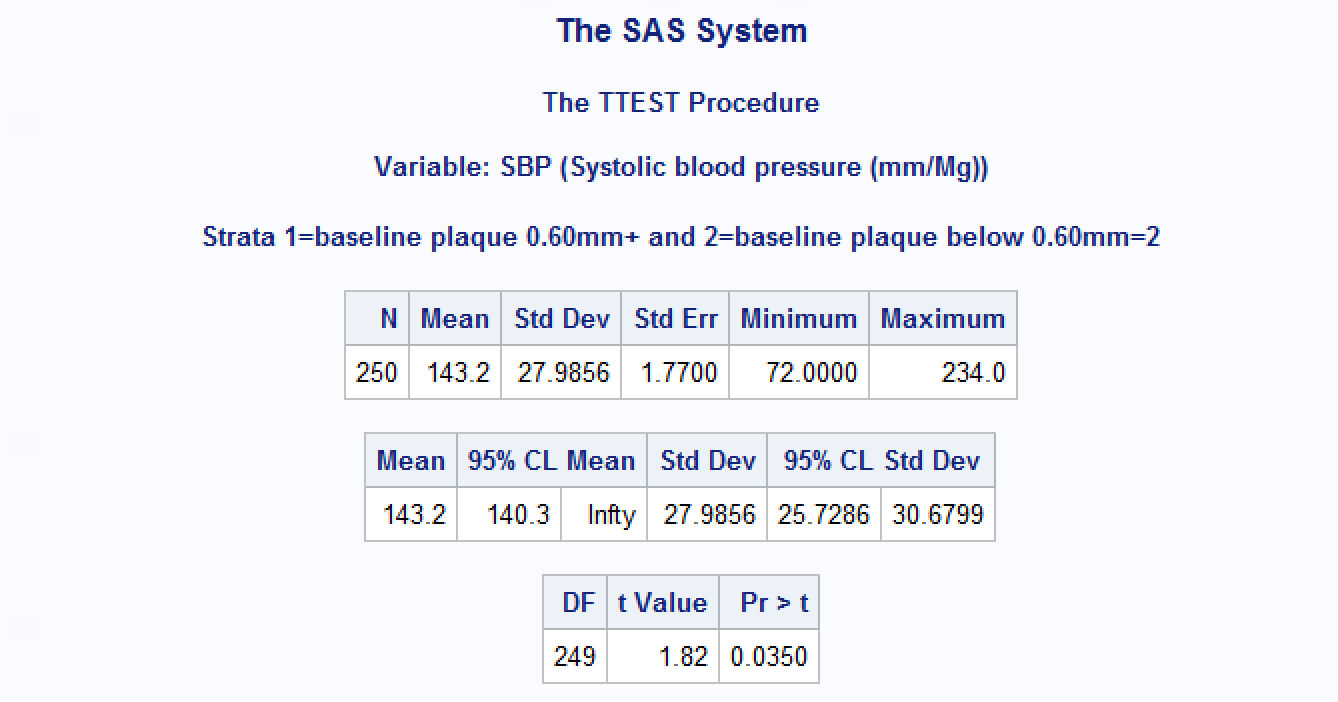
Var SBP ;

where visit = **0** ;

by Strata;

**run**;

/\*So, the true mean systolic blood pressure at baseline > 140 mm/HG for Strata = 2\*/

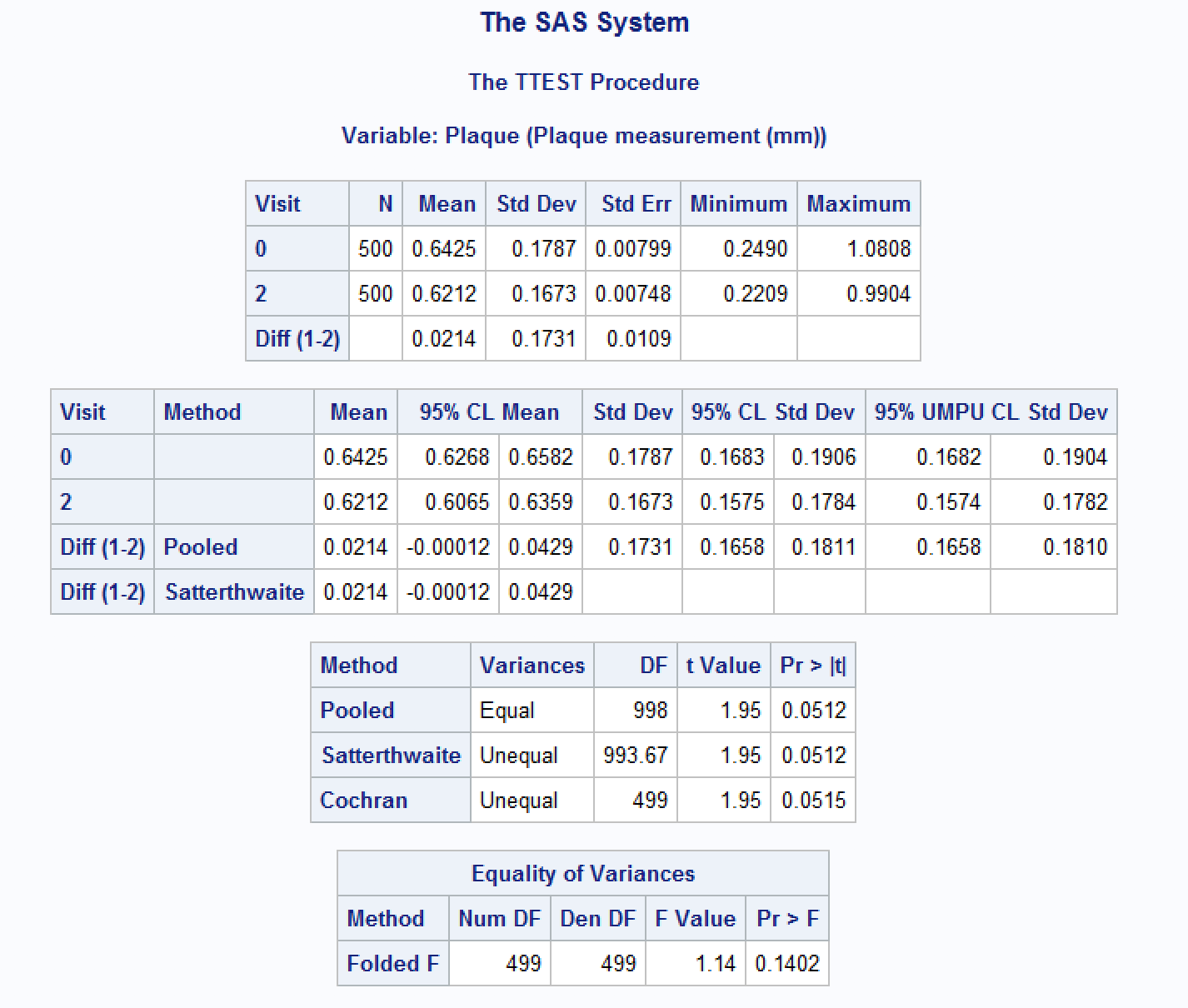


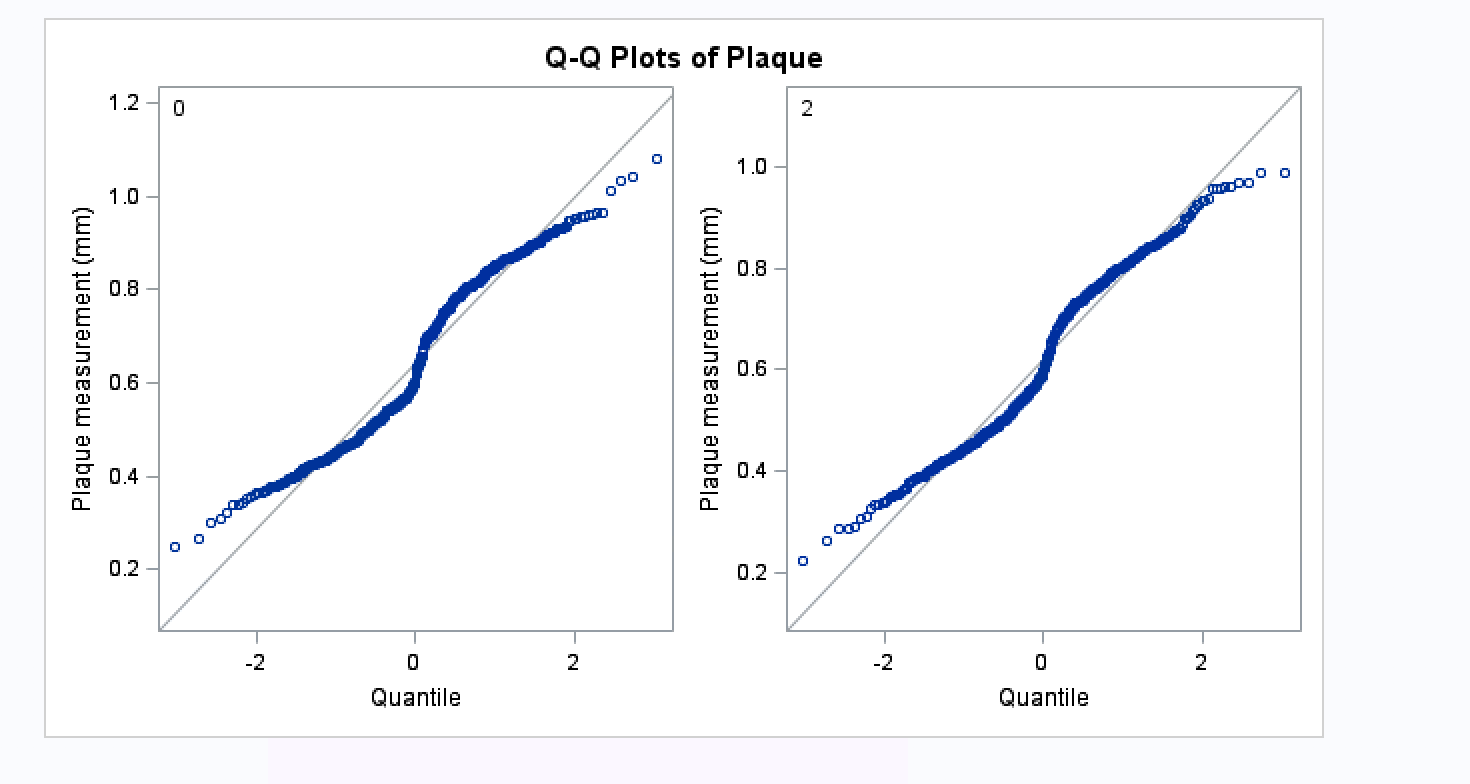
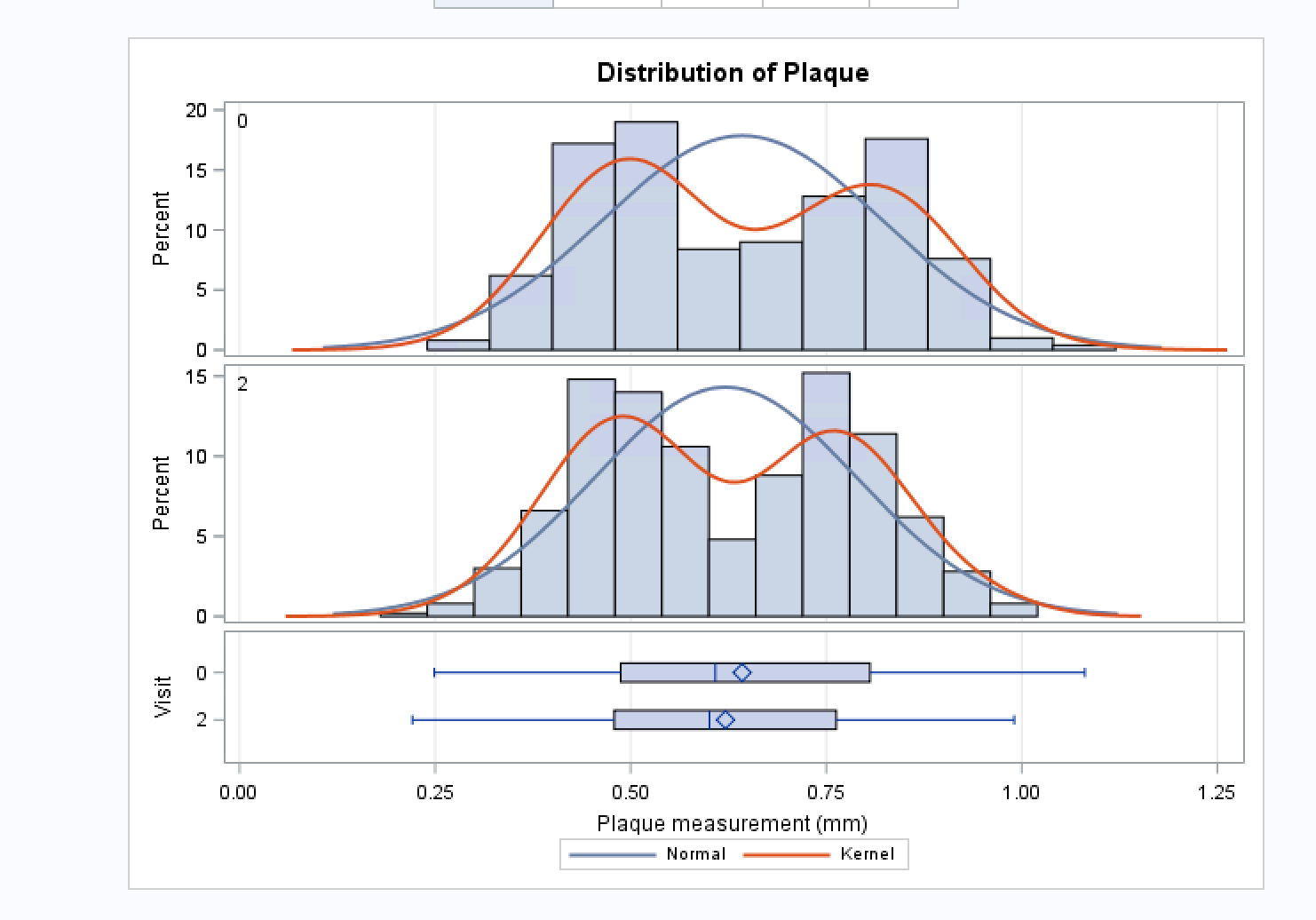
/\*Question 3B

H0 = For all people, mean plaque value difference at 0th visit and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.1402 , which is greater than 0.05 , we do not reject the null.\*/





ods graphics on;

**proc** **ttest** cochran ci=equal umpu;

where visit in (**0**,**2**);

class visit;

var Plaque;

**run**;

ods graphics off;

/\*Question 3C.\*/

**proc** **ttest** cochran ci=equal umpu;

where visit in (**1**,**2**) and Treatment = **1** and strata = **2**;

class visit;

var Plaque;

title ' Treatment = 1 & Strata = 2';

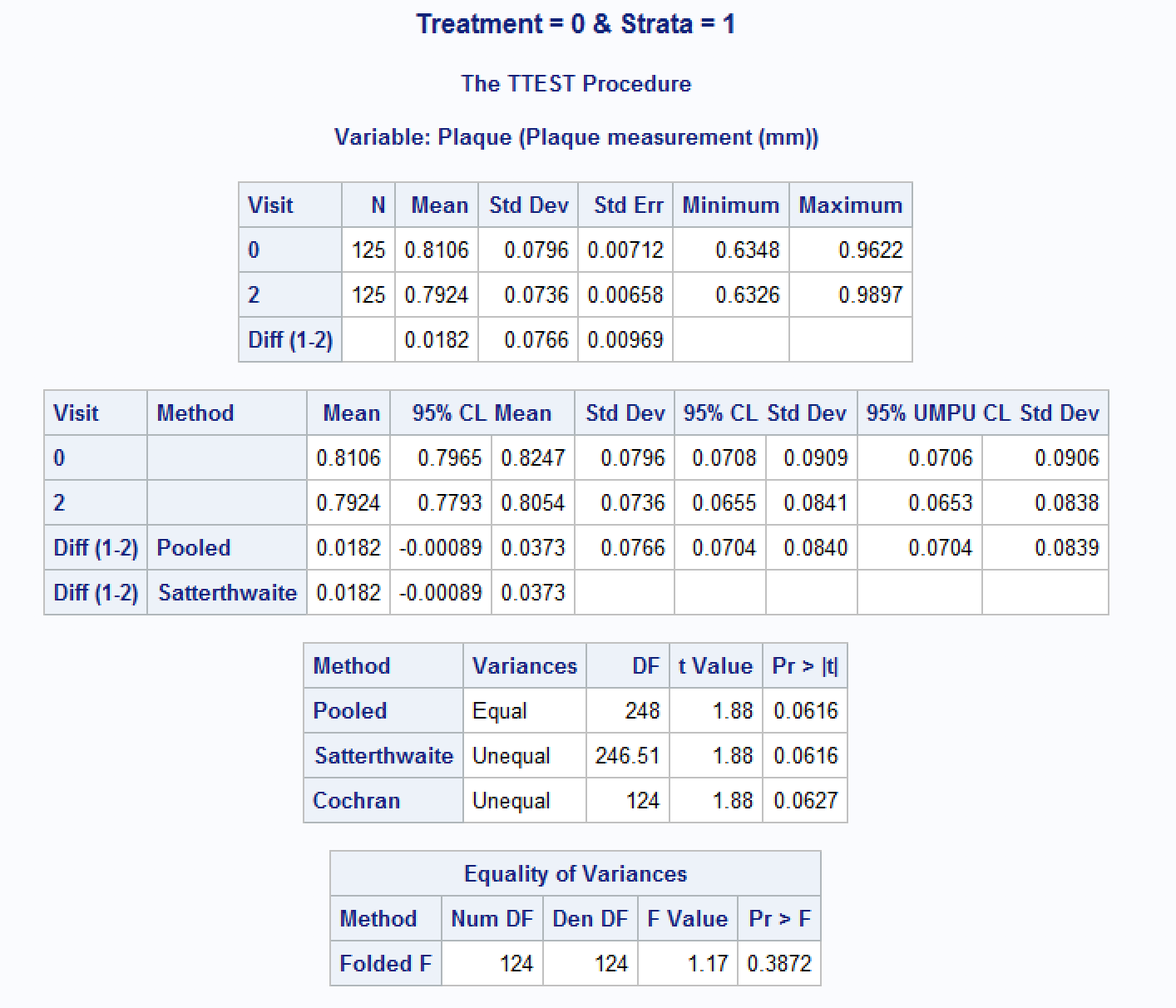
**run**;

/\*H0 = For people who did not receive treatment and whose Plaque level < 0.6,

the mean plaque value difference before the treatment and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.3872 , which is greater than 0.05 , we do not reject the null.\*/



/\*H0 = For people who did not receive treatment and whose Plaque level > 0.6,

the mean plaque value difference before the treatment and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.1562 , which is greater than 0.05 , we do not reject the null.\*/



/\*H0 = For people who did receive the treatment and whose Plaque level > 0.6, the mean plaque value difference before the treatment and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.0194 , which is less than 0.05 , so we reject the null.\*/

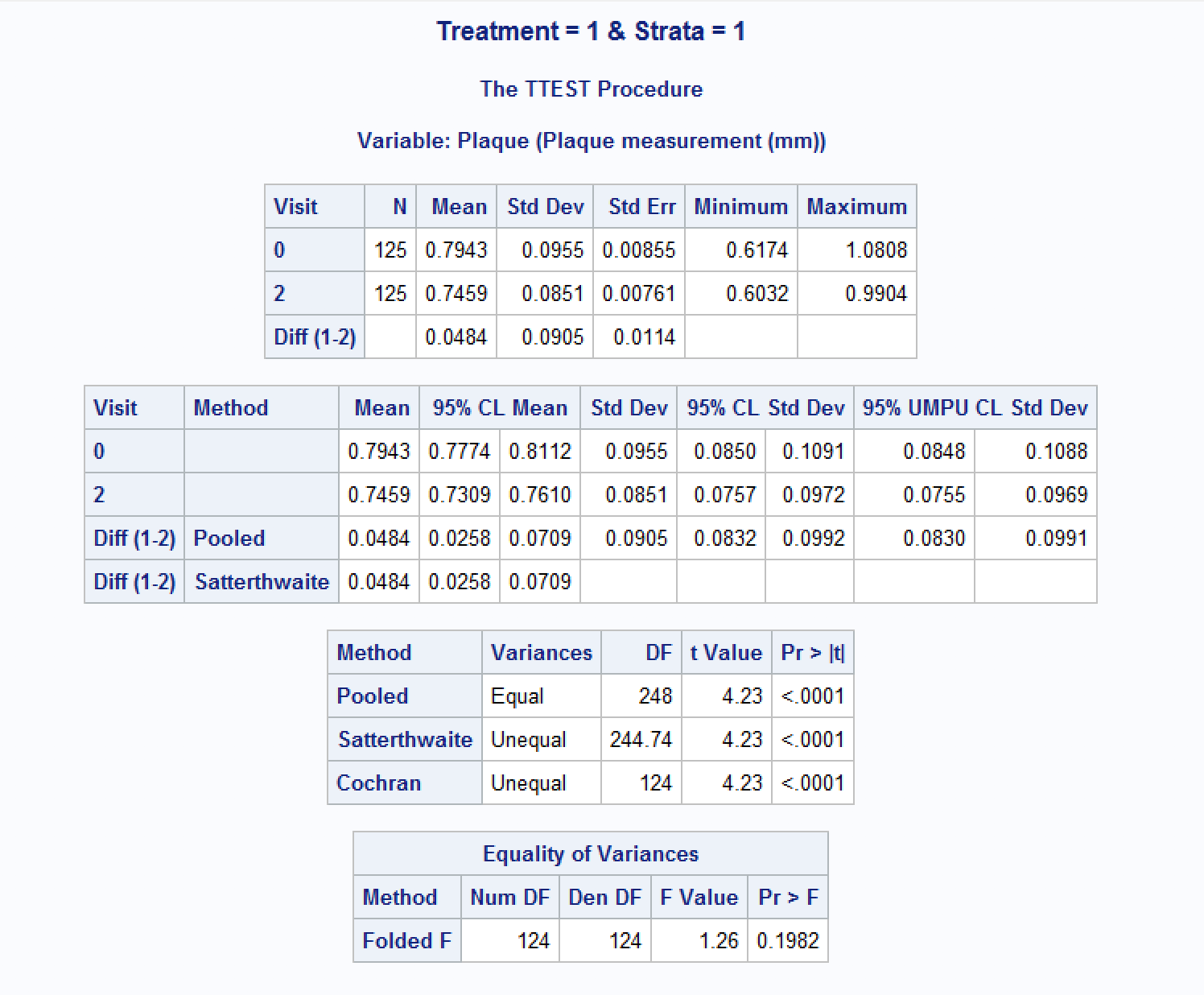


/\*H0 = For people who did receive the treatment and whose Plaque level < 0.6,

the mean plaque value difference before the treatment and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.1982 , which is greater than 0.05 , we do not reject the null.\*/



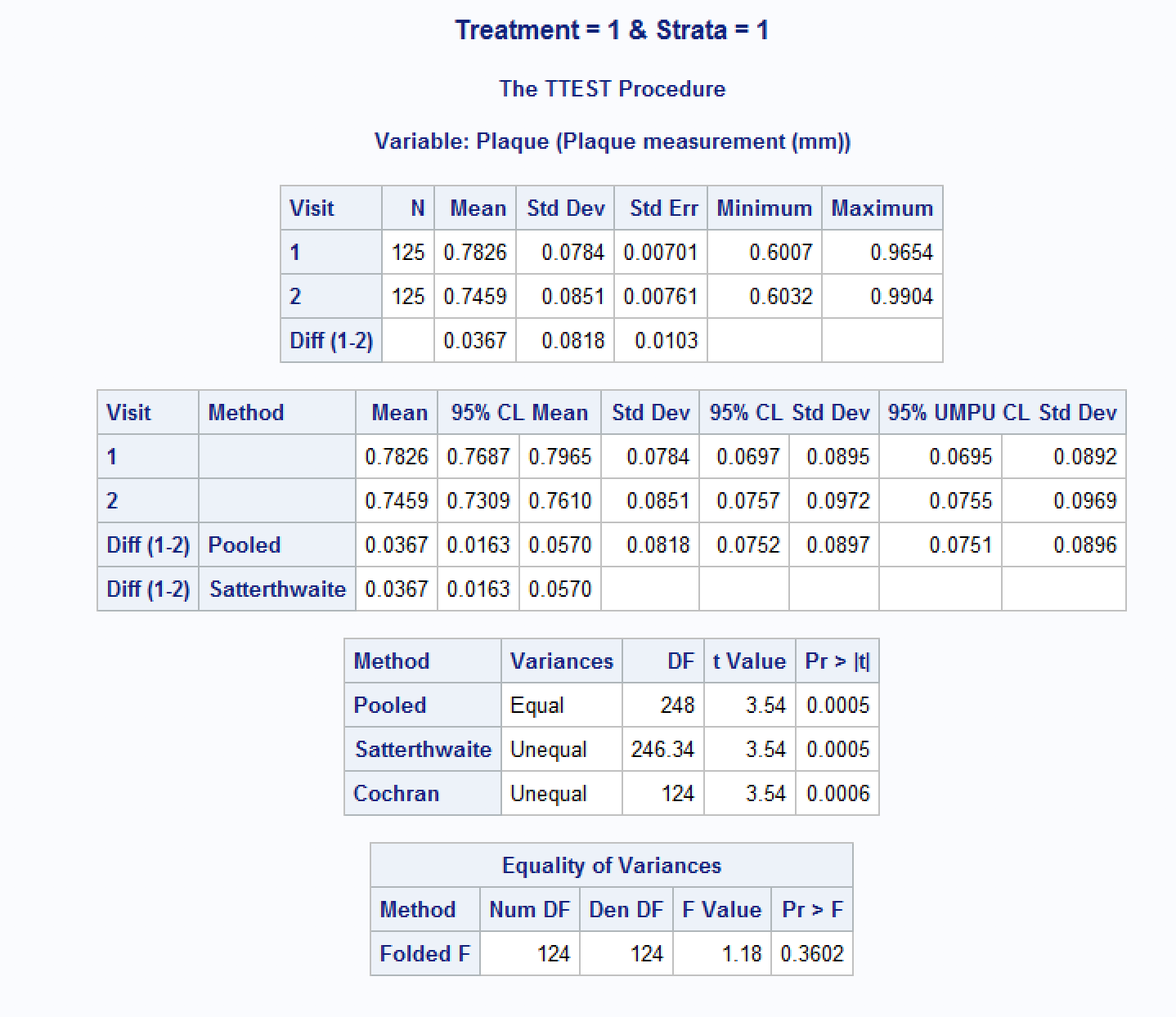
/\*Question 3D

H0 = For people who did receive the treatment and whose Plaque level < 0.6,

the mean plaque value difference before and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.3602 , which is greater than 0.05, so we do not reject the null.\*/



/\*H0 = For people who did not receive the treatment and whose Plaque level < 0.6, the mean plaque value difference before and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.0581 , which is greater than 0.05 , so we reject the null.\*/



/\*H0 = For people who did not receive the treatment and whose Plaque level > 0.6, the mean plaque value difference before and after the 2nd visit is not significant.

H1 = It is significant

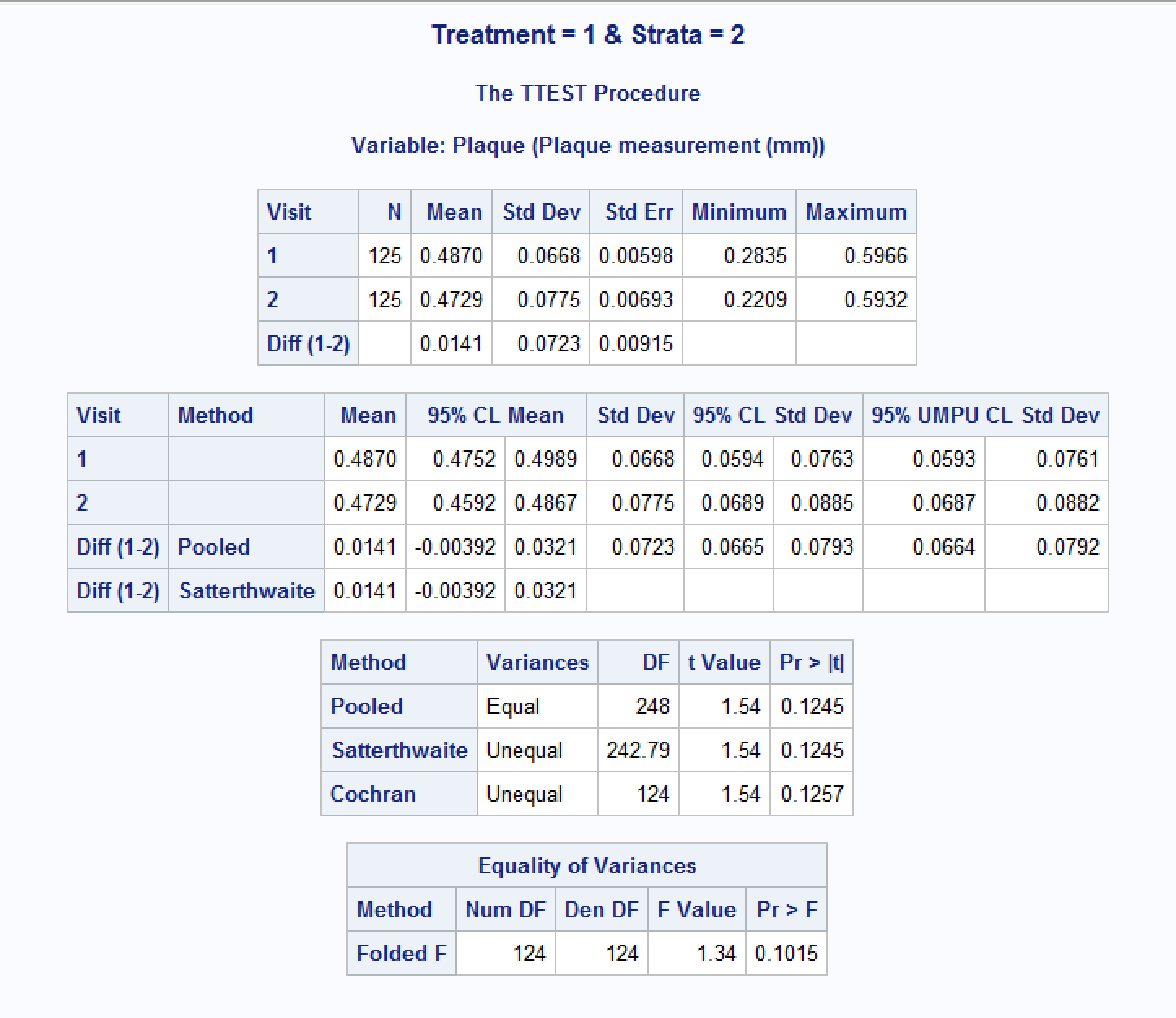
Since p val = 0.6955 , which is greater than 0.05 , so we do not reject the null.\*/



/\*H0 = For people who did receive the treatment and whose Plaque level > 0.6, the mean plaque value difference before and after the 2nd visit is not significant.

H1 = It is significant

Since p val = 0.1015 , which is greater than 0.05 , so we do not reject the null.\*/



/\*Q.4\*/

LIBNAME HW3 'C:\Users\mxt164730\Desktop\HW3';

**DATA** HW3.Lefties;

INFILE 'C:\Users\mxt164730\Desktop\HW3\lefties.sas7bdat';

**Proc** **print** data= HW3.Lefties;

**RUN**;

/\*Q.4 a. Using Chi Square testing to test whether there is an association between the writing hand and kicking foot preference.

Ho: Variables hand and foot are independent.

Ha: Variables hand and foot are not independent. \*/

**proc** **freq** data=HW3.Lefties order=data;

tables hand\*foot/ expected chisq;

output out=ChiSqDatafoot pchi lrchi;

title 'Chi-Square Tests for Hand and Foot';

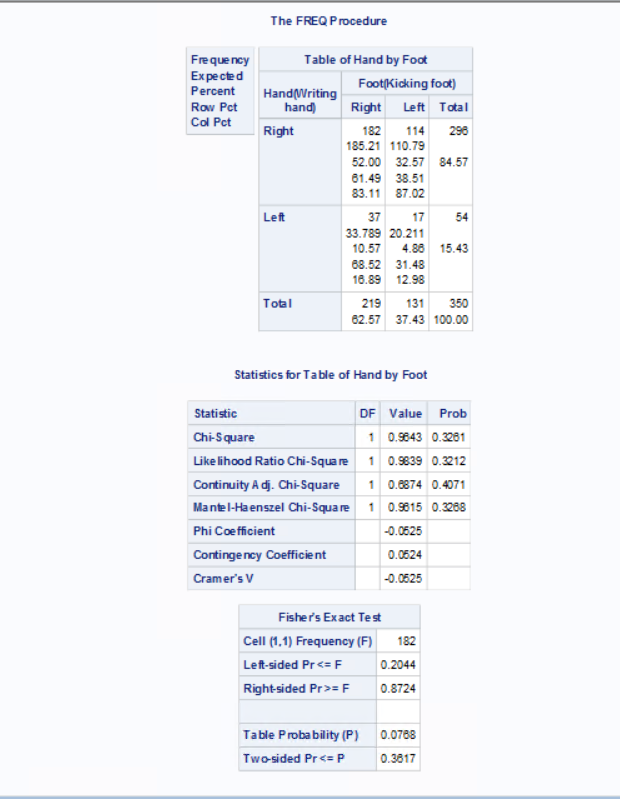
**run**;

**proc** **print** data=ChiSqDatafoot;

title1 'Chi-Square Statistics for Hand and Foot';

title2 'Output Data Set from the FREQ Procedure';

**Run**;



/\*Q.4 b. Using Chi Square testing to test whether there is an association between the writing hand and mousing hand preference.

Ho: Variables hand and mouse are independent.

Ha: Variables hand and mouse are not independent. \*/

**proc** **freq** data=HW3.Lefties order=data;

tables hand\*mouse/ expected chisq;

output out=ChiSqDataMouse pchi lrchi;

title 'Chi-Square Tests for Hand and Mouse';

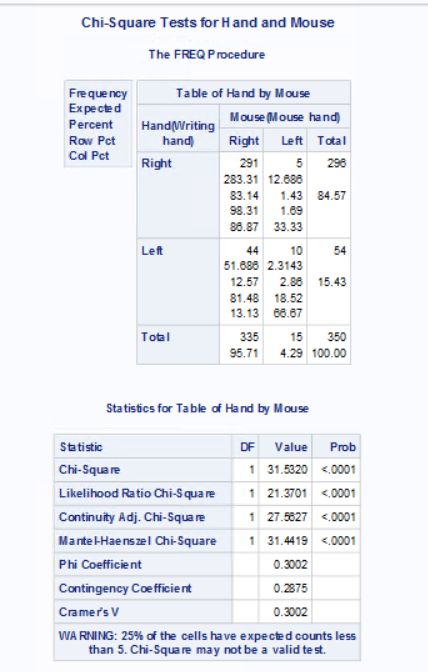
**run**;

**proc** **print** data=ChiSqDataMouse;

title1 'Chi-Square Statistics for Hand and Mouse';

title2 'Output Data Set from the FREQ Procedure of Hand and Mouse';

**Run**;



\*Q.4 c.Using Chi Square testing to test whether there is an association between the writing hand preference and gender.

Ho: Variables hand and gender are independent.

Ha: Variables hand and gender are not independent. \*/

**proc** **freq** data=HW3.Lefties order=data;

tables hand\*gender/ expected chisq;

output out=ChiSqDataGender pchi lrchi;

title 'Chi-Square Tests for Hand and Gender';

**run**;

**proc** **print** data=ChiSqDataGender;

title1 'Chi-Square Statistics for Hand and Gender';

title2 'Output Data Set from the FREQ Procedure of Hand and Gender';

**Run**;



/\*Q4. d. \*/

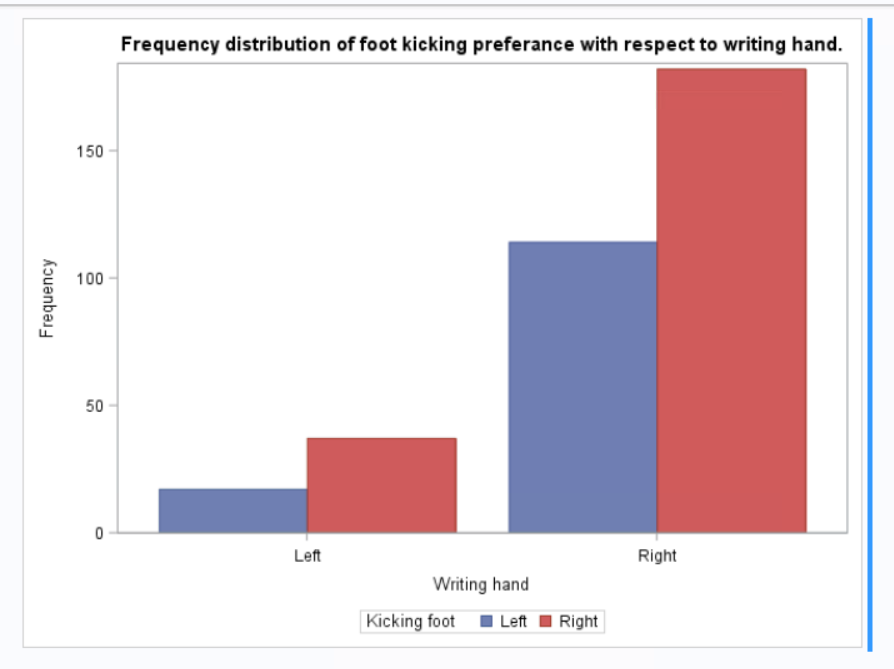
/\*part a\*/

**Proc** **sgplot** data= HW3.Lefties;

vbar hand / group= foot groupdisplay= cluster;

title1 'Frequency distribution of foot kicking preferance with respect to writing hand.';

**Run**;



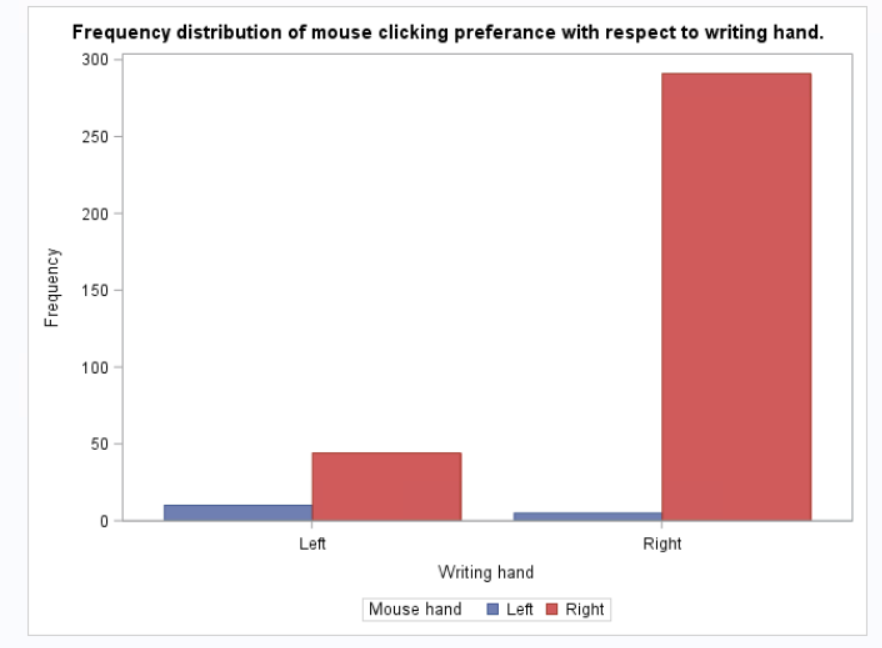
/\*part b\*/

**Proc** **sgplot** data= HW3.Lefties;

vbar hand / group= mouse groupdisplay= cluster;

title1 'Frequency distribution of mouse clicking preferance with respect to writing hand.';

**run**;



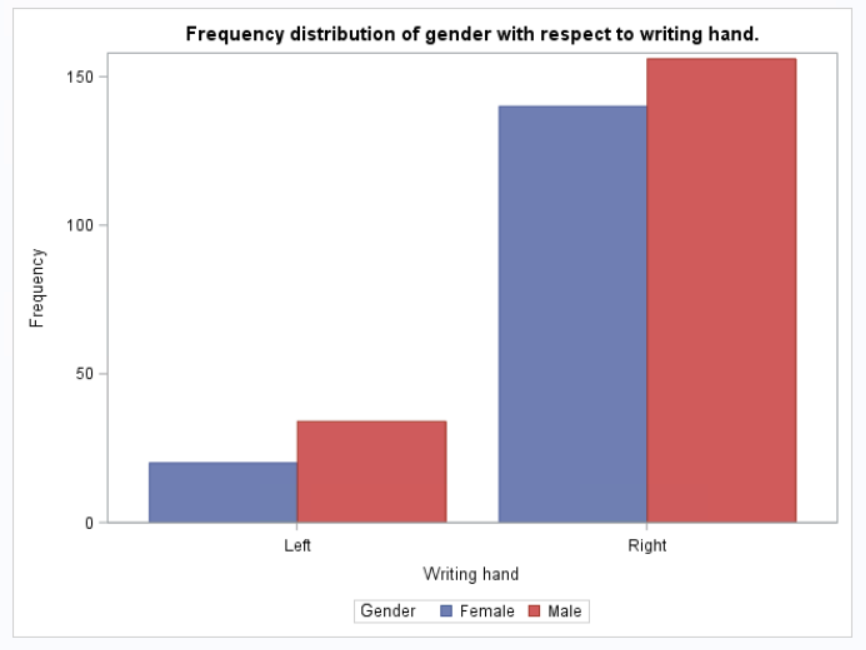
/\*part c\*/

**Proc** **sgplot** data= HW3.Lefties;

vbar hand / group= gender groupdisplay= cluster;

title1 'Frequency distribution of gender with respect to writing hand.';

**Run**;



/\*Q.4 e. Interpretation of the Chi-square tests in terms of p-value.

A. P-value is significantly higher than 0.05 suggesting that the null hypothesis is not rejected. The variables of writing hand and kicking foot are independent.

B. P-value is significantly lower than 0.05 suggesting that the null hypothesis is rejected. This means that the variables of writing hand and clicking mouse are not independent.

C. P-value is higher than 0.05 suggesting that the null hypothesis is not rejected. The variables of writing hand and gender are independent. \*/