STAT 6348
Applied Multivariate Analysis
Spring 2024
Project 2

This project is individual work. So do not consult with anybody in or out of class. You can ask me questions.

Sign on this page below and attach with your project. You project will not be graded without it.

This project is entirely my work. I have not discussed about this project with anybody in or out of class. I understand and have complied with the academic integrity policies written in the *Handbook of Operating Procedures* **of UT Dallas** https://policy.utdallas.edu/utdsp5003**.**

I understand that if any academic misconduct is suspected, it will be referred to the Office of Community Standards and Conduct https://conduct.utdallas.edu/.

YOUR NAME John Kenney

DATE 03/26/1997

YOUR SIGNATURE John Kenney

Directions:

- Arrange your report in the following order: (1) Answers, (2) Outputs, and (3) Code. If using R markdown, code and output may be interspersed. Otherwise, attach your R script at the end by simply copying-pasting (just the commands – not the whole output) with comments included to indicate the parts.

- Type your answers including comments. These should not exceed more than 5 pages.

- Attach <u>only the part of output that has been asked for</u> in the question.

- Your plots should be properly labeled and in a presentable format (e.g., properly labeled axes and without unnecessary white space). Do not present very small plots where it is difficult to see the pattern.

1. Consider the data in Table 6.8 of JW. The description of the data is given in Exercise 6.17 (you don't have to answer the questions in this exercise). Construct as well as plot a 95% confidence region (ellipsoid) for difference of means between Arabic and Word formats? Based on this region, do the means for the two formats differ? Also, construct Bonferroni 95% simultaneous confidence intervals for each individual mean difference and use them to test if the means for the two formats differ. Compare the conclusions obtained using the confidence region and simultaneous confidence intervals.

2. Consider column_3C.dat data set consisting of orthopedic patients classified into 3 classes (normal, disk hernia or spondylolisthesis) or 2 classes (normal or abnormal). The following convention is used for the class labels: DH (Disk Hernia), Spondylolisthesis (SL), and Normal (NO). Each patient has six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine (in this order): pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius, and grade of spondylolisthesis. We will work with only three attributes – lumbar lordosis angle, sacral slope, and pelvic radius.

   (a) Make scatterplot matrix of the three attributes with side-by-side boxplots of the two classes of patients in the diagonal (use different plotting symbols and colors to represent the two classes of patients; make sure you choose appropriate scales for x- and y-axes so that all observations for both types get plotted; add a legend in the plot to distinguish the two types). Use the plot to compare normal and abnormal classes.

   (b) Check for univariate and multivariate normality of the three attributes. If normality appears to be violated, explore transformations that may help. For each variable, include only one transformation and the corresponding plot (after transformation) that appears to be the most helpful.

(c) Retain the transformations from the previous part and test at 5% level whether the two classes differ in terms of the two means jointly. You may assume that the variance-covariance matrices for the two types are the same. Write the relevant hypotheses, test statistic, p-value, and conclusion.

(d) Make the scatterplot of part (a) but now representing the three classes in one plot (use the original scale for the plots). Use this and the previous plots to comment if two classes are sufficient for classification or three classes provide extra information.

(e) Conduct a MANOVA with three classes assuming equality of the variance-covariance matrices and interpret the results at 5% level (use transformed scale, same as in part c). Include all steps of hypothesis test as mentioned in part (c). Obtain 95% simultaneous CI for differences in mean components and interpret the results.

Stat 6348 Project 2

John Kenney

2024-03-26

Answers

Question 1

95% Confidence Region
$\{\mu_D : 32 * 9.997992e^{-05} * (142.2500 - \mu_1)^2 + 32 * 0.0002411671 * (164.6719 - \mu_2)^2 - 2 * 32 * -1.026697e^{-04} * (142.2500 - \mu_1) * (164.6719 - \mu_2) \leq 6.809\}$

Zero is not contained in our 95% confidence region therefore we can say that the means of the two formats differ. We can also observe that the major axis is 6.42 times longer than the minor axis.

Since the zero is not contained in either of the intervals we can say that there is a difference in the individual means of the DiffParity and SameParity based on the difference of the Arabic and Word formats.

The results of the 95% Simultaneous Bonferroni CI agree with the result from the 95% confidence region (ellipsoid). Therefore it is reasonable to conclude that there is a difference in mean response time of correctly identifying number parity for the two formats Word and Arabic.

Question 2

Question 2 (a)

The normal Patients seem to be grouped fairly close together. For Pelvic Radius there are several outliers in the abnormal class. The two groups seem to have different centers. Also Sacral Slope and Lumbar Lordosis Angle seem to be have a positive linear relationship.

Question 2 (b)

For fixing univariate Normality
Transformations applied are as follows:
Lumbar Lordosis Angle -> (Lumbar Lordosis Angle)$^{1/3}$
Sacral Slope -> $\sqrt{\text{Sacral Slope}}$
Pelvic Radius -> Pelvic Radius (No transformation applied)

The multivariate normality assumption seems to be fairy adequate except for the two large outliers. Observations that are multivariate Outliers: 86, 116, 163, 168, 180, 181, 198, 202

Question 2 (c)

Assume $\Sigma_{Normal} = \Sigma_{Abnormal}$ and $\alpha = 0.05$
$H_0 : \boldsymbol{\mu_{Normal}} - \boldsymbol{\mu_{Abnormal}} = \mathbf{0}$ vs $H_A : \boldsymbol{\mu_{Normal}} - \boldsymbol{\mu_{Abnormal}} \neq \mathbf{0}$
$T^2 = (\overline{\boldsymbol{X}}_{Normal} - \overline{\boldsymbol{X}}_{Normal})^\top [(\frac{1}{n_1} + \frac{1}{n_2})\boldsymbol{S}]^{-1} (\overline{\boldsymbol{X}}_{Normal} - \overline{\boldsymbol{X}}_{Normal}) = 66.09492$
$F^* = \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{p,n_1+n_2-p-1}(\alpha) = 7.953987$
$T^2 = 66.095 > 7.954 = F^*$, Therefore, we reject the Null Hypothesis and conclude that the means of Normal Patients and Abnormal Patients are different.

Our P-value $= 7.209788e - 13 < 0.05 = \alpha$ which is very small. Therefore, we can conclude with a 95% confidence level that the means of the Normal Patients and Abnormal Patients are different.

Question 2 (d)

There is a more significant difference in the centers of the different classes based on the box plots. In addition, the 3 classes seem to be fairly separated in the scatter plots. In comparison to the two classes we can see in the three classes boxplots that DH's median is always below Normal and SL's median is above Normal except for pelvic radius. This tells us that when we have two classes we will be closer to the the normal center compared with three distinct classes. In conclusion, I believe that two classes are not sufficient for the data because three classes provide extra information on the distributions of the patients.

Question 2 (e)

Assume $\Sigma_{NO} = \Sigma_{DH} = \Sigma_{SL}$ and $\alpha = 0.05$
$H_0 : \boldsymbol{\mu}_{NO} = \boldsymbol{\mu}_{DH} = \boldsymbol{\mu}_{SL}$ vs H_A : at least one mean vector is different
$\Lambda^* = \frac{|W|}{|B+W|}$
$F = \left(\frac{\Sigma_l n_l - p - 2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) = 51.85378$
$\left(\frac{\Sigma_l n_l - p - 2}{p}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p,2(\Sigma_l n_l - p - 2)}$
$Critical = F_{2p,2(\Sigma_l n_l - p - 2)}(\alpha) = 2.113426$
$F = 51.85378 > 2.113426 = Critical$, Therefore, we reject the Null Hypothesis and conclude that at least one of the means are different for the three classes of patients.

Our P-value $< 2.2e - 16 < 0.05 = \alpha$ which is very small. Therefore, we can conclude with a 95% confidence level that at least one of the means are different for the three classes of patients.

We can see from the from the table for the difference of means 95% Bonferroni CI that DH - SL for the variable Pelvic Radius is the only interval that contains zero indicating the difference is not significant for this comparison. The other comparisons of the mean components do not contain zero indicating there is a significant difference in the mean components for the 3 class variables.

Outputs

Question 1

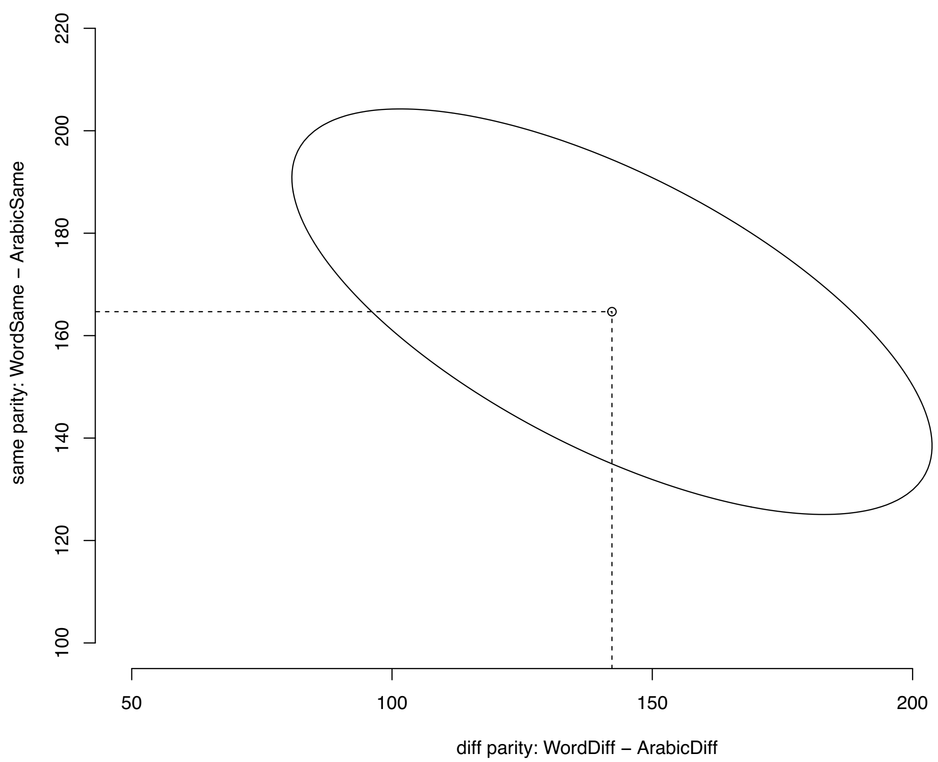

Figure 1: 95% Confidence Region for the Difference in means of Arabic and Word Formats

Table 1: 95% Bonferroni CI for the Individual Means

	Mean	95% Bonferoni CI
DiffParity	142.25	(86.739 , 197.761)
SameParity	164.672	(128.93 , 200.414)

Question 2

Question 2 (a)

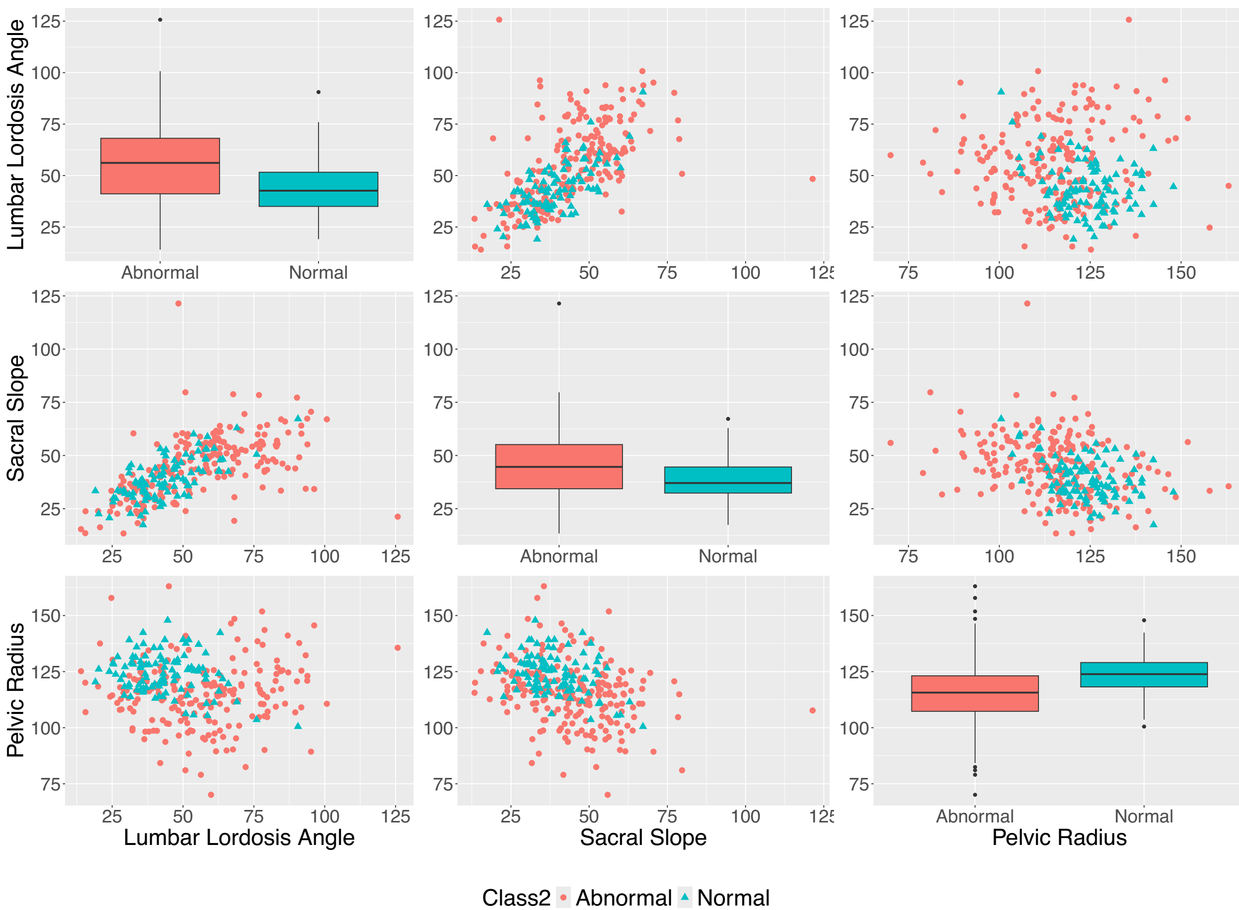

Figure 2: ScatterPlot Matrix given Patient Class

Question 2 (b)

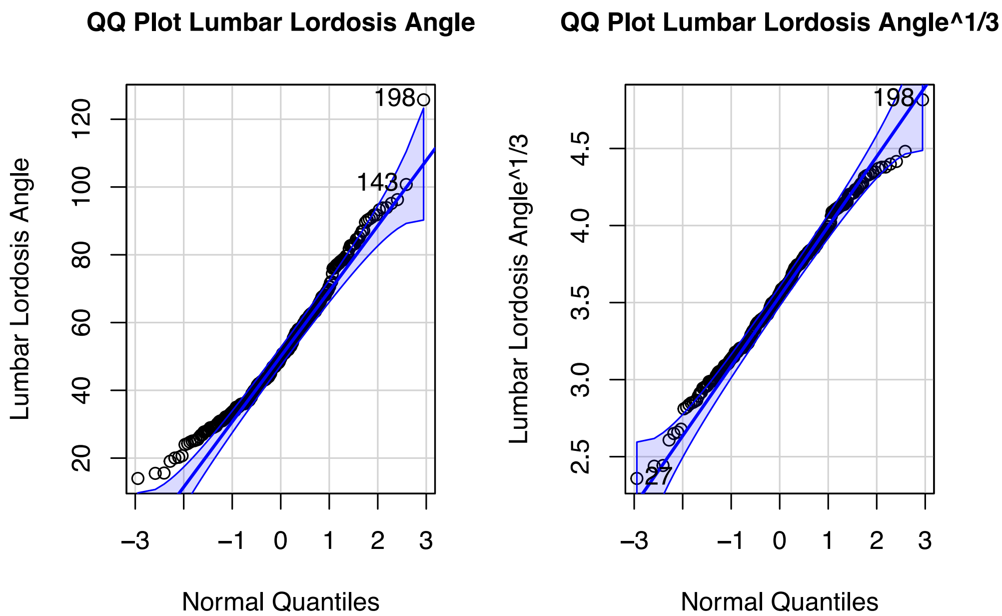

Figure 3: Transformation applied is (Lumbar Lordosis Angle)$^{1/3}$

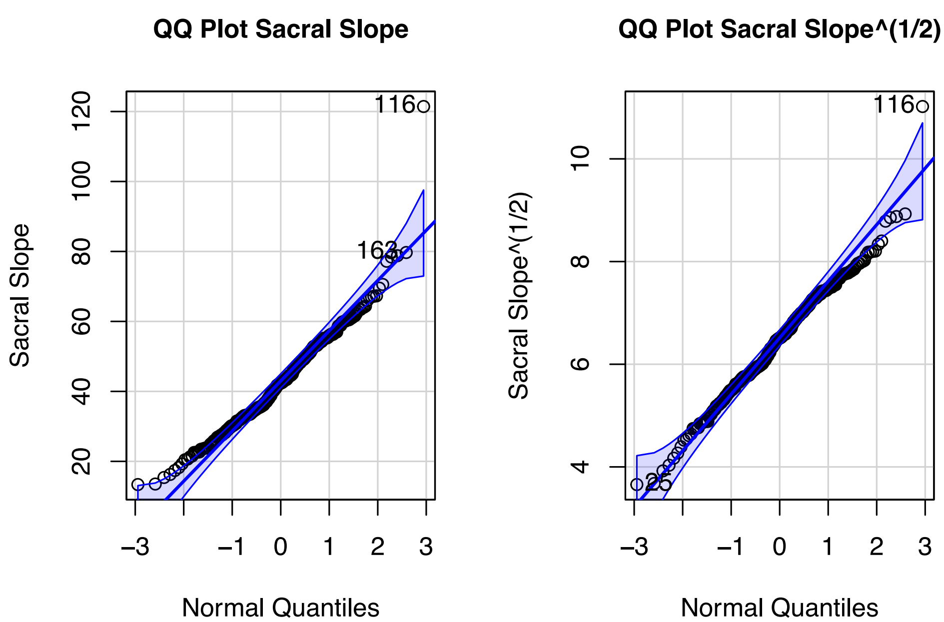

Figure 4: Transformation applied is $\sqrt{\text{Sacral Slope}}$

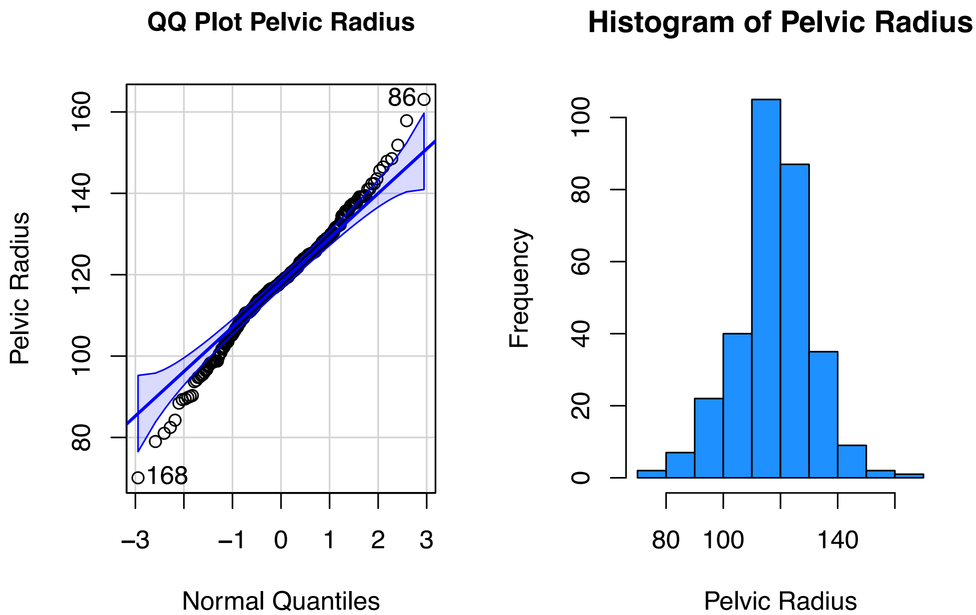

Figure 5: No transformation applied for Pelvic Radius

```
## [1] "Observations that are multivariate Outliers"
## [1]  86 116 163 168 180 181 198 202
```

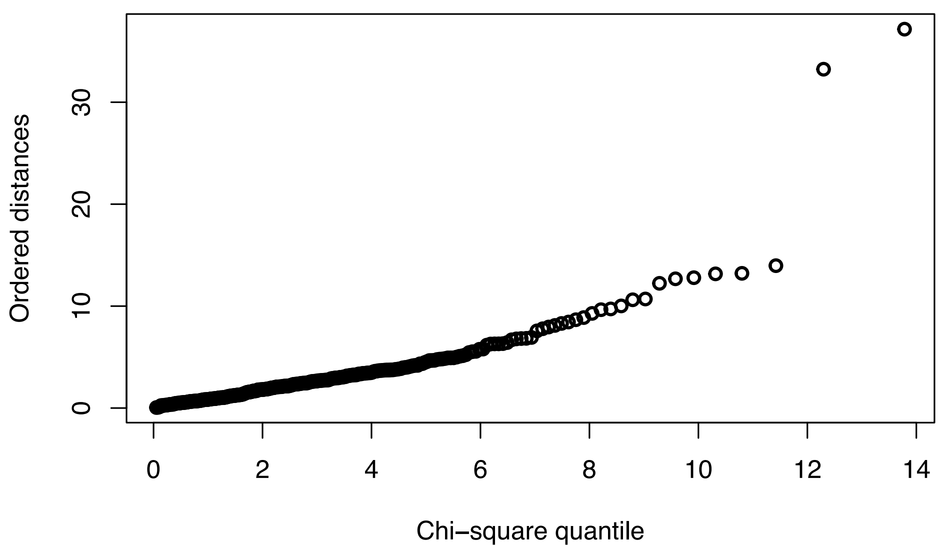

Figure 6: Chi-square plot on the Transformed variables

Question 2 (c)

Question 2 (d)

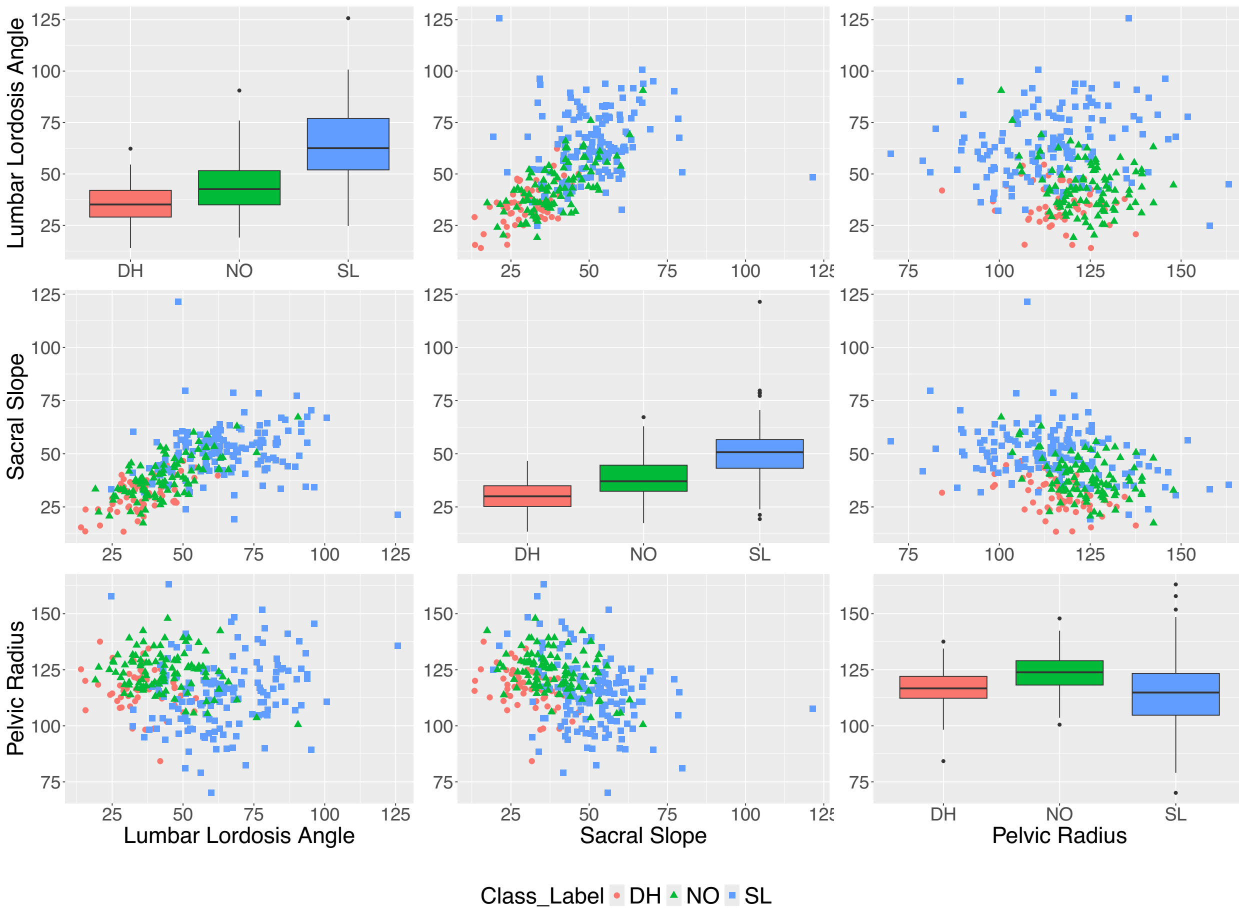

Figure 7: ScatterPlot Matrix given the three Patient Classes

Question 2 (e)

Table 2: MANOVA Table

Source of Variation	SSCP Matrix	d.f.
Treatments	$B = \begin{bmatrix} 24.42924 & 54.68592 & -215.0291 \\ 54.68592 & 124.80770 & -389.4338 \\ -215.02914 & -389.43380 & 5426.5048 \end{bmatrix}$	$g-1=3-1=2$
Residuals	$W = \begin{bmatrix} 30.48061 & 31.49275 & 38.72387 \\ 31.49275 & 196.84666 & -1043.21468 \\ 38.72387 & -1043.21468 & 49377.50484 \end{bmatrix}$	$\Sigma_{l=1}^{g} n_l - g = 310 - 3 = 307$
Total	$B+W = \begin{bmatrix} 54.90986 & 86.17867 & -176.3053 \\ 86.17867 & 321.65435 & -1432.6485 \\ -176.30527 & -1432.64848 & 54804.0096 \end{bmatrix}$	$\Sigma_{l=1}^{g} n_l - 1 = 310 - 1 = 309$

Table 3: 95% Simultaneous Bonferroni CI for diff of means component

	Lumbar Lordosis Angle	Sacral Slope	Pelvic Radius
NO - DH	(0.0753 , 0.3626)	(0.368 , 1.0983)	(1.6326 , 13.1998)
NO - SL	(-0.5722 , -0.345)	(-1.1764 , -0.599)	(4.8005 , 13.9452)
DH - SL	(-0.8119 , -0.5431)	(-1.9625 , -1.2793)	(-3.4534 , 7.3667)

Code
Question 1

```
tab68 <-
↪  read.table("https://utdallas.box.com/shared/static/nkskglp1sfzkipm8vxslrzlkbnd2n3x4.dat",
                    col.names = c("WordDiff","WordSame","ArabicDiff","ArabicSame"))
# number of subjects is 32
# asked is 2 numbers where both even or odd "same" or "diff"

# Arabic = ("2","3")
# Word = ("two","three")

# median reaction time of correct responses for each subject

# X1 = median reaction time for word format-different parity combination
# X2 = median reaction time for word format-same parity combination
# X1 = median reaction time for arabic format-different parity combination
# X1 = median reaction time for arabic format-same parity combination
```

```
diff<-tab68$WordDiff-tab68$ArabicDiff
#diff
same<-tab68$WordSame-tab68$ArabicSame
#same
arabicword<-data.frame(diff,same)
#arabicword
```

```
n <- 32
p <- 2
mean<-colMeans(arabicword)
#mean
var.cov<-cov(arabicword)
#var.cov
var.inv<-solve(var.cov)
#var.inv
mu<-matrix(c(0,0),nrow=2)
#mu
#as.matrix(mean)
T2 <- n*t(as.matrix(mean))%*% var.inv %*% as.matrix(mean)
#T2
crit.value <- (((n-1)*p)/(n-p))*qf(0.05,2,32, lower.tail=F)
#crit.value
```

```
ellipse<-function(mu1,mu2)
{
32*9.997992e-05*(142.2500 -mu1)^2+32*
↪  0.0002411671*(164.6719-mu2)^2-2*32*-1.026697e-04*(142.2500-mu1)*(164.6719-mu2)
}
mu1<-seq(50,225,1)
mu2<-seq(100,225,1)
z <- t(sapply(mu1,ellipse,mu2))
```

```
contour(mu1,mu2,z, levels=(31*2)*qf(0.05,2,32, lower.tail=F)/30, drawlabels=F, axes=T,
↪  frame.plot=F, xlab="diff parity: WordDiff - ArabicDiff", ylab="same parity: WordSame
↪  - ArabicSame")
points(142.2500, 164.6719 ,pch=1)
segments(0,164.6719,142.2500, 164.6719,lty=2)
segments(142.2500,0,142.2500, 164.6719, lty=2)
```

```
#eigen(var.cov)$values[1]/eigen(var.cov)$values[2]
```

```
#Test for H0: mu_d = 0
p <- 2
n <- 32
crit.value<-qt(1-.05/(2*p),n-1)
#crit.value

#95% Simultaneous CI

summary <-
↪  rbind(c(round(mean[1],3),paste("(",round(mean[1]-(crit.value)*sqrt(var.cov[1,1]/n),3),
↪  ",",round(mean[1]+(crit.value)*sqrt(var.cov[1,1]/n),3),")")),
                 c(round(mean[2],3),paste("(
                 ↪  ",round(mean[2]-(crit.value)*sqrt(var.cov[2,2]/n),3), ",",
                 ↪  round(mean[2]+(crit.value)*sqrt(var.cov[2,2]/n),3),")")))
row.names(summary) <- c("DiffParity","SameParity")
colnames(summary) <- c("Mean","95% Bonferoni CI")
#summary
library(kableExtra)
knitr::kable(summary,booktabs = T,caption = "$95\\%$ Bonferroni CI for the Individual
↪  Means", align = "cc") %>%
  kable_styling(latex_options = c("hold_position"))
```

```
datafull <-
↪  read.table("https://utdallas.box.com/shared/static/90jodfx6t2xv0yi1b2tpz1z2bbzqftpm.dat",
↪
                       col.names = c("Pelvic_Incidence", "Pelvic_Tilt",
                       ↪  "LumbarLordosisAngle", "Sacral_Slope", "Pelvic_Radius",
                       ↪  "Spondylolisthesis","Class_Label"))
data <- datafull[,c("LumbarLordosisAngle","Sacral_Slope","Pelvic_Radius","Class_Label")]
#data$Class_Label <- factor(data$Class_Label)

# 3 Class Labels
# DH: Disk Hernia
# SL: Spondylolisthesis
# NO: Normal

# 2 class labels
# normal
# abnormal
```

```r
data$Class2 <- ifelse(data$Class_Label == "NO","Normal","Abnormal")
#data$Class2 <- factor(data$Class2)
#data$Class_Label <- factor(data$Class_Label)
#data
#summary(data)
```

Question 2
Question 2 (a)

```r
library(ggpubr)
library(gridExtra)
textsize <- 26
p1 <- ggplot(data, aes(x = Class2, y = LumbarLordosisAngle, group = Class2,fill =
↪  Class2)) +
  geom_boxplot() + theme(legend.position="none")+ xlab(NULL)+ylab("Lumbar Lordosis
  ↪  Angle")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p2 <- ggplot(data, aes(x = Sacral_Slope, y = LumbarLordosisAngle, color = Class2,pch =
↪  Class2)) +
  geom_point(size = 3) + theme(legend.position="none")+ xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p3 <- ggplot(data, aes(x = Pelvic_Radius, y = LumbarLordosisAngle, color = Class2,pch =
↪  Class2)) +
  geom_point(size = 3) + theme(legend.position="none") + xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p4 <- ggplot(data, aes(x = LumbarLordosisAngle, y = Sacral_Slope, color = Class2,pch =
↪  Class2)) +
  geom_point(size = 3) + theme(legend.position="none") + xlab(NULL) + ylab("Sacral
  ↪  Slope")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p5 <- ggplot(data, aes(x = Class2, y = Sacral_Slope, group = Class2,fill = Class2)) +
```

```
  geom_boxplot() + theme(legend.position="none")+ xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p6 <- ggplot(data, aes(x = Pelvic_Radius, y = Sacral_Slope, color = Class2,pch = Class2))
↪  +
  geom_point(size = 3)+ theme(legend.position="none")+ xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p7 <- ggplot(data, aes(x = LumbarLordosisAngle, y = Pelvic_Radius, color = Class2,pch =
↪  Class2)) +
  geom_point(size = 3) + theme(legend.position="none")+xlab("Lumbar Lordosis
  ↪  Angle")+ylab("Pelvic Radius")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p8 <-  ggplot(data, aes(x = Sacral_Slope, y = Pelvic_Radius, color = Class2,pch =
↪  Class2)) +
  geom_point(size = 3) + theme(legend.position="bottom")+ ylab( NULL)+xlab("Sacral
  ↪  Slope")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p9 <- ggplot(data, aes(x = Class2, y = Pelvic_Radius, group = Class2,fill = Class2)) +
  geom_boxplot() + theme(legend.position="none")+ ylab( NULL)+xlab("Pelvic Radius")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title


get_only_legend <- function(plot) {

# get tabular interpretation of plot
plot_table <- ggplot_gtable(ggplot_build(plot))

#  Mark only legend in plot
```

```r
legend_plot <- which(sapply(plot_table$grobs, function(x) x$name) == "guide-box")

# extract legend
legend <- plot_table$grobs[[legend_plot]]

# return legend
return(legend)
}
# extract legend from plot1 using above function
legend <- get_only_legend(p8)
p8 <- p8+theme(legend.position="none")
g <- ggarrange(p1, p2, p3, p4,p5,p6,p7,p8,p9, ncol=3,nrow = 3)
```

```r
ggarrange(g,legend,nrow = 2, heights = c(10, 1))
```

Question 2 (b)

```r
transdata <- data
#transdata[,1:3] <- scale(transdata[,1:3])
```

```r
par(mfrow = c(1,2))
library(MASS)
library(car)
qqPlot(data$LumbarLordosisAngle,xlab = "Normal Quantiles",ylab="Lumbar Lordosis Angle") +
↪  title("QQ Plot Lumbar Lordosis Angle", cex.main = 1)
#hist(data$LumbarLordosisAngle,xlab = "Lumbar Lordosis Angle",main="Histogram of Lumbar
↪  Lordosis Angle",col = "dodgerblue")

powerTransform(data$LumbarLordosisAngle)

#QQ Plot of the Box-Cox transformed data with the specified power
#qqnorm(bcPower(data$LumbarLordosisAngle, 0.3251683 ), main="QQ Plot after Box-Cox
↪  Transformation")
#qqline(bcPower(data$LumbarLordosisAngle,0.3027923))
#shapiro.test(bcPower(df$Age,0.3027923))

x <- data$LumbarLordosisAngle
powerTransform(x)
t <- powerTransform(x)$lambda[[1]]
#QQ Plot of the Box-Cox transformed data with the specified power
#qqnorm(bcPower(x, t ), main="QQ Plot after Box-Cox Transformation")
#qqline(bcPower(x,t))
#qqPlot(bcPower(x,t),xlab = "Normal Quantiles",ylab="variable") + title("QQ Plot",
↪  cex.main = 1)
#hist(bcPower(x,t),xlab = "varaible",main="Histogram of variable",col = "dodgerblue")
#shapiro.test(bcPower(x,t))


#par(mfrow = c(1,2))
```

```r
transdata$LumbarLordosisAngle <- (data$LumbarLordosisAngle)^(t)
qqPlot(transdata$LumbarLordosisAngle,xlab = "Normal Quantiles",ylab="Lumbar Lordosis
↪  Angle^1/3") + title("QQ Plot Lumbar Lordosis Angle^1/3", cex.main = 1)
```

```r
#hist(transdata$LumbarLordosisAngle,xlab = "Lumbar Lordosis Angle^(1/3)",main="Histogram
↪  of Lumbar Angle^(1/3)",col = "dodgerblue")

#x <- (df$Age)^(1/3)
#shapiro.test(x)
## Normal Probability plot(q-q plot)
#par(mfrow = c(1,2))
#library(MASS)
#library(car)
#qqPlot(x,xlab = "Normal Quantiles",ylab="Age") + title("QQ Plot", cex.main = 1)
#hist(x,xlab = "Age",main="Histogram of Age",col = "dodgerblue")
```

```r
par(mfrow = c(1,2))
library(MASS)
library(car)
x <- data$Sacral_Slope

qqPlot(x,xlab = "Normal Quantiles",ylab="Sacral Slope") + title("QQ Plot Sacral Slope",
↪  cex.main = 1)
#hist(x,xlab = "Sacral Slope",main="Histogram of Sacral Slope",col = "dodgerblue")

powerTransform(x)
#par(mfrow = c(1,2))
t <- powerTransform(x)$lambda[[1]]
#qqPlot(bcPower(x,t),xlab = "Normal Quantiles",ylab="variable") + title("QQ Plot",
↪  cex.main = 1)
#hist(bcPower(x,t),xlab = "varaible",main="Histogram of variable",col = "dodgerblue")



#par(mfrow = c(1,2))
transdata$Sacral_Slope <- (data$Sacral_Slope)^(.5)
qqPlot(transdata$Sacral_Slope,xlab = "Normal Quantiles",ylab="Sacral Slope^(1/2)") +
↪  title("QQ Plot Sacral Slope^(1/2)", cex.main = 1)
```

```r
#hist(transdata$Sacral_Slope,xlab = "Sacral Slope^(1/2)",main="Histogram of Sacral
↪  Slope^(1/2)",col = "dodgerblue")
```

```r
par(mfrow = c(1,2))
library(MASS)
library(car)
x <- data$Pelvic_Radius

qqPlot(x,xlab = "Normal Quantiles",ylab="Pelvic Radius") + title("QQ Plot Pelvic Radius",
↪  cex.main = 1)
```

```
hist(x,xlab = "Pelvic Radius",main="Histogram of Pelvic Radius",col = "dodgerblue")
```

```
#powerTransform(x)
#par(mfrow = c(1,2))
#t <- powerTransform(x)$lambda[[1]]
#qqPlot(bcPower(x,t),xlab = "Normal Quantiles",ylab="variable") + title("QQ Plot",
↪  cex.main = 1)
#hist(bcPower(x,t),xlab = "varaible",main="Histogram of variable",col = "dodgerblue")



#par(mfrow = c(1,2))
#transdata$Pelvic_Radius<- (data$Pelvic_Radius)^(1.5)
#qqPlot(transdata$Pelvic_Radius,xlab = "Normal Quantiles",ylab="Pelvic Radius^(3/2)") +
↪  title("QQ Plot Pelvic Radius^(3/2)", cex.main = 1)
#hist(transdata$Pelvic_Radius,xlab = "Pelvic Radius^(3/2)",main="Histogram of Pelvic
↪  Radius^(3/2)",col = "dodgerblue")
```

```
#Function for calculating chi-square plot (copy and paste)
chisplot <- function(x) {
    if (!is.matrix(x)) stop("x is not a matrix")

    ### determine dimensions
    n <- nrow(x)
    p <- ncol(x)
    #
    xbar <- apply(x, 2, mean)
    S <- var(x)
    S <- solve(S)
    index <- (1:n)/(n+1)
    #
    xcent <- t(t(x) - xbar)
    di <- apply(xcent, 1, function(x,S) x %*% S %*% x,S)

    print("Observations that are multivariate Outliers")
    print(which(di > qchisq(.99,p)))
    quant <- qchisq(index,p)
    plot(quant, sort(di), ylab = "Ordered distances",
         xlab = "Chi-square quantile", lwd=2,pch=1)
}
#par(mfrow = c(1,2))
#chisplot(as.matrix(data[,1:3]))
chisplot(as.matrix(transdata[,1:3]))
```

Question 2 (c)

```
Normdata <- subset(transdata,transdata$Class2 == "Normal")[,1:3]
#colMeans(Normdata)

abnormdata <- subset(transdata,transdata$Class2 == "Abnormal")[,1:3]
```

```r
#colMeans(abnormdata)

#var(Normdata)

#var(abnormdata)
n1 <- dim(Normdata)[1]
#n1
n2 <- dim(abnormdata)[1]
#n2

pooled <- ((n1-1)*var(Normdata) + (n2-1)*var(abnormdata))/(n1+n2-2)
#pooled
T2 <- (n1*n2)/(n1+n2)*t(colMeans(Normdata) - colMeans(abnormdata)) %*% solve(pooled) %*%
↪  (colMeans(Normdata) - colMeans(abnormdata))
#T2
p <- dim(Normdata)[2]
#p
critical_value <- qf(0.95,p,n1+n2-p-1)*(n1+n2-2)*p/(n1+n2-p-1)
#critical_value
F_val <- (n1+n2-p-1)*T2/((n1+n2-2)*p)
#F_val

pval = 1 - pf(F_val,p,n1+n2-p-1)
#pval
```

Question 2 (d)

```r
library(ggpubr)
library(gridExtra)
textsize <- 26
p1 <- ggplot(data, aes(x = Class_Label, y = LumbarLordosisAngle, group = Class_Label,fill
↪  = Class_Label)) +
  geom_boxplot() + theme(legend.position="none")+ xlab(NULL)+ylab("Lumbar Lordosis
  ↪  Angle")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p2 <- ggplot(data, aes(x = Sacral_Slope, y = LumbarLordosisAngle, color = Class_Label,pch
↪  = Class_Label)) +
  geom_point(size = 3) + theme(legend.position="none")+ xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p3 <- ggplot(data, aes(x = Pelvic_Radius, y = LumbarLordosisAngle, color =
↪  Class_Label,pch = Class_Label)) +
```

```
  geom_point(size = 3) + theme(legend.position="none") + xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p4 <- ggplot(data, aes(x = LumbarLordosisAngle, y = Sacral_Slope, color = Class_Label,pch
↪  = Class_Label)) +
  geom_point(size = 3) + theme(legend.position="none") + xlab(NULL) + ylab("Sacral
  ↪  Slope")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p5 <- ggplot(data, aes(x = Class_Label, y = Sacral_Slope, group = Class_Label,fill =
↪  Class_Label)) +
  geom_boxplot() + theme(legend.position="none")+ xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p6 <- ggplot(data, aes(x = Pelvic_Radius, y = Sacral_Slope, color =Class_Label,pch =
↪  Class_Label)) +
  geom_point(size = 3)+ theme(legend.position="none")+ xlab(NULL)+ ylab( NULL)+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p7 <- ggplot(data, aes(x = LumbarLordosisAngle, y = Pelvic_Radius, color =Class_Label,pch
↪  = Class_Label)) +
  geom_point(size = 3) + theme(legend.position="none")+xlab("Lumbar Lordosis
  ↪  Angle")+ylab("Pelvic Radius")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p8 <-  ggplot(data, aes(x = Sacral_Slope, y = Pelvic_Radius, color = Class_Label,pch =
↪  Class_Label)) +
  geom_point(size = 3) + theme(legend.position="bottom")+ ylab( NULL)+xlab("Sacral
  ↪  Slope")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
```

```
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title

p9 <- ggplot(data, aes(x =Class_Label, y = Pelvic_Radius, group = Class_Label,fill =
↪  Class_Label)) +
  geom_boxplot() + theme(legend.position="none")+ ylab( NULL)+xlab("Pelvic Radius")+
  theme(text=element_text(size=textsize), #change font size of all text
        axis.title=element_text(size=textsize), #change font size of axis titles
        plot.title=element_text(size=textsize), #change font size of plot title
        legend.text=element_text(size=textsize), #change font size of legend text
        legend.title=element_text(size=textsize))#change font size of legend title


get_only_legend <- function(plot) {

# get tabular interpretation of plot
plot_table <- ggplot_gtable(ggplot_build(plot))

#  Mark only legend in plot
legend_plot <- which(sapply(plot_table$grobs, function(x) x$name) == "guide-box")

# extract legend
legend <- plot_table$grobs[[legend_plot]]

# return legend
return(legend)
}
# extract legend from plot1 using above function
legend <- get_only_legend(p8)
p8 <- p8+theme(legend.position="none")
g <- ggarrange(p1, p2, p3, p4,p5,p6,p7,p8,p9, ncol=3,nrow = 3)
```

```
ggarrange(g,legend,nrow = 2, heights = c(10, 1))
```

Question 2 (e)

```
transdata$Class_Label <- factor(transdata$Class_Label)
manova.3.classes<-manova(cbind(transdata$LumbarLordosisAngle, transdata$Sacral_Slope,
↪  transdata$Pelvic_Radius) ~ transdata$Class_Label)
#summary(manova.3.classes, test="Wilks")
#summary(manova.3.classes, test="Wilks")$SS

B <- summary(manova.3.classes, test="Wilks")$SS$`transdata$Class_Label`
W <- summary(manova.3.classes, test="Wilks")$SS$Residuals
t <- B+W

wilk <- det(W)/det(t)
#wilk
f <- ((dim(transdata)[1] - p -2)/(p))*((1 - sqrt(wilk))/(sqrt(wilk)))
```

```r
#f
crit <- qf(.95,2*p,2*(dim(transdata)[1] - p -2))
#crit
```

```
\captionof{table}{MANOVA Table}
\begin{tabular}{ c c c}
\toprule
\textbf{Source of Variation} & \textbf{SSCP Matrix} & \textbf{d.f.}\\
\midrule\\
\addlinespace[-2ex]
\text{Treatments} & $B = \begin{bmatrix}  24.42924 &  54.68592 & -215.0291 \\
54.68592  &124.80770& -389.4338 \\
-215.02914 &-389.43380& 5426.5048 \end{bmatrix}$ & $g-1  = 3 - 1 = 2$\\
\addlinespace[1.5ex]
\midrule
\addlinespace[1.5ex]
\text{Residuals} & $W = \begin{bmatrix}  30.48061   & 31.49275 &   38.72387 \\
31.49275 &  196.84666 &-1043.21468 \\
38.72387& -1043.21468 &49377.50484 \end{bmatrix}$ &
$\Sigma^g_{l = 1}n_l - g  =  310 - 3 = 307$\\

\addlinespace[1.5ex]
\midrule
\addlinespace[1.5ex]
\text{Total} & $B+W = \begin{bmatrix}  54.90986 &   86.17867 & -176.3053 \\
86.17867  & 321.65435 & -1432.6485 \\
-176.30527 & -1432.64848 & 54804.0096 \end{bmatrix}$ &
$\Sigma^g_{l = 1}n_l - 1  =  310 - 1 = 309$\\
\bottomrule
\end{tabular}
```

```r
NOdata <- subset(transdata,transdata$Class_Label == "NO")
#dim(NOdata)

DHdata <- subset(transdata,transdata$Class_Label == "DH")
#dim(DHdata)

SLdata <- subset(transdata,transdata$Class_Label == "SL")
#dim(SLdata)


mu <- colMeans(transdata[,1:3])
#mu

mu_NO <- colMeans(NOdata[,1:3])
#mu_NO

mu_DH <- colMeans(DHdata[,1:3])
#mu_DH
```

```
mu_SL <- colMeans(SLdata[,1:3])
#mu_SL
```

```
n <- dim(transdata)[1]
n_NO <- dim(NOdata)[1]
n_DH <- dim(DHdata)[1]
n_SL <- dim(SLdata)[1]
p <- 3
g <- 3
alpha <- 0.05

#Lumbar Lordosis Angle

#CI for Lumbar Lordosis Angle between groups NO and DH
CI.LumbarLordosisAngle.grpNODH<-paste("( ",

  ↪  round(mu_NO[1]-mu_DH[1]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[1,1]/(n-g)*(1/n_NO+1/n_DH)),4),
  ↪
  ", ",

  ↪  round(mu_NO[1]-mu_DH[1]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[1,1]/(n-g)*(1/n_NO+1/n_DH)),4),
  ")")
#CI.LumbarLordosisAngle.grpNODH

#CI for Lumbar Lordosis Angle between groups NO and SL
CI.LumbarLordosisAngle.grpNOSL<-paste("(",

  ↪  round(mu_NO[1]-mu_SL[1]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[1,1]/(n-g)*(1/n_NO+1/n_SL)),4),
  ↪
  ",",

  ↪  round(mu_NO[1]-mu_SL[1]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[1,1]/(n-g)*(1/n_NO+1/n_SL)),4),
  " )")
#CI.LumbarLordosisAngle.grpNOSL

#CI for Lumbar Lordosis Angle between groups DH and SL
CI.LumbarLordosisAngle.grpDHSL<-paste("(",

  ↪  round(mu_DH[1]-mu_SL[1]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[1,1]/(n-g)*(1/n_DH+1/n_SL)),4),
  ↪
  ",",

  ↪  round(mu_DH[1]-mu_SL[1]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[1,1]/(n-g)*(1/n_DH+1/n_SL)),4),
  " )")
#CI.LumbarLordosisAngle.grpDHSL


# Sacral_Slope
```

```
#CI for Sacral_Slope between groups NO and DH
CI.Sacral_Slope.grpNODH<-paste("( ",
  ↪ round(mu_NO[2]-mu_DH[2]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[2,2]/(n-g)*(1/n_NO+1/n_DH)),4),
  ↪
  " , ",
  ↪ round(mu_NO[2]-mu_DH[2]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[2,2]/(n-g)*(1/n_NO+1/n_DH)),4),
  ")")
#CI.Sacral_Slope.grpNODH

#CI for Sacral_Slope between groups NO and SL
CI.Sacral_Slope.grpNOSL<-paste("(",
  ↪ round(mu_NO[2]-mu_SL[2]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[2,2]/(n-g)*(1/n_NO+1/n_SL)),4),
  ",",
  ↪ round(mu_NO[2]-mu_SL[2]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[2,2]/(n-g)*(1/n_NO+1/n_SL)),4),
  " )")
#CI.Sacral_Slope.grpNOSL

#CI for Sacral_Slope between groups DH and SL
CI.Sacral_Slope.grpDHSL<-paste("(",
  ↪ round(mu_DH[2]-mu_SL[2]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[2,2]/(n-g)*(1/n_DH+1/n_SL)),4),
  ",",
  ↪ round(mu_DH[2]-mu_SL[2]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[2,2]/(n-g)*(1/n_DH+1/n_SL)),4),
  " )")
#CI.Sacral_Slope.grpDHSL



# Pelvic_Radius


#CI for Pelvic_Radius between groups NO and DH
CI.Pelvic_Radius.grpNODH<-paste("(",
  ↪ round(mu_NO[3]-mu_DH[3]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[3,3]/(n-g)*(1/n_NO+1/n_DH)),4),
  ↪
        ",",
  ↪ round(mu_NO[3]-mu_DH[3]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[3,3]/(n-g)*(1/n_NO+1/n_DH)),4),
        ")")
#CI.Pelvic_Radius.grpNODH

#CI for Pelvic_Radius between groups NO and SL
CI.Pelvic_Radius.grpNOSL<-paste("(",
  ↪ round(mu_NO[3]-mu_SL[3]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[3,3]/(n-g)*(1/n_NO+1/n_SL)),4),
```

```
  ",",
  ↪  round(mu_NO[3]-mu_SL[3]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[3,3]/(n-g)*(1/n_NO+1/n_SL)),4),
      ")")
#CI.Pelvic_Radius.grpNOSL

#CI for Pelvic_Radius between groups DH and SL
CI.Pelvic_Radius.grpDHSL<-paste("(",
  ↪  round(mu_DH[3]-mu_SL[3]-qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[3,3]/(n-g)*(1/n_DH+1/n_SL)),4),
  ",",
  ↪  round(mu_DH[3]-mu_SL[3]+qt(1-alpha/(p*g*(g-1)),n-g)*sqrt(W[3,3]/(n-g)*(1/n_DH+1/n_SL)),4),
      ")")
#CI.Pelvic_Radius.grpDHSL


# table creation

tab <- cbind(rbind(CI.LumbarLordosisAngle.grpNODH, CI.LumbarLordosisAngle.grpNOSL,
↪  CI.LumbarLordosisAngle.grpDHSL),
             rbind(CI.Sacral_Slope.grpNODH, CI.Sacral_Slope.grpNOSL,
             ↪  CI.Sacral_Slope.grpDHSL),
             rbind(CI.Pelvic_Radius.grpNODH, CI.Pelvic_Radius.grpNOSL,
             ↪  CI.Pelvic_Radius.grpDHSL))
row.names(tab) <- c("NO - DH","NO - SL","DH - SL")
#col.names(tab) <- c("Lumbar Lordosis Angle","Sacral Slope","Pelvic_Radius")
#tab
library(kableExtra)
knitr::kable(tab,booktabs = T,col.names = c("Lumbar Lordosis Angle","Sacral
↪  Slope","Pelvic Radius"),caption = "$95\\%$ Simultaneous Bonferroni CI for diff of
↪  means component", align = "ccc") %>%
  kable_styling(latex_options = c("hold_position"))
```