Project 2

Analysis A: Bone Data

A) Checking for MVN:

Multivariate Normality (Dominant):

Test	H	P value	MVN
Royston	7.22671	0.06358	YES

Table1 MVN for dominant

chi-square Q-Q plot of Dominant bones
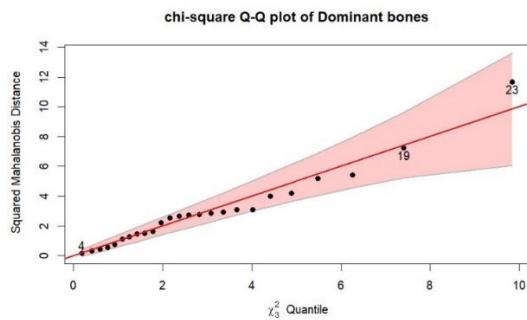
Multivariate Normality (Non-Dominant):

Test	H	P value	MVN
Royston	2.00584	0.52766	YES

Table2 MVN for dominant

chi-square Q-Q plot of Non Dominant bones
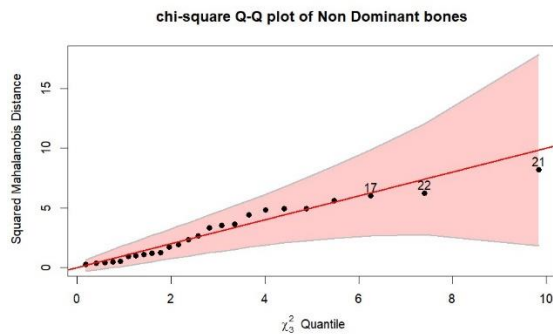
Hotelling T2 test:

data: cat1 and cat2(Dominant & Non dominant)

Test	t2	df1	df2	P-value
Hotellingt2	0.29523	3	46	0.8286
Chi sq	0.92421	3		0.8196

Table3 HT test results

B) CI:

	Cis-Hotelling		Cis-Chi-Sq		Cis-Bonferroni	
	LCBS	**UCBS**	**LCBS**	**UCBS**	**LCBS**	**UCBS**
Radius	-0.067	0.118	-0.062	0.113	-0.052	0.103
Humerous	-0.172	0.287	-0.159	0.274	-0.134	0.25
Ulna	-0.078	0.099	-0.073	0.094	-0.063	0.084

Table4 Confidence intervals for dominant & non dominant

C) Report for Analysis A: Bone Data

The Chi-square Q-Q plots and Royston test suggest the data is multivariate normal for both dominant and non-dominant bone groups.

The Hoteling's T2 test and Chi-square test both yield high p-values (0.8286 and 0.8196 respectively), indicating no significant difference in the mineral content of the dominant vs non-dominant bones overall.

The 95% simultaneous confidence intervals for the difference in mineral content between dominant and non-dominant bones all contain 0, further confirming no significant differences for any of the individual bones (radius, humerus, ulna).
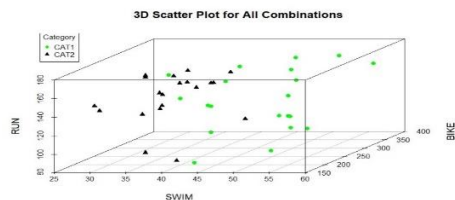
I agree with both of the test & CI. Because the interval results and hoteling t2 test concludes that dominant and non-dominant bones do not differ significantly in mineral content.

B1 analysis: Triathlon Data

A) Bivariate between two age categories
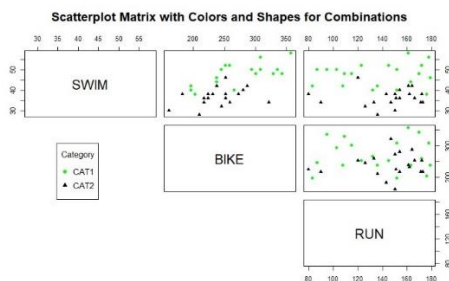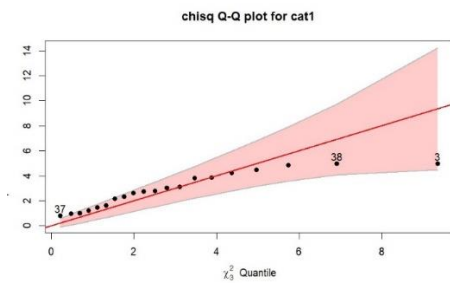

3D graph between two age categories

B) Checking for MVN:

Multivariate Normality (Cat1):

Test	H	P value	MVN
Royston	4.094535	0.253542	YES

Table5 MVN for cat1

chisq Q-Q plot for cat1

Multivariate Normality (Cat2):

Test	H	P value	MVN
Royston	7.202779	0.066244	YES

Table6 MVN for cat2

chisq Q-Q plot for cat2
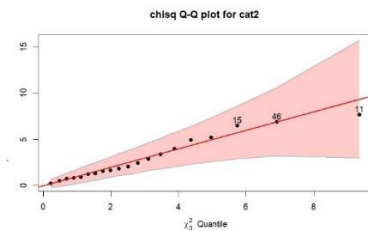
Hotelling t2 test:

Test	t2	df1	df2	P-value
Hotellingt2	17.81	3	36	2.96E-07
Chi-sq	56.399	3		3.45E-12

Table7 HT for cat1 & cat2

C) CI:

Categories	swim		bike		Run	
	LCBS	**UCBS**	**LCBS**	**UCBS**	**LCBS**	**UCBS**
Cat1 & cat2	7.108	14.892	-2.55	66.25	-31.172	14.772

Table8 CI for cat1 & cat2

D) Report:

The Chi-square Q-Q plots, and Royston test suggest the data is Fairly considered as multivariate normal, with minor deviation for category 2 run times.

The Hoteling's T2 test and Chi-square test both yield very low p-values (2.96e-07 and 3.45e-12 respectively), indicating a highly significant difference in performance between age categories 1 and 2 overall.

The 95% simultaneous confidence intervals for the difference in performance between categories 1 and 2 exclude 0 for swim times, but include 0 for bike and run times. This suggests categories differ significantly on swim performance, but not conclusively for bike and run individually.
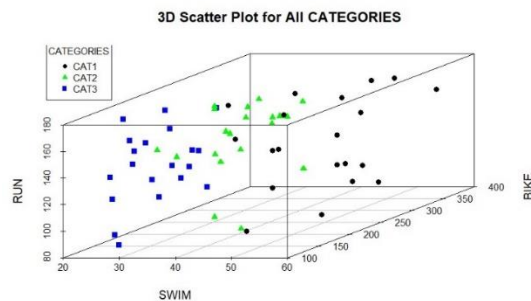
Conclusion: I agree with the both tests & CI. Because categories 1 and 2 differ significantly in overall performance, primarily driven by swim times.

Analysis B2: Trithalon Data

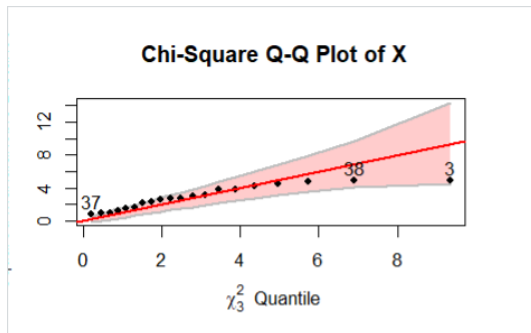### A) Bivariate between three age categories
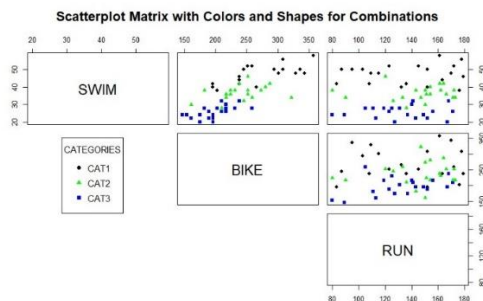
3D graph between two age categories

B) Checking for MVN:

Multivariate Normality (Cat1):

Test	H	P value	MVN
Royston	4.0945	0.2535	YES

Table9 MVN for cat1

Chi Sq Q-Q plot for cat1:

Multivariate Normality (Cat2):

Test	H	P value	MVN
Royston	7.2028	0.0662	YES

Table10 MVN for cat2

Chi Sq Q-Q plot for cat2:
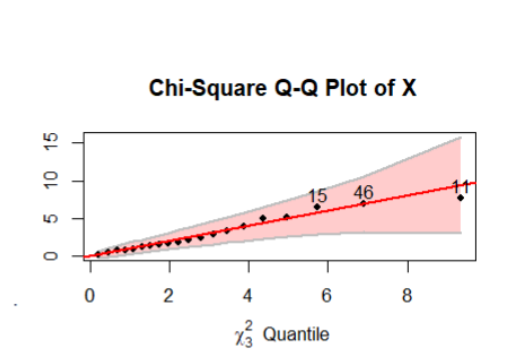
Multivariate Normality (Cat3):

Test	H	P value	MVN
Royston	0.9462	0.8218	YES

Table11 MVN for cat3

Chi Sq Q-Q plot for cat3:

Chi-Square Q-Q Plot of X
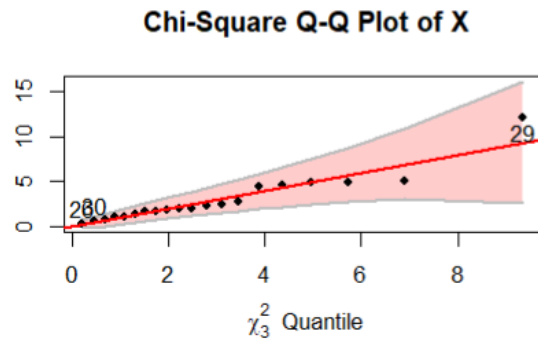
Manova Test Result:

categories	Wilks approx	F	num DF	den DF	Pr(>F)
CAT1 - CAT2	0.46279	21.281	3	55	8.407e-09
CAT1 - CAT3	0.18891	78.713	3	55	< 2.2e-16
CAT2 - CAT3	0.48994	19.086	3	55	3.935e-08

C) CI:

Categories	swim		bike		Run	
	LCBS	**UCBS**	**LCBS**	**UCBS**	**LCBS**	**UCBS**
cat1 & cat2	7.108	14.892	-2.55	66.25	-31.172	14.772
cat1 & cat3	17.58	25.81	36.16	107.33	-20.71	29.91
cat2 & cat3	6.58	14.81	4.31	75.48	-12.51	38.11

Table12 CIs for cat 1,2&3

D) Report:

The chi-sq Q-Q plot & Royston test gives p-values high than considered 0.05 it shows multivariate normality for all age categories.

The MANOVA test indicates highly significant differences between the age categories overall by checking the P-values.

The simultaneous 95% confidence intervals for the difference in performance between age categories indicate significant differences in performance across all three sports.

Conclusion: I agree with the both tests & CI. Because all age categories exhibit significant differences in overall performance across all three sports, indicating that age plays a role in athletic performance in this triathlon dataset.

R Codes:

```
library(readxl)

library(mvtnorm)

library(MVN)
```

```r
library(heplots)

source("projectfile.R")

BonesData = read_excel("BonesData.xlsx")

TriathlonData = read_excel("TriathlonData.xlsx")

# 3 variables in Bones Data

D  = data.frame(BonesData$DomRadius, BonesData$DomHumerus , BonesData$DomUlna )

ND = data.frame(BonesData$NonDomRadius , BonesData$NonDomHumerus ,BonesData$NonDomUlna )

library(DescTools)

HotellingsT2Test(D,ND) # Hotelling's test

HotellingsT2Test(D,ND, test = 'chi') # ChiSq test

cqplot(D, id.n=3,main="chi-square Q-Q plot of Dominant bones") # ChiSq QQPlot for Dominant Bones

mvn(D, mvnTest = "royston")

mvn(D, mvnTest = "hz")

cqplot(ND, id.n=3,main="chi-square Q-Q plot of Non Dominant bones") # ChiSq QQPlot for Non Dominant Bones

mvn(ND, mvnTest = "royston")

mvn(ND, mvnTest = "hz")

MVN2Sample.HT.CIs.f(D, ND,conf.level = 0.95,alpha=.05,mu0=rep(0,ncol(D)),

        ContrastMAT=NULL,SigDig=3,var.eq=TRUE)


#B1 analysis: Triathlon Data

df <- subset(TriathlonData,TriathlonData$CATEGORY != "CAT3")

combinations_colors <- c("CAT1" = "green", "CAT2" = "black")

combinations_shapes <- c("CAT1" = 19, "CAT2" = 17)


# Scatterplot Matrix with Colors and Shapes for Combinations

pairs(~ SWIM + BIKE + RUN, data = df,

    col = combinations_colors[df$CATEGORY],

    pch = combinations_shapes[df$CATEGORY],

    main = "Scatterplot Matrix with Colors and Shapes for Combinations",

    lower.panel = NULL)


# Add legend for combinations

legend("bottomleft", legend = unique(df$CATEGORY), title = "Category",

    col = combinations_colors, pch = combinations_shapes,inset = c(0.12, 0.32), cex = 0.8)

library(scatterplot3d)
```

```r
new_data <- data.frame(
  Combination = df$CATEGORY,
  SWIM = df$SWIM,
  BIKE = df$BIKE,
  RUN = df$RUN
)
scatterplot3d(new_data$SWIM, new_data$BIKE, new_data$RUN,
        color = combinations_colors[new_data$Combination], # Color for all points
        pch = combinations_shapes[new_data$Combination],      # Shape for all points
        xlab = "SWIM",
        ylab = "BIKE",
        zlab = "RUN",
        main = "3D Scatter Plot for All Combinations")
legend("topleft",
     legend = unique(new_data$Combination), # Unique combinations
     col = combinations_colors,      # Colors corresponding to combinations
     pch = combinations_shapes,      # Shapes corresponding to combinations
     title = "Category",
     inset=c(0.001,-0.01),
 cex = 0.8)


cat1 = subset(data.frame(TriathlonData$SWIM, TriathlonData$BIKE, TriathlonData$RUN),TriathlonData$CATEGORY == "CAT1")
cat2 = subset(data.frame(TriathlonData$SWIM, TriathlonData$BIKE, TriathlonData$RUN),TriathlonData$CATEGORY == "CAT2")
cat3 = subset(data.frame(TriathlonData$SWIM, TriathlonData$BIKE, TriathlonData$RUN),TriathlonData$CATEGORY == "CAT3")


library(DescTools)
HotellingsT2Test(cat1,cat2) # Hotelling's test
HotellingsT2Test(cat1,cat2, test = 'chi') # ChiSq test


cqplot(cat1, id.n=3,main="chisq Q-Q plot for cat1") # ChiSq QQPlot for cat1
mvn(cat1, mvnTest = "royston")
mvn(cat1, mvnTest = "hz")


cqplot(cat2, id.n=3,main="chisq Q-Q plot for cat2") # ChiSq QQPlot for cat2
mvn(cat2, mvnTest = "royston")
```

```
mvn(cat2, mvnTest = "hz")


MVN2Sample.HT.CIs.f(cat1,cat2,conf.level = 0.95,alpha=.05,mu0=rep(0,ncol(cat1)),

           ContrastMAT=NULL,SigDig=3,var.eq=TRUE)
```

 # B2 Analysis:

```
df <- TriathlonData

combinations_colors <- c("CAT1" = "black", "CAT2" = "green","CAT3" = "blue")

combinations_shapes <- c("CAT1" = 19, "CAT2" = 17, "CAT3" = 15)


# Scatterplot Matrix with Colors and Shapes for Combinations

pairs(~ SWIM + BIKE + RUN, data = df,

    col = combinations_colors[df$CATEGORY],

    pch = combinations_shapes[df$CATEGORY],

    main = "Scatterplot Matrix with Colors and Shapes for Combinations",

    lower.panel = NULL)


# Add legend for combinations

legend("bottomleft", legend = unique(df$CATEGORY), title = "CATEGORIES",

     col = combinations_colors, pch = combinations_shapes,inset = c(0.12, 0.32), cex = 0.8)


library(scatterplot3d)

new_data <- data.frame(

 Combination = df$CATEGORY,

 SWIM = df$SWIM,

 BIKE = df$BIKE,

 RUN = df$RUN

)


scatterplot3d(new_data$SWIM, new_data$BIKE, new_data$RUN,

       color = combinations_colors[new_data$Combination], # Color for all points

       pch = combinations_shapes[new_data$Combination],     # Shape for all points

       xlab = "SWIM",

       ylab = "BIKE",

       zlab = "RUN",
```

```r
            main = "3D Scatter Plot for All CATEGORIES")


legend("topleft",

    legend = unique(df$CATEGORY), # Unique combinations

    col = combinations_colors,      # Colors corresponding to combinations

    pch = combinations_shapes,      # Shapes corresponding to combinations

    title = "CATEGORIES",

    inset=c(0.001,-0.01),

    cex = 0.8)


# MANOVA with car package

library(car)

# Computing Means and Var/Cov Mats per groups

VMeans = statList(new_data[,-1],new_data[,1],FUN = colMeans)

VMat = statList(new_data[,-1],new_data[,1],FUN = var)

Ns = table(new_data[,1]) # Sample sizes

p = 3

g = 3

n = nrow(new_data)

# Manually computing the W matrix

W = (Ns[1]-1)*VMat[[1]] + (Ns[2]-1)*VMat[[2]] + (Ns[3]-1)*VMat[[3]]

library(heplots)

library(MVN)

# Split data by group

new_data_gr = split(new_data,new_data[,1])

# Checking Multivariate normality for each group

par(mfrow = c(2,2))

for(i in 1:g){

 X = new_data_gr[[i]][,-1]

 cqplot(X, id.n=3)

 print(mvn(X, mvnTest = "royston"))

 print(mvn(X, mvnTest = "hz"))

}
```