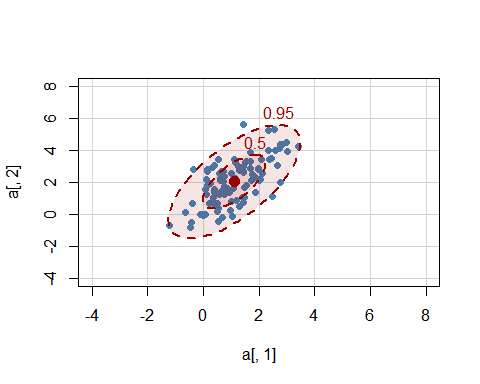
Question 3,4

Wed Jan 23 2019

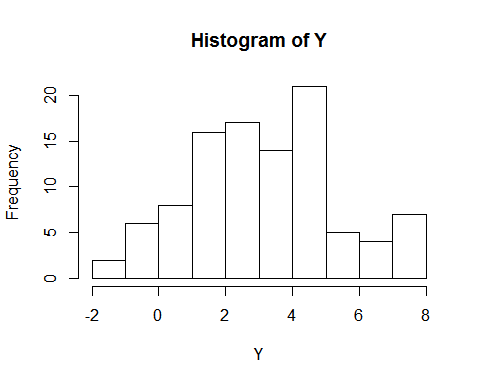
####Question 3#######  
library(mnormt)

var\_=cbind(c(1,1),c(1,2))  
a=rmnorm(100,mean=c(1,2),varcov=var\_,sqrt=NULL)  
library(car)

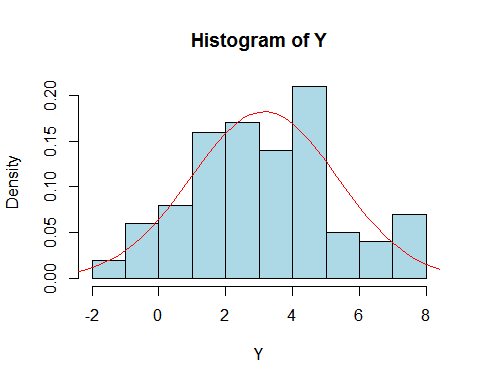
dataEllipse(a[,1], a[,2],   
 xlim = c(-4, 8), ylim = c(-4, 8),   
 pch=19, col = c("steelblue", "#990000"), lty=2,  
 ellipse.label=c(0.5, 0.95), levels = c(0.5, 0.95),  
 fill=TRUE, fill.alpha=0.1)



Y<- a[,1]+a[,2]  
hist(Y)



m<-mean(Y)  
v<- var(Y)  
#3d  
std<-sqrt(v)  
hist(Y,col="light blue",main="Histogram of Y",freq=F)  
curve(dnorm(x,mean=m,sd=std),-5,20, add=T,col= "red")



Question 2

Nadim

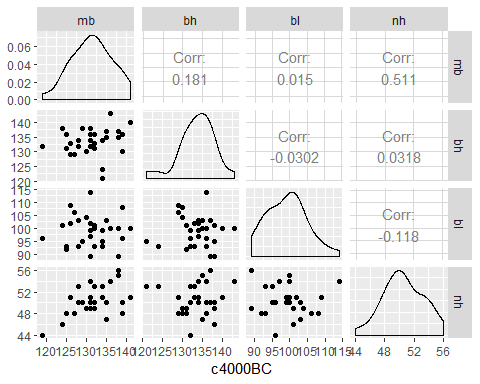
Wed Jan 23 17:46:37 2019

library(HSAUR3)

library(ggplot2)

library(GGally)

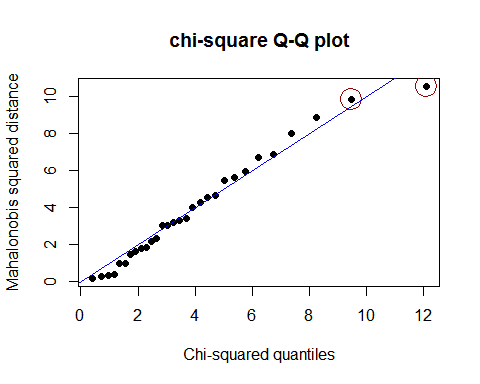
epochs<-skulls$epoch  
skull\_cov<- skulls[epochs=="c4000BC",2:5]  
ggpairs(skull\_cov,xlab="c4000BC")



chisquare.plot <- function(x,mark){  
 p <- ncol(x)  
 n<-nrow(x)  
 xbar<- colMeans(x)  
 s <- cov(x)  
   
 ###Mahalonobis Dist  
   
 x.cent <- scale(x,center =T, scale = F)  
 d2<- diag(x.cent%\*%solve(s)%\*%t(x.cent))  
   
 qchi <- qchisq((1:n-0.5)/n, df =p)  
 sortd <- sort(d2)  
   
 plot(qchi,sortd,pch=19,xlab="Chi-squared quantiles",ylab="Mahalonobis squared distance",main="chi-square Q-Q plot")  
   
 points(qchi[(n-mark+1):n],sortd[(n-mark+1):n],cex=3,col="#990000")  
   
 return((sortd[(n-mark+1):n]))  
}  
  
chisquare.plot(x=skull\_cov,mark = 2)

## 12 29   
## 9.861683 10.573099

abline(0,1,col="blue")

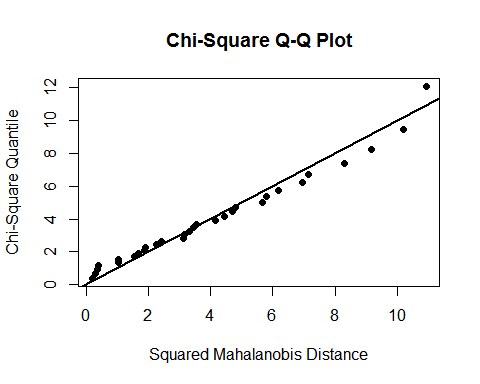


# epochs<-skulls$epoch  
# skull\_cov<- skulls[epochs=="c4000BC",2:5]  
  
z\_col<- list()  
#z\_col<-matrix(list(),nrow=length(skull\_cov$mb),ncol=4)  
for (i in 1:4){  
 z\_col[i]<- list((scale(skull\_cov[i])))  
 skull\_cov[i+4]<- z\_col[i]  
 #skull\_cov<-cbind(z\_col[-i],skull\_cov)  
  
}  
  
colnames(skull\_cov)<-c("mb","bh","bl","nh","Z\_mb","z\_bh","z\_bl","z\_nh")  
sk<-skull\_cov[,1:4]  
s<-cov(sk)  
x.cent <- scale(sk,center =T, scale = F)  
d2<- diag(x.cent%\*%solve(s)%\*%t(x.cent))  
skull\_cov<-cbind(d2=d2,skull\_cov)  
  
apply(skull\_cov[,2:5], 2, shapiro.test)

## $mb  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.98136, p-value = 0.8603  
##   
##   
## $bh  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.95664, p-value = 0.2536  
##   
##   
## $bl  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.97314, p-value = 0.6282  
##   
##   
## $nh  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.97481, p-value = 0.6772

library(MVN)

mvn(skull\_cov[, 2:5], mvnTest = "royston", multivariatePlot = "qq")



## $multivariateNormality  
## Test H p value MVN  
## 1 Royston 2.752767 0.603866 YES  
##   
## $univariateNormality  
## Test Variable Statistic p value Normality  
## 1 Shapiro-Wilk mb 0.9814 0.8603 YES   
## 2 Shapiro-Wilk bh 0.9566 0.2536 YES   
## 3 Shapiro-Wilk bl 0.9731 0.6282 YES   
## 4 Shapiro-Wilk nh 0.9748 0.6772 YES   
##   
## $Descriptives  
## n Mean Std.Dev Median Min Max 25th 75th Skew  
## mb 30 131.36667 5.129249 131 119 141 128.00 134.75 -0.16642216  
## bh 30 133.60000 4.469051 134 121 143 131.25 136.00 -0.64720446  
## bl 30 99.16667 5.884423 100 89 114 95.00 102.75 0.31717217  
## nh 30 50.53333 2.763473 50 44 56 49.00 53.00 -0.08670975  
## Kurtosis  
## mb -0.4879548  
## bh 0.8488047  
## bl -0.2756768  
## nh -0.4538837

###Q4#### -- part b  
library(mnormt)  
var\_cov = cbind(c(2,1,1),c(1,1,1),c(1,1,3))  
r\_new <- rmnorm(100,mean=c(1,2,3),varcov=var\_cov,sqrt=NULL)  
cor(r\_new)

## [,1] [,2] [,3]  
## [1,] 1.0000000 0.7565691 0.3970326  
## [2,] 0.7565691 1.0000000 0.5299395  
## [3,] 0.3970326 0.5299395 1.0000000

scaled <- scale(r\_new)   
cov(scaled)

## [,1] [,2] [,3]  
## [1,] 1.0000000 0.7565691 0.3970326  
## [2,] 0.7565691 1.0000000 0.5299395  
## [3,] 0.3970326 0.5299395 1.0000000

Question 1.R

Nadim

Wed Jan 23 22:40:28 2019

dat<- read.table("https://www.stat.ncsu.edu/people/maity/courses/st537-S2019/data/T4-3.DAT", header=F)  
colnames(dat) <- c("x1", "x2", "x3", "x4", "d2")  
n <- nrow(dat)  
p <- ncol(dat)  
head(dat)

## x1 x2 x3 x4 d2  
## 1 1889 1651 1561 1778 0.60  
## 2 2403 2048 2087 2197 5.48  
## 3 2119 1700 1815 2222 7.62  
## 4 1645 1627 1110 1533 5.21  
## 5 1976 1916 1614 1883 1.40  
## 6 1712 1712 1439 1546 2.22

######Question 1#######  
i<-which(dat$d2==max(dat$d2))  
dat <- dat[-c(i),]  
j<-which(dat$d2==max(dat$d2))  
dat<- dat[-c(j),]

max(dat$d2)

## [1] 9.9

apply(dat[,1:4], 2, shapiro.test)

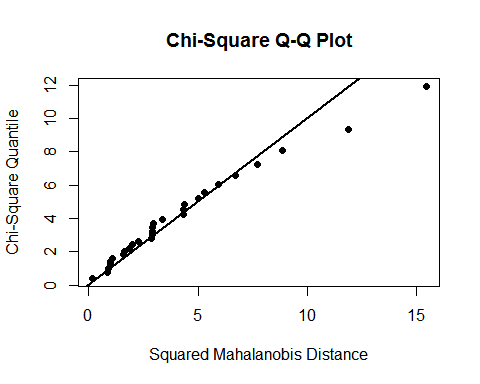
## $x1  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.98469, p-value = 0.9439  
##   
##   
## $x2  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.96636, p-value = 0.4871  
##   
##   
## $x3  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.96717, p-value = 0.507  
##   
##   
## $x4  
##   
## Shapiro-Wilk normality test  
##   
## data: newX[, i]  
## W = 0.96598, p-value = 0.4779

library(MVN)

## Warning: package 'MVN' was built under R version 3.5.2

## sROC 0.1-2 loaded

mvn(dat[, 1:4], mvnTest = "royston", multivariatePlot = "qq")



## $multivariateNormality  
## Test H p value MVN  
## 1 Royston 1.098338 0.6271166 YES  
##   
## $univariateNormality  
## Test Variable Statistic p value Normality  
## 1 Shapiro-Wilk x1 0.9847 0.9439 YES   
## 2 Shapiro-Wilk x2 0.9664 0.4871 YES   
## 3 Shapiro-Wilk x3 0.9672 0.5070 YES   
## 4 Shapiro-Wilk x4 0.9660 0.4779 YES   
##   
## $Descriptives  
## n Mean Std.Dev Median Min Max 25th 75th Skew  
## x1 28 1865.929 262.1619 1857.5 1325 2403 1711.50 2049.75 0.08994538  
## x2 28 1697.964 244.8618 1663.0 1170 2301 1593.25 1847.50 0.39091767  
## x3 28 1488.643 253.1536 1466.0 1002 2087 1307.25 1617.25 0.49661284  
## x4 28 1710.250 277.9986 1674.5 1176 2234 1528.75 1876.25 0.25921958  
## Kurtosis  
## x1 -0.5084972  
## x2 0.1961808  
## x3 0.0516768  
## x4 -0.6484407

chisquare.plot <- function(x,mark){  
 p <- ncol(x)  
 n<-nrow(x)  
 xbar<- colMeans(x)  
 s <- cov(x)  
   
 ###Mahalonobis Dist  
   
 x.cent <- scale(x,center =T, scale = F)  
 d2<- diag(x.cent%\*%solve(s)%\*%t(x.cent))  
   
 qchi <- qchisq((1:n-0.5)/n, df =p)  
 sortd <- sort(d2)  
   
 plot(qchi,sortd,pch=19,xlab="Chi-squared quantiles",ylab="Mahalonobis squared distance",main="chi-square Q-Q plot")  
   
 points(qchi[(n-mark+1):n],sortd[(n-mark+1):n],cex=3,col="#990000")  
   
 return((sortd[(n-mark+1):n]))  
 }  
  
 a<-chisquare.plot(x=dat[,1:4],mark = 2)  
 a[1]

## 21   
## 11.45119

a[2]

## 3   
## 14.89312

# plot(dat$x1,dat$x2 ,  
 # xlim = c(1000, 3500), ylim = c(800, 3000),   
 # pch=19, col = c("steelblue"))  
 abline(0,1,col="blue")

