question\_1.R

Nadim

Wed Feb 13 22:21:07 2019

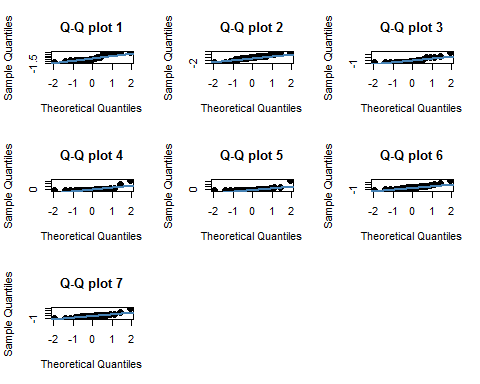
## Question 1 ##   
## part a ##  
dat<- read.table("hemangioma.txt",header = TRUE)  
scaled\_dat <- scale(dat[, -1], scale = TRUE)  
  
par(mfrow=c(3,3))  
for(i in 1:7){  
 qqnorm(scaled\_dat[,i],main = paste0("Q-Q plot ",i),pch=19,cex=1.5)  
 qqline(scaled\_dat, col = "steelblue", lwd = 2)  
}  
  
## Creating data ellipse to detect outliers ##  
 library(car)

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

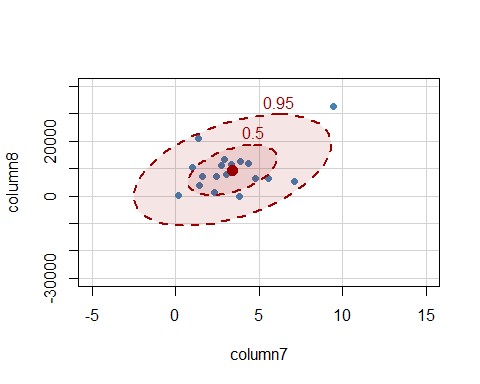
## Loading required package: carData

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

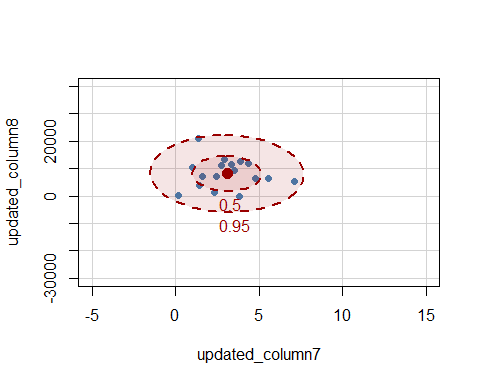
column7 <- dat[,7]  
 column8 <- dat[,8]  
 par(mfrow=c(1,1))



dataEllipse(column7, column8,   
 xlim = c(-5,15), ylim = c(-30000, 40000),   
 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)



# Data point 11 seems to be the clear outlier   
# Therefore, removing that data point from the original data   
updated\_dat <- dat[-11,]  
updated\_column7 <- updated\_dat[,7]  
updated\_column8 <- updated\_dat[,8]  
par(mfrow=c(1,1))  
dataEllipse(updated\_column7, updated\_column8,   
 xlim = c(-5,15), ylim = c(-30000, 40000),   
 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)



# Performing EFA  
EFA\_dat <- factanal(scaled\_dat,factors=3)  
EFA\_dat

##   
## Call:  
## factanal(x = scaled\_dat, factors = 3)  
##   
## Uniquenesses:  
## RB p16 DLK Nanog C.Myc EZH2 IGF.2   
## 0.050 0.293 0.005 0.609 0.005 0.490 0.249   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## RB 0.141 -0.144 0.954   
## p16 0.366 0.757   
## DLK -0.163 0.961 -0.211   
## Nanog 0.559 0.275   
## C.Myc 0.841 0.295 -0.448   
## EZH2 0.682 0.193   
## IGF.2 0.780 0.377   
##   
## Factor1 Factor2 Factor3  
## SS loadings 2.274 1.757 1.269  
## Proportion Var 0.325 0.251 0.181  
## Cumulative Var 0.325 0.576 0.757  
##   
## Test of the hypothesis that 3 factors are sufficient.  
## The chi square statistic is 1.86 on 3 degrees of freedom.  
## The p-value is 0.603

EFA\_updated\_dat <- factanal(updated\_dat,factors=3)  
EFA\_updated\_dat

##   
## Call:  
## factanal(x = updated\_dat, factors = 3)  
##   
## Uniquenesses:  
## Age RB p16 DLK Nanog C.Myc EZH2 IGF.2   
## 0.412 0.051 0.164 0.222 0.703 0.353 0.421 0.197   
##   
## Loadings:  
## Factor1 Factor2 Factor3  
## Age 0.765   
## RB -0.168 0.927 -0.249   
## p16 0.896 -0.163   
## DLK 0.845 -0.235   
## Nanog 0.541   
## C.Myc 0.766 -0.233   
## EZH2 -0.126 0.447 -0.603   
## IGF.2 0.613 0.541 0.366   
##   
## Factor1 Factor2 Factor3  
## SS loadings 2.530 1.760 1.186  
## Proportion Var 0.316 0.220 0.148  
## Cumulative Var 0.316 0.536 0.685  
##   
## Test of the hypothesis that 3 factors are sufficient.  
## The chi square statistic is 2.02 on 7 degrees of freedom.  
## The p-value is 0.959