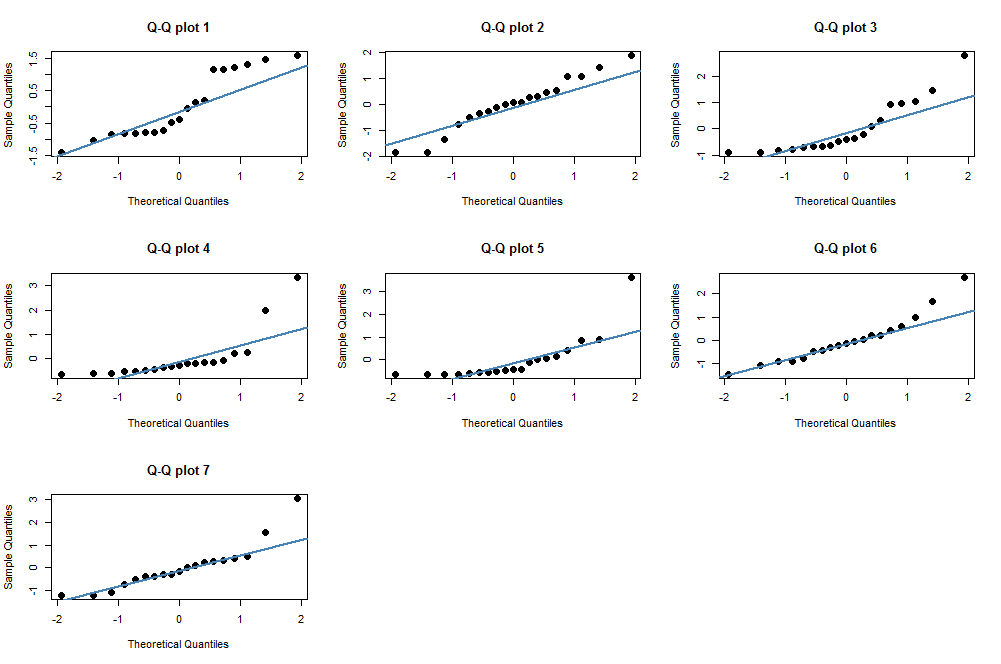
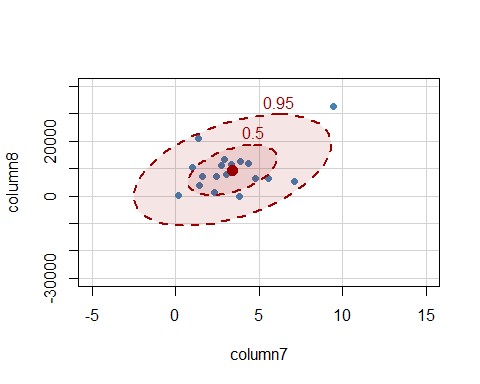
## Question 1 ##   
## part a ##  
dat<- read.table("hemangioma.txt",header = TRUE)  
scaled\_dat <- scale(dat[, -1], scale = TRUE)  
  
par(mfrow=c(3,3))  
for(i in 1:7){  
 qqnorm(scaled\_dat[,i],main = paste0("Q-Q plot ",i),pch=19,cex=1.5)  
 qqline(scaled\_dat, col = "steelblue", lwd = 2)  
}  
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

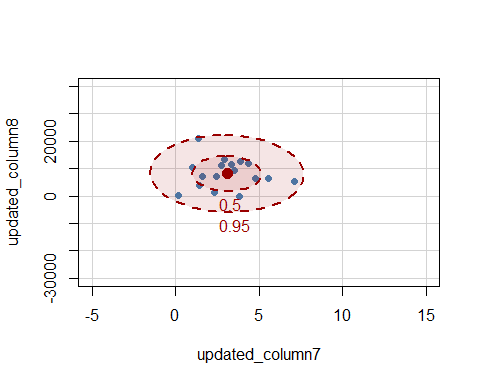
column7 <- dat[,7]  
 column8 <- dat[,8]  
 par(mfrow=c(1,1))



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

We can see a clear improvement in the model fit as the p-value has increased to 0.959 approx. from 0.60 in the EFA of the raw data

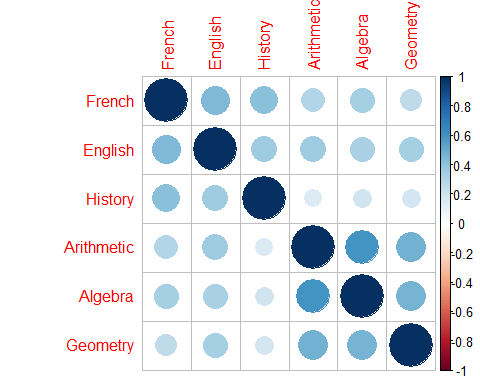
Question 2

library(sem)

lt <- readMoments("EverittEx5.5.txt", diag = T)  
R <- (lt + t(lt)) - diag(1, 6)  
colnames(R) <- c("French", "English", "History", "Arithmetic", "Algebra", "Geometry")  
rownames(R) <- c("French", "English", "History", "Arithmetic", "Algebra", "Geometry")  
  
library(corrplot)

library(psych)

corrplot(R)



#model fit-1  
model.1<-fa(r=R,nfactors = 2,n.obs = 220,fm = "ml",rotate = "none")  
model.1$loadings

##   
## Loadings:  
## ML1 ML2   
## French 0.558 0.425  
## English 0.569 0.286  
## History 0.392 0.450  
## Arithmetic 0.738 -0.279  
## Algebra 0.718 -0.209  
## Geometry 0.595 -0.133  
##   
## ML1 ML2  
## SS loadings 2.204 0.603  
## Proportion Var 0.367 0.101  
## Cumulative Var 0.367 0.468

###**We can see that factors here are not very much interpretable**#model fit-2  
model.2<-fa(r=R,nfactors = 2,n.obs = 220,fm = "ml",rotate = "varimax")  
model.2$loadings

##   
## Loadings:  
## ML1 ML2   
## French 0.233 0.661  
## English 0.319 0.551  
## History 0.591  
## Arithmetic 0.770 0.172  
## Algebra 0.715 0.220  
## Geometry 0.570 0.215  
##   
## ML1 ML2  
## SS loadings 1.593 1.215  
## Proportion Var 0.265 0.202  
## Cumulative Var 0.265 0.468

**####after rotation of the factors we can make clear interpretations that  
####the first factor focuses more on the arthmetic abilities and the second** **factor  
#### focusses more on the linguistic abilities**  
factanal(covmat =R,factors = 2,n.obs = 220)#R:

##   
## Call:  
## factanal(factors = 2, covmat = R, n.obs = 220)  
##   
## Uniquenesses:  
## French English History Arithmetic Algebra Geometry   
## 0.508 0.595 0.644 0.377 0.440 0.628   
##   
## Loadings:  
## Factor1 Factor2  
## French 0.233 0.661   
## English 0.319 0.551   
## History 0.591   
## Arithmetic 0.770 0.172   
## Algebra 0.715 0.220   
## Geometry 0.570 0.215   
##   
## Factor1 Factor2  
## SS loadings 1.593 1.215  
## Proportion Var 0.265 0.202  
## Cumulative Var 0.265 0.468  
##   
## Test of the hypothesis that 2 factors are sufficient.  
## The chi square statistic is 2.18 on 4 degrees of freedom.  
## The p-value is 0.703

**#### From the P-value we can clearly see that that the model is a good fit for the data  
#### since p>0.05(by a very huge margin)**

Question 3

library(sem)

model.1<-specify.model(file="mode.txt.txt")

model.1

#####Model for the question

## Path Parameter StartValue  
## 1 doctor -> X1 lambda11   
## 2 doctor -> X3 lambda31   
## 3 doctor -> X4 lambda41   
## 4 doctor -> X8 lambda81   
## 5 patient -> X2 lambda21   
## 6 patient -> X5 lambda51   
## 7 patient -> X6 lambda61   
## 8 patient -> X7 lambda71   
## 9 X1 <->X1 psi1   
## 10 X2 <->X2 psi2   
## 11 X3 <->X3 psi3   
## 12 X4 <->X4 psi4   
## 13 X5 <->X5 psi5   
## 14 X6 <->X6 psi6   
## 15 X7 <->X7 psi7   
## 16 X8 <->X8 psi8   
## 17 doctor <->doctor <fixed> 1   
## 18 patient <->patient <fixed> 1   
## 19 doctor <->patient rho

lt <- readMoments("EverittEx7.1.txt", diag = T)  
R <- (lt + t(lt)) - diag(1, 9)  
R2 <- R[-9, -9]  
R2

## X1 X2 X3 X4 X5 X6 X7 X8  
## X1 1.00 -0.04 0.61 0.45 0.03 -0.29 -0.30 0.45  
## X2 -0.04 1.00 -0.07 -0.12 0.49 0.43 0.30 -0.31  
## X3 0.61 -0.07 1.00 0.59 0.03 -0.13 -0.24 0.59  
## X4 0.45 -0.12 0.59 1.00 -0.08 -0.21 -0.19 0.63  
## X5 0.03 0.49 0.03 -0.08 1.00 0.47 0.41 -0.14  
## X6 -0.29 0.43 -0.13 -0.21 0.47 1.00 0.63 -0.13  
## X7 -0.30 0.30 -0.24 -0.19 0.41 0.63 1.00 -0.26  
## X8 0.45 -0.31 0.59 0.63 -0.14 -0.13 -0.26 1.00

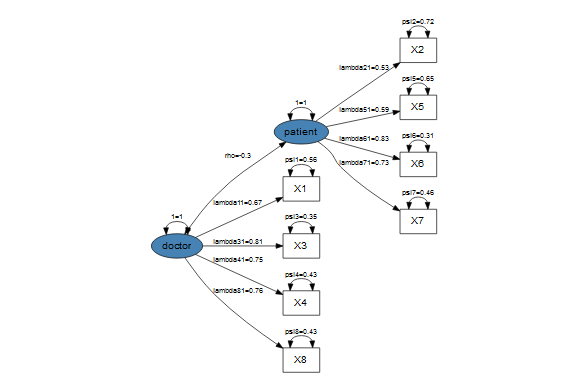
doc\_patient\_sem<- sem::sem(model = model.1, S = R2, N = 123)  
table<-summary(doc\_patient\_sem)  
  
library(DiagrammeR)

pathDiagram(doc\_patient\_sem,ignore.double = FALSE, edge.labels = "both", file="fitted",output.type = "dot",node.colors = c("steelblue","transparent"))grViz("fitted.dot")  
###Create a confidence in  
##lower limit  
-0.3049759+(1.96\* 0.10136386)

## [1] -0.1063027

##Upper limit  
-0.3049759-(1.96\* 0.10136386)

## [1] -0.5036491



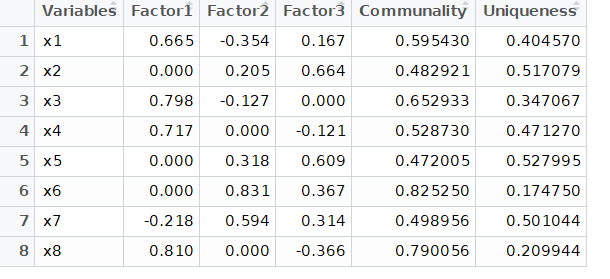
question\_4.R

**#a we can see 8 manifest variables  
# 3 common factors  
#b how to take the uniqueness from the partial output?**efa\_result<-data.frame("Variables"=c("x1","x2","x3","x4","x5","x6","x7","x8"),  
 "Factor1" = c(0.665,0,0.798,0.717,0,0,-0.218,0.810),   
 "Factor2" = c(-0.354,0.205,-0.127,0,0.318,0.831,0.594,0),   
 "Factor3" = c(0.167,0.664,0,-0.121,0.609,0.367,0.314,-0.366))  
  
**#Communality = Row sum of Manifest variables factors**  
efa\_result$Communality<-(efa\_result$Factor1)^2+(efa\_result$Factor2)^2+(efa\_result$Factor3)^2  
**#Uniqueness = 1 - Communality**  
efa\_result$Uniqueness<-1-efa\_result$Communality  
efa\_result$Communality

## [1] 0.595430 0.482921 0.652933 0.528730 0.472005 0.825250 0.498956 0.790056

efa\_result$Uniqueness

## [1] 0.404570 0.517079 0.347067 0.471270 0.527995 0.174750 0.501044 0.209944



**#c technically yes but no because the p-value lies on the border line  
#d 0.289 and 0.457  
#e *Factor 2 has highest correlation with X7. The loading value of X7 is higher under Factor* 2.**