Appendix

Data Cleaning

Chi-square plot of the Mahalanobis distances

AreaDroughtNew=AreaDrought   
   
*#Get column means + cov*   
xbar=colMeans(AreaDroughtNew)   
S=cov(AreaDroughtNew)   
   
*#using the mahalanobis function*   
d2=mahalanobis(AreaDroughtNew,xbar,S)   
*#finding the quantiles*   
quantiles=qchisq((1:nrow(AreaDroughtNew)-1/2)/nrow(AreaDroughtNew),df=ncol(AreaDroughtNew))   
sd2=sort(d2)   
*#plotting quantiles versus ordered squared distances*   
plot(quantiles,sd2,xlab='Quantile',ylab='Ordered Square Distances')   
abline(a=0,b=1)

Obtaining Outliers + Removing Outliers

*#text(quantiles,sd2,labels=row.names(AreaDroughtNew),pos = 1,cex=0.4)*   
outliers <- which(d2> qchisq(0.975, ncol(S)))   
outliers

AreaDroughtClean=AreaDrought[-outliers,]

PCA Analysis

Running PCA

areaDroughtPCA=princomp(AreaDroughtClean,cor=T)   
   
   
print(summary(areaDroughtPCA,loading=T),cut=0.4)

Checking that correlations between components are 0

AreaDroughtScore=areaDroughtPCA$scores[,1:3]   
*#round(AreaDroughtScore,2)*   
   
round(cor(AreaDroughtScore),2)

Cluster Analysis

Single Linkage

AreaDroughtNew=AreaDrought   
   
AreaDroughtScale=scale(AreaDroughtNew)   
AreaDroughtDist=dist(AreaDroughtScale)   
   
AreaDroughtHClust=hclust(AreaDroughtDist,method='single')   
plot(AreaDroughtHClust,main="Single Linkage of Area Drought")

plot(rev(AreaDroughtHClust$height),xlim = c(0,100))

Average Linkage

AreaDroughtHClust=hclust(AreaDroughtDist,method='average')   
plot(AreaDroughtHClust,main="Average Linkage of Area Drought")

ct=cutree(AreaDroughtHClust,k=3)   
table(ct)

Complete

AreaDroughtHClust=hclust(AreaDroughtDist,method='complete')   
ct=cutree(AreaDroughtHClust,k=6)   
table(ct)

plot(AreaDroughtNew[,c(1,5)],col=ct,main="Colored based on Complete Linkage")

K-Means

plot.wgss(AreaDroughtScale,maxc=30)

AreaDroughtKMean=kmeans(AreaDroughtScale,centers=3,nstart=10)   
table(AreaDroughtKMean$cluster)

plot(AreaDroughtNew,col=AreaDroughtKMean$cluster,main="Colored based on Kmeans")

K-means column Means

clusterMeans=AreaDroughtKMean$centers   
rownames(clusterMeans)=c("Clust1","Clust2","Clust3")   
library(knitr)   
   
kable(clusterMeans,caption="Column Means for Each Cluster")

*Column Means for Each Cluster*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | PAD1 | FCET | AGSBillions | YIR | PPIF | PPD1 |
| Clust1 | 0.9796945 | -0.9644277 | 0.3643360 | 0.2583599 | 0.7260874 | 0.9668258 |
| Clust2 | 0.2101822 | 0.6028553 | -0.0661409 | 0.4880050 | -0.7390956 | -0.0573961 |
| Clust3 | -0.8815986 | -0.2798341 | -0.1233286 | -0.8455916 | 0.6149338 | -0.4946865 |

Model Based Clustering

library(mclust)

mc=Mclust(AreaDroughtScale)   
table(mc$classification)

plot(mc, what="BIC")

plot(mc,what="classification",c(1,2))

plot(mc,what="uncertainty",c(1,2))

Exploratory Factor Analysis

AreaDroughtFactorAnalysis=factanal(AreaDroughtScale, factors=3)   
   
AreaDroughtFactorAnalysis

print(AreaDroughtFactorAnalysis$loadings,cut=.4)

f.loading=AreaDroughtFactorAnalysis$loadings[,1:2]   
   
corHat=f.loading%\*%t(f.loading)+diag(AreaDroughtFactorAnalysis$uniquenesses)   
corr=cor(AreaDroughtScale)   
   
   
print('This is the Root mean Square Error')

rmse=sqrt(mean((corHat-corr)^2))   
rmse

Confirmatory Factor Analysis

library(sem)

modelCFA=specifyModel(text="   
                            Drought->PAD1,lambda1,NA   
                            Drought->PPD1,lambda2,NA   
                            Inflation->AGSBillions,lambda3,NA   
                            Inflation->YIR,lambda4,NA   
                            Food\_Supply->FCET,lambda5,NA   
                            Food\_Supply->PPIF,lambda6,NA   
                               
                            Drought<->Inflation,rho,NA   
                            Inflation<->Food\_Supply,rho2,NA   
                            Food\_Supply<->Drought,rho3,NA   
                               
                               
                            PAD1<->PAD1,theta1,NA   
                            FCET<->FCET,theta2,NA   
                            AGSBillions<->AGSBillions,theta3,NA   
                            YIR<->YIR,theta4,NA   
                            PPIF<->PPIF,theta5,NA   
                            PPD1<->PPD1,theta6,NA   
                            Drought<->Drought,NA,1   
                            Inflation<->Inflation,NA,1   
                            Food\_Supply<->Food\_Supply,NA,1")

covthing=cov(AreaDrought)   
covthing

*ability\_sem=sem(modelCFA,covthing,nrow(AreaDrought))*   
*summary(ability\_sem)*