ACP_state.x77

Irma Eunice Martínez de la Cruz

2022-04-04

`{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)`

ANALISIS FACTORIAL

1.- Lectura de la matriz de datos

```
x<-as.data.frame(state.x77)
```

#2.- Quitar los espacios de los nombres

```
colnames(x)[4]="Life.Exp"
colnames(x)[6]= "HS.Grad"
```

#3.- Separa n (estados) y p (variables)

```
n<-dim(x)[1]
p<-dim(x)[2]
```

4.- Generacion de un scater plot para la

Visualización de variables originales

```
pairs(x, col="pink", pch=19, main="Matriz original")
```

Transformación de alguna varibles

1.- Aplicamos logaritmo para las columnas 1,3 y 8

```
x[,1]<-log(x[,1])
colnames(x)[1]<-"Log-Population"

x[,3]<-log(x[,3])
colnames(x)[3]<-"Log-Illiteracy"

x[,8]<-log(x[,8])
colnames(x)[8]<-"Log-Area"
```

Grafico scater para la visualizacion de la

Matriz original con 3 variables que se incluyeron

```
pairs(x,col="red", pch=19, main="Matriz original")
```

Nota: Como las variables tiene diferentes unidades

de medida, se va a implementar la matriz de

correlaciones para estimar la matriz de carga

Reduccion de la dimensionalidad

Análsis Factorial de componentes principales (PCFA)

1.- Calcular la matriz de medias y de correlaciones

Matriz de medias

```
mu<-colMeans(x)
mu
```

Matriz de correlaciones

```
R<-cor(x)
R
```

2.- Reducción de la dimensionalidad mediante

Análisis factorial de componentes principales (PCFA).

1.- Calcular los valores y vectores propios.

```
eR<-eigen(R)
```

2.- Valores propios

```
eigen.val<-eR$values
eigen.val
```

3.- Vectores propios

```
eigen.vec<-eR$vectors
eigen.vec
```

4.- Calcular la proporcion de variabilidad

```
prop.var<-eigen.val/sum(eigen.val)
prop.var
```

5.- Calcular la proporcion de variabilidad acumulada

```
prop.var.acum<-cumsum(eigen.val)/sum(eigen.val)
prop.var.acum
```

Estimacion de la matriz de carga

Nota: se estima la matriz de carga usando los

autovalores y autovectores.

se aplica la rotación varimax

Primera estimación de Lamda mayuscula

se calcula multiplicando la matriz de los

3 primeros autovectores por la matriz diagonal

formada por la raiz cuadrada de los primeros

3 autovalores.

```
L.est.1<-eigen.vec[,1:3] %*% diag(sqrt(eigen.val[1:3]))
L.est.1
```

Rotación varimax

```
L.est.1.var<-varimax(L.est.1)
L.est.1.var
```

Estimación de la matriz de los errores

1.- Estimación de la matriz de perturbaciones

```
Psi.est.1<-diag(diag(R-as.matrix(L.est.1.var$loadings)%*% t(as.matrix(L.est.1.var$loadings))))
Psi.est.1
```

2.- Se utiliza el método Análisis de factor principal (PFA)

para estimación de autovalores y autovectores

```
RP<-R-Psi.est.1
RP
```

Calculo de la matriz de autovalores y autovectores

```
eRP<-eigen(RP)
```

Autovalores

```
eigen.val.RP<-eRP$values
eigen.val.RP
```

Autovectores

```
eigen.vec.RP<-eRP$vectors
eigen.val.RP
```

Proporcion de variabilidad

```
prop.var.RP<-eigen.val.RP/ sum(eigen.val.RP)
prop.var.RP
```

Proporcion de variabilidad acumulada

```
prop.var.RP.acum<-cumsum(eigen.val.RP)/ sum(eigen.val.RP)
prop.var.RP.acum
```

Estimación de la matriz de cargas

con rotación varimax

```
L.est.2<-eigen.vec.RP[,1:3] %*% diag(sqrt(eigen.val.RP[1:3]))
L.est.2
```

Rotacion varimax

```
L.est.2.var<-varimax(L.est.2)
```

Estimación de la matriz de covarianzas de los errores.

```
Psi.est.2<-diag(diag(R-as.matrix(L.est.2.var$loadings)%*% t(as.matrix(L.est.2.var$loadings))))
Psi.est.2
```

Obtencion de los scores de ambos métodos

PCFA

```
FS.est.1<-scale(x)%*% as.matrix(L.est.1.var$loadings)
FS.est.1
```

PFA

```
FS.est.2<-scale(x)%*% as.matrix (L.est.2.var$loadings)
FS.est.2
```

Graficamos ambos scores

```
par(mfrow=c(2,1))
```

Factor I y II

```
pl1<-plot(FS.est.1[,1], FS.est.1[,2], xlab="primer factor",
          ylab="segundo factor", main="scores con factor I y II con PCFA",
          pch=19, col="blue")
text(FS.est.1[,1], FS.est.1[,2], labels = rownames(x), pos=4, col="green")
```

Factor I y III

```
pl2<-plot(FS.est.1[,1], FS.est.1[,3], xlab="Primer factor",
          ylab="Tercer factor", main="scores con factor I y III con PCFA",
          pch=19, col="blue")
text(FS.est.1[,1], FS.est.1[,3], labels = rownames(x), pos=4, col="red")
```

Factor II y III

```
pl3<-plot(FS.est.1[,2], FS.est.1[,3], xlab="Segundo factor",
          ylab="Tercer factor", main="scores con factor II y III con PCFA",
          pch=19, col="blue")
text(FS.est.1[,2], FS.est.1[,3], labels = rownames(x), pos=4, col="black")
```