Análisis Multivariado

Producto académico 02

Kevin Heberth Haquehua Apaza

23 de julio del 2025

Tabla de Contenidos

[Examen AED multivariado, ACP y AFE 1](#_Toc204194776)

[Ejercicio 1: 1](#_Toc204194777)

[Solución 2](#_Toc204194778)

[Ejercicio 2: 7](#_Toc204194779)

[Solución 7](#_Toc204194780)

[Ejercicio 3: 10](#_Toc204194781)

[Solución 10](#_Toc204194782)

# Examen AED multivariado, ACP y AFE

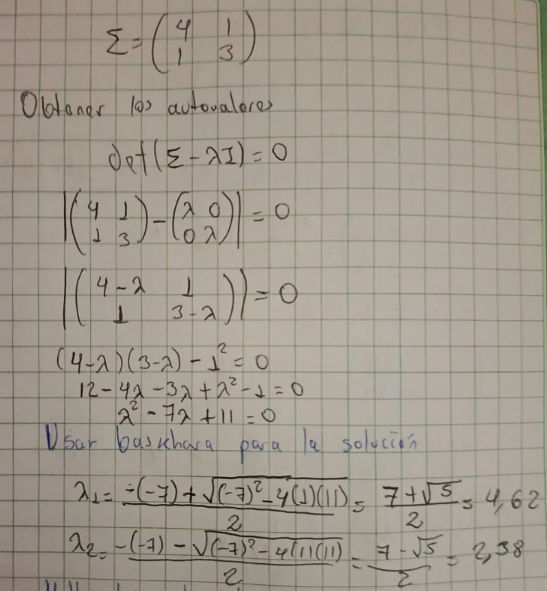
## Ejercicio 1:

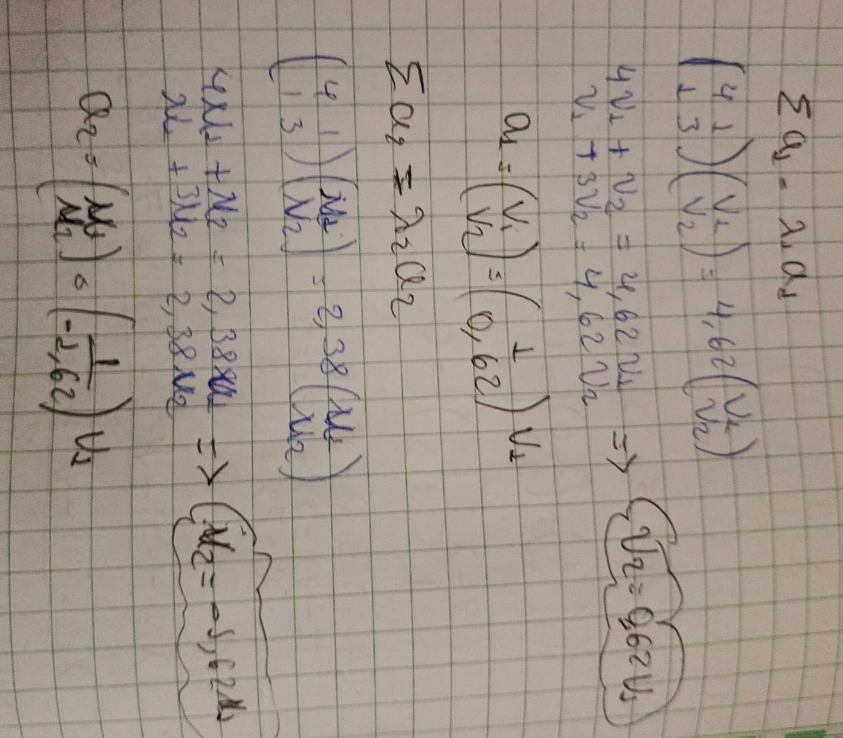
Sea un vector aleatorio con matriz de varianza-covarianza dada por

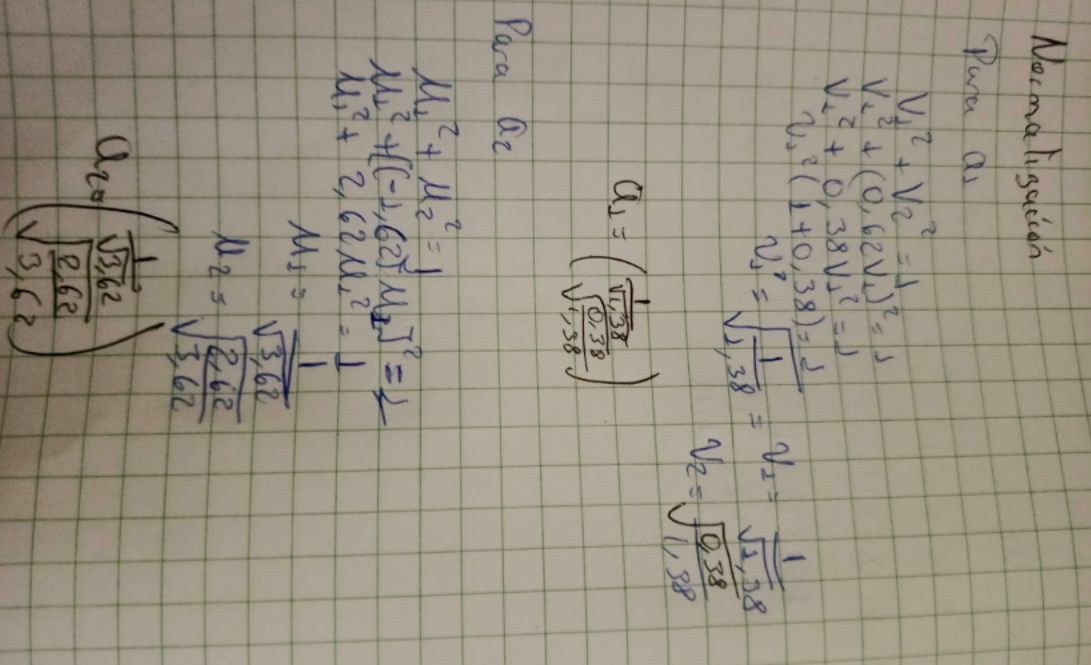
1. Determine las componentes principales y .
2. Calcule la proporción de la varianza total explicada por la primera componente principal.
3. Calcule la matriz de correlaciones a partir de la matriz de covarianzas y determine las componentes principales y a partir de . Calcule la proporción de la varianza total explicada por
4. Calcule la correlación entre las variables y las componentes principales, es decir, calcule

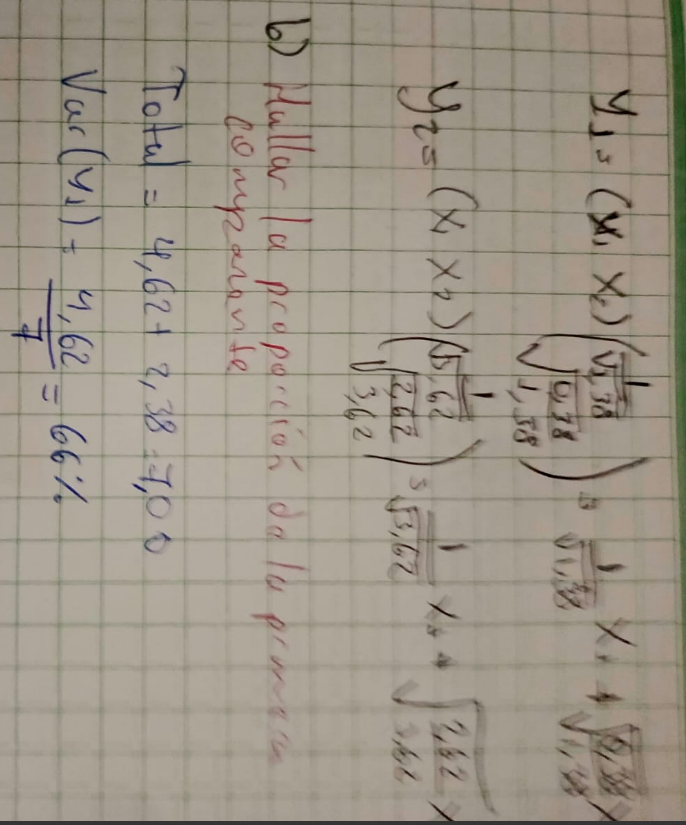
### Solución

a y b)









1. **Calcule la matriz de correlaciones a partir de la matriz de covarianzas y determine las componentes principales y a partir de . Calcule la proporción de la varianza total explicada por**

S <- matrix(c(4, 1, 1, 3), 2, 2) ; S

## [,1] [,2]  
## [1,] 4 1  
## [2,] 1 3

Calcular la matriz de correlaciones

D\_inv <- diag(1 / sqrt(diag(S)))  
R <- D\_inv %\*% S %\*% D\_inv  
R

## [,1] [,2]  
## [1,] 1.0000000 0.2886751  
## [2,] 0.2886751 1.0000000

Ahora hallar los componentes principales

pca <- princomp(covmat = R, cor = TRUE)

Y las varianzas explicadas

explained\_var <- pca$sdev^2 / sum(pca$sdev^2)  
explained\_var

## Comp.1 Comp.2   
## 0.6443376 0.3556624

En el cálculo manual no hya tanta diferencia, debido al cálculo con decimales.

1. **Calcule la correlación entre las variables y las componentes principales, es decir, calcule**

L <- unclass(pca$loadings) ; L

## Comp.1 Comp.2  
## [1,] 0.7071068 0.7071068  
## [2,] 0.7071068 -0.7071068

# Autovalores  
lambda <- pca$sdev^2  
  
# Correlaciones  
cor <- L %\*% diag(sqrt(lambda))  
cor

## [,1] [,2]  
## [1,] 0.8027064 0.5963744  
## [2,] 0.8027064 -0.5963744

## Ejercicio 2:

La siguiente tabla muestra los datos sobre la longitud de huesos registrados de 20 jóvenes a los 8, 8.5, 9 y 9.5 años respectivamente; Verificar si alguno de los individuos es considerado un dato atípico multivariado. Realizar la comprobación paso a paso como se realizó en clase (matricialmente), además tienes que comprobarlo con la función directa en el R

data <- data.frame(X1\_y8 = c(47.8, 46.4, 46.3, 45.1, 47.6, 52.5, 51.2, 49.8,  
 48.1, 45.0, 51.2, 48.5, 52.1, 48.2, 49.6, 50.7,  
 47.2, 53.3, 46.2, 46.3),  
 X2\_y8.5 = c(48.8, 47.3, 46.8, 45.3, 48.5, 53.2, 53.0, 50.0,  
 50.8, 47.0, 51.4, 49.2, 52.8, 48.9, 50.4, 51.7,  
 47.7, 54.6, 47.5, 47.6),  
 X3\_y9 = c(49.0, 47.7, 47.8, 46.1, 48.9, 53.3, 54.3, 50.3,  
 52.3, 47.3, 51.6, 53.0, 53.7, 49.3, 51.2, 52.7,  
 48.4, 55.1, 48.1, 51.3),  
 X3\_y9.5 = c(49.7, 48.4, 48.5, 47.2, 49.3, 53.7, 54.5, 52.7,  
 54.4, 48.3, 51.9, 55.5, 55.0, 49.8, 51.8, 53.3,  
 49.5, 55.3, 48.4, 51.8))

### Solución

Hallemos manualmente

# Calcular vector de medias  
mu <- colMeans(data) ; mu

## X1\_y8 X2\_y8.5 X3\_y9 X3\_y9.5   
## 48.655 49.625 50.570 51.450

data\_centered <- t(apply(data, 1, function(x) x - mu))  
data\_centered <- as.matrix(data\_centered) ; data\_centered

## X1\_y8 X2\_y8.5 X3\_y9 X3\_y9.5  
## [1,] -0.855 -0.825 -1.57 -1.75  
## [2,] -2.255 -2.325 -2.87 -3.05  
## [3,] -2.355 -2.825 -2.77 -2.95  
## [4,] -3.555 -4.325 -4.47 -4.25  
## [5,] -1.055 -1.125 -1.67 -2.15  
## [6,] 3.845 3.575 2.73 2.25  
## [7,] 2.545 3.375 3.73 3.05  
## [8,] 1.145 0.375 -0.27 1.25  
## [9,] -0.555 1.175 1.73 2.95  
## [10,] -3.655 -2.625 -3.27 -3.15  
## [11,] 2.545 1.775 1.03 0.45  
## [12,] -0.155 -0.425 2.43 4.05  
## [13,] 3.445 3.175 3.13 3.55  
## [14,] -0.455 -0.725 -1.27 -1.65  
## [15,] 0.945 0.775 0.63 0.35  
## [16,] 2.045 2.075 2.13 1.85  
## [17,] -1.455 -1.925 -2.17 -1.95  
## [18,] 4.645 4.975 4.53 3.85  
## [19,] -2.455 -2.125 -2.47 -3.05  
## [20,] -2.355 -2.025 0.73 0.35

# Calcular matriz de covarianza e inversa  
S <- cov(data) ; S

## X1\_y8 X2\_y8.5 X3\_y9 X3\_y9.5  
## X1\_y8 6.329974 6.189079 5.777000 5.548158  
## X2\_y8.5 6.189079 6.449342 6.153421 5.923421  
## X3\_y9 5.777000 6.153421 6.918000 6.946316  
## X3\_y9.5 5.548158 5.923421 6.946316 7.464737

S\_inv <- solve(S) ; S\_inv

## X1\_y8 X2\_y8.5 X3\_y9 X3\_y9.5  
## X1\_y8 2.6751033 -2.9161240 0.5007711 -0.1402566  
## X2\_y8.5 -2.9161240 4.3649625 -2.2211535 0.7706152  
## X3\_y9 0.5007711 -2.2211535 4.6620565 -2.9479467  
## X3\_y9.5 -0.1402566 0.7706152 -2.9479467 2.3699236

# Calcular distancia de Mahalanobis manualmente (paso a paso)  
dist\_manual <- (data\_centered) %\*% S\_inv %\*% t(data\_centered)  
dist\_manual<-(diag(dist\_manual)) ; dist\_manual

## [1] 0.7589305 1.2973512 1.7542420 3.8608496 0.8776488 2.8221570  
## [7] 4.0573285 8.1105862 10.9511973 5.8461957 2.8393025 10.5795182  
## [13] 2.5797150 0.6625258 0.3324791 0.8462319 1.1141959 4.4104027  
## [19] 2.1249099 10.1742322

Ahora hallemos con la función directa del R

# Calcular con la función base de R  
dist\_r <- mahalanobis(data, center = mu, cov = S) ; dist\_r

## [1] 0.7589305 1.2973512 1.7542420 3.8608496 0.8776488 2.8221570  
## [7] 4.0573285 8.1105862 10.9511973 5.8461957 2.8393025 10.5795182  
## [13] 2.5797150 0.6625258 0.3324791 0.8462319 1.1141959 4.4104027  
## [19] 2.1249099 10.1742322

Comparemos

# Comparar resultados  
resultado <- data.frame(  
 Observacion = rownames(data),  
 Dist\_Mahal\_Manual = dist\_manual,  
 Dist\_Mahal\_R = dist\_r,  
 Diferencia = abs(dist\_manual - dist\_r)  
)  
  
print(resultado)

## Observacion Dist\_Mahal\_Manual Dist\_Mahal\_R Diferencia  
## 1 1 0.7589305 0.7589305 0.000000e+00  
## 2 2 1.2973512 1.2973512 0.000000e+00  
## 3 3 1.7542420 1.7542420 0.000000e+00  
## 4 4 3.8608496 3.8608496 0.000000e+00  
## 5 5 0.8776488 0.8776488 0.000000e+00  
## 6 6 2.8221570 2.8221570 0.000000e+00  
## 7 7 4.0573285 4.0573285 8.881784e-16  
## 8 8 8.1105862 8.1105862 0.000000e+00  
## 9 9 10.9511973 10.9511973 0.000000e+00  
## 10 10 5.8461957 5.8461957 0.000000e+00  
## 11 11 2.8393025 2.8393025 0.000000e+00  
## 12 12 10.5795182 10.5795182 1.776357e-15  
## 13 13 2.5797150 2.5797150 4.440892e-16  
## 14 14 0.6625258 0.6625258 0.000000e+00  
## 15 15 0.3324791 0.3324791 5.551115e-17  
## 16 16 0.8462319 0.8462319 0.000000e+00  
## 17 17 1.1141959 1.1141959 2.220446e-16  
## 18 18 4.4104027 4.4104027 8.881784e-16  
## 19 19 2.1249099 2.1249099 0.000000e+00  
## 20 20 10.1742322 10.1742322 0.000000e+00

Practimente no se observan diferencias significativas. Ahora veamos los individuos considerados atípicos

# Umbral usando el percentil 97.5 de chi-cuadrado con 4 grados de libertad  
umbral <- qchisq(0.975, df = 4) ; umbral

## [1] 11.14329

# Agregar al dataframe  
resultado$dist\_maha <- dist\_r  
resultado$Atipico <- ifelse(resultado$dist\_maha > umbral, "Sí", "No")  
  
# Mostrar resultado  
print(resultado)

## Observacion Dist\_Mahal\_Manual Dist\_Mahal\_R Diferencia dist\_maha Atipico  
## 1 1 0.7589305 0.7589305 0.000000e+00 0.7589305 No  
## 2 2 1.2973512 1.2973512 0.000000e+00 1.2973512 No  
## 3 3 1.7542420 1.7542420 0.000000e+00 1.7542420 No  
## 4 4 3.8608496 3.8608496 0.000000e+00 3.8608496 No  
## 5 5 0.8776488 0.8776488 0.000000e+00 0.8776488 No  
## 6 6 2.8221570 2.8221570 0.000000e+00 2.8221570 No  
## 7 7 4.0573285 4.0573285 8.881784e-16 4.0573285 No  
## 8 8 8.1105862 8.1105862 0.000000e+00 8.1105862 No  
## 9 9 10.9511973 10.9511973 0.000000e+00 10.9511973 No  
## 10 10 5.8461957 5.8461957 0.000000e+00 5.8461957 No  
## 11 11 2.8393025 2.8393025 0.000000e+00 2.8393025 No  
## 12 12 10.5795182 10.5795182 1.776357e-15 10.5795182 No  
## 13 13 2.5797150 2.5797150 4.440892e-16 2.5797150 No  
## 14 14 0.6625258 0.6625258 0.000000e+00 0.6625258 No  
## 15 15 0.3324791 0.3324791 5.551115e-17 0.3324791 No  
## 16 16 0.8462319 0.8462319 0.000000e+00 0.8462319 No  
## 17 17 1.1141959 1.1141959 2.220446e-16 1.1141959 No  
## 18 18 4.4104027 4.4104027 8.881784e-16 4.4104027 No  
## 19 19 2.1249099 2.1249099 0.000000e+00 2.1249099 No  
## 20 20 10.1742322 10.1742322 0.000000e+00 10.1742322 No

Se observa que no se tiene ningun dato atípicos registrado, sin embargo la observación 9, 11 y 20 están cercas a considerarse outliers multivariados.

## Ejercicio 3:

En el Excel *(pregunta3.xlsx)* se muestran los valores de cinco variables obtenidas en 20 alumnos que quieren entrar a alguna universidad del consejo de rectores. Las variables en estudio son la distancia en kilómetros al lugar del colegio en el que estudiaban (DIST), el promedio de horas que hacían actividad física a la semana (EF), índice de masa corporal (IMC), IQ (coeficiente intelectual) y NEM (promedio de notas con el cual postulan a las universidades). Se quiere determinar las relaciones existentes entre dichas variables intentando reducir la dimensionalidad del problema vía un análisis factorial exploratorio.

1. Sobre el conjunto de datos halle si se cumple o no la normalidad multivariada.
2. Verifique si existe presencia de valores outliers para la data.
3. Realice el análisis factorial exploratorio vía componentes principales e interprete y justifique sus conclusiones.

### Solución

Leamos la data

library(readxl)  
pregunta3 <- read\_excel(here("pregunta3.xlsx"))

1. **Sobre el conjunto de datos halle si se cumple o no la normalidad multivariada.**

Veamos si los datos siguen una distribución multivariada con los test de Mardia, Henze-Zirkler y Royston

#Libreria a utilizar  
library(MVN)

## Warning: package 'MVN' was built under R version 4.5.1

Y <- pregunta3[,-1]

Mardia = mvn(Y, mvn\_test = "mardia")  
Mardia$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Mardia Skewness 59.425 0.006 asymptotic ✗ Not normal  
## 2 Mardia Kurtosis 0.976 0.329 asymptotic ✓ Normal

La prueba de Mardia indica que no se tiene una distribución normal, sigamos con el test de Henze-Zirkler

HZ =mvn(Y, mvn\_test = "hz")   
HZ$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Henze-Zirkler 0.899 0.068 asymptotic ✓ Normal

El test de Henz indica que es normal, ahora el test de Royston

Roy =mvn(Y, mvn\_test = "royston")   
Roy$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Royston 37.453 <0.001 asymptotic ✗ Not normal

En dos se indican que no se tiene una distribución normal y en uno al momento se indico que sigue una distribución normal. Veamos con los otros test

HW =mvn(Y, mvn\_test = "hw")   
HW$multivariate\_normality

## Test Statistic p.value Method MVN  
## 1 Henze-Wagner 0.859 0.041 asymptotic ✗ Not normal

HW =mvn(Y, mvn\_test = "doornik\_hansen")   
HW$multivariate\_normality

## Test Statistic df p.value Method MVN  
## 1 Doornik-Hansen 141.944 10 <0.001 asymptotic ✗ Not normal

HW =mvn(Y, mvn\_test = "energy")   
HW$multivariate\_normality

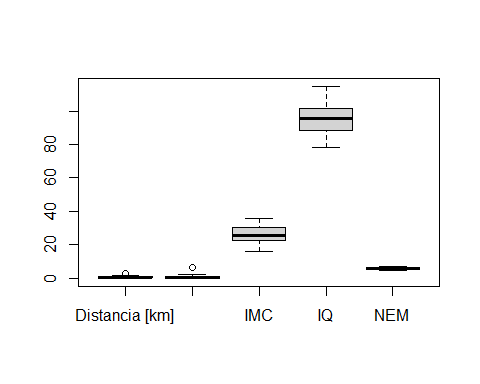
## Test Statistic p.value Method MVN  
## 1 E-Statistic 1.34 0.006 bootstrap ✗ Not normal

La mayor parte de los test indicaron que no se sigue una distribución normal multivariada, por lo tanto los datos no siguen una distribución normal multivariada. No se cumple la normalidad multivariada.

1. **Verifique si existe presencia de valores outliers para la data.**

Veamos primeramente a nivel univariado

boxplot(pregunta3[,-1])



Parece que distancia y EF tienen valores outliers a nivel univariado.

Ahora veamos a nivel multivariado a través de la distancia clásica de Mahalanobich

# Librerias a utilizar  
library(mvoutlier)

## Warning: package 'mvoutlier' was built under R version 4.5.1

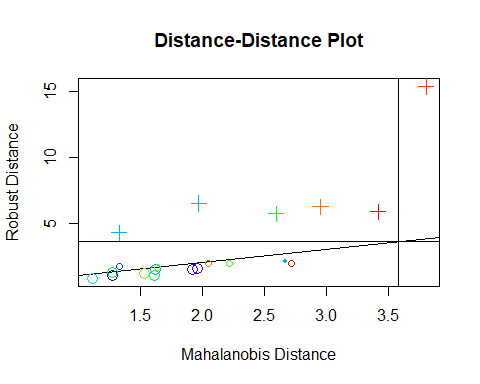
## Cargando paquete requerido: sgeostat

##   
## Adjuntando el paquete: 'mvoutlier'

## The following object is masked \_by\_ '.GlobalEnv':  
##   
## Y

library(aplpack)

distances <- dd.plot(pregunta3[,-1], quan=1/2, alpha=0.025)



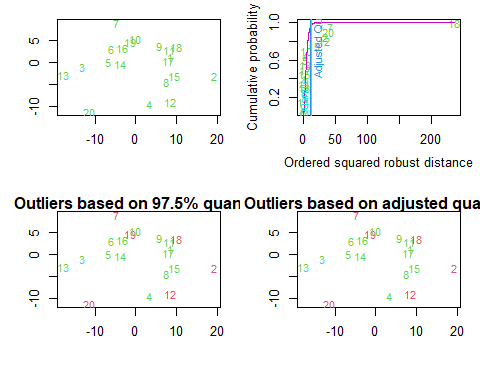
out1=data.frame(pregunta3,pred=as.numeric(distances$outliers))  
subset(out1,pred==1)

## Individuo Distancia..km. EF..hrs. IMC IQ NEM pred  
## 2 2 3.1442 0.4168 31.6282 77.9768 5.1 1  
## 7 7 0.2746 1.5676 15.9921 100.2147 5.5 1  
## 12 12 0.3059 0.2488 36.0617 89.4410 5.7 1  
## 18 18 0.1206 6.5881 22.7429 86.5072 5.7 1  
## 19 19 0.1266 1.2532 20.9371 97.3890 5.9 1  
## 20 20 0.4173 0.0742 35.5664 109.5347 6.3 1

Se muestran los individuos los cuales son outliers multivariados, ahora con la distancia de Mahalanobich robusta

outliers <- aq.plot(pregunta3[,-1], delta=qchisq(0.975, df = ncol(pregunta3[,-1])), quan = 1/2, alpha = 0.05)

## Projection to the first and second robust principal components.  
## Proportion of total variation (explained variance): 0.9194741



out2=data.frame(pregunta3,pred=as.numeric(outliers$outliers))  
subset(out2,pred==1)

## Individuo Distancia..km. EF..hrs. IMC IQ NEM pred  
## 2 2 3.1442 0.4168 31.6282 77.9768 5.1 1  
## 7 7 0.2746 1.5676 15.9921 100.2147 5.5 1  
## 12 12 0.3059 0.2488 36.0617 89.4410 5.7 1  
## 18 18 0.1206 6.5881 22.7429 86.5072 5.7 1  
## 19 19 0.1266 1.2532 20.9371 97.3890 5.9 1  
## 20 20 0.4173 0.0742 35.5664 109.5347 6.3 1

Se tienen los mismo individuos (2, 7, 12, 18, 19 y 20) como outliers multivariados.

1. **Realice el análisis factorial exploratorio vía componentes principales e interprete y justifique sus conclusiones.**

Ver si existen datos NA

colSums(is.na(pregunta3))

## Individuo Distancia [km] EF [hrs] IMC IQ   
## 0 0 0 0 0   
## NEM   
## 0

No se tienen datos perdidos, empecemos realizando la prueba de esfericidad de Bartlet

library(psych)

##   
## Adjuntando el paquete: 'psych'

## The following object is masked from 'package:MVN':  
##   
## mardia

cortest.bartlett(cor(Y), n=nrow(Y))

## $chisq  
## [1] 26.14193  
##   
## $p.value  
## [1] 0.00355391  
##   
## $df  
## [1] 10

El valor es significativo, justifica el uso de reducción de datos, veamos el test KMO

KMO(Y)

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = Y)  
## Overall MSA = 0.37  
## MSA for each item =   
## Distancia [km] EF [hrs] IMC IQ NEM   
## 0.57 0.21 0.23 0.40 0.40

Los valores son menores a 0.5 a excepción de Distancia [km] por lo que se puede extraer para que se considere aceptable la aplicación del análisis factorial al conjunto de datos

data\_AFE <- Y[,-1]

Veamos el test de bartlet y KMO

cortest.bartlett(cor(data\_AFE), n=nrow(data\_AFE))

## $chisq  
## [1] 19.15131  
##   
## $p.value  
## [1] 0.003915573  
##   
## $df  
## [1] 6

Significativo, ahora veamos el KMO

KMO(data\_AFE)

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = data\_AFE)  
## Overall MSA = 0.29  
## MSA for each item =   
## EF [hrs] IMC IQ NEM   
## 0.23 0.13 0.34 0.33

Ahora si es justificable el uso de un análisis factorial exploratorio, realizemos con la rotación principal

facto=principal(r=data\_AFE,nfactors=4,rotate="none")  
facto=principal(r=data\_AFE,nfactors=4,rotate="varimax")  
facto

## Principal Components Analysis  
## Call: principal(r = data\_AFE, nfactors = 4, rotate = "varimax")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## RC1 RC2 RC3 RC4 h2 u2 com  
## EF [hrs] -0.01 -0.05 0.99 -0.15 1 0.0e+00 1.1  
## IMC 0.14 0.98 -0.05 -0.10 1 -2.2e-15 1.1  
## IQ 0.40 -0.15 -0.21 0.88 1 -2.2e-15 1.6  
## NEM 0.92 0.18 0.00 0.35 1 -3.8e-15 1.4  
##   
## RC1 RC2 RC3 RC4  
## SS loadings 1.03 1.03 1.02 0.92  
## Proportion Var 0.26 0.26 0.26 0.23  
## Cumulative Var 0.26 0.51 0.77 1.00  
## Proportion Explained 0.26 0.26 0.26 0.23  
## Cumulative Proportion 0.26 0.51 0.77 1.00  
##   
## Mean item complexity = 1.3  
## Test of the hypothesis that 4 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0   
## with the empirical chi square 0 with prob < NA   
##   
## Fit based upon off diagonal values = 1

facto$loadings

##   
## Loadings:  
## RC1 RC2 RC3 RC4   
## EF [hrs] 0.987 -0.152  
## IMC 0.136 0.984 -0.101  
## IQ 0.404 -0.148 -0.212 0.878  
## NEM 0.920 0.182 0.348  
##   
## RC1 RC2 RC3 RC4  
## SS loadings 1.028 1.026 1.022 0.924  
## Proportion Var 0.257 0.257 0.255 0.231  
## Cumulative Var 0.257 0.513 0.769 1.000

Por la mayor explicación de la varianza, se recomienda usar 3 factores

facto=principal(r=data\_AFE,nfactors=3,rotate="varimax")  
facto

## Principal Components Analysis  
## Call: principal(r = data\_AFE, nfactors = 3, rotate = "varimax")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## RC1 RC2 RC3 h2 u2 com  
## EF [hrs] -0.10 -0.05 0.99 0.99 0.011 1.0  
## IMC 0.03 0.99 -0.05 0.98 0.022 1.0  
## IQ 0.88 -0.26 -0.30 0.93 0.075 1.4  
## NEM 0.92 0.28 0.06 0.94 0.065 1.2  
##   
## RC1 RC2 RC3  
## SS loadings 1.63 1.12 1.07  
## Proportion Var 0.41 0.28 0.27  
## Cumulative Var 0.41 0.69 0.96  
## Proportion Explained 0.43 0.29 0.28  
## Cumulative Proportion 0.43 0.72 1.00  
##   
## Mean item complexity = 1.2  
## Test of the hypothesis that 3 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.04   
## with the empirical chi square 0.39 with prob < NA   
##   
## Fit based upon off diagonal values = 0.99

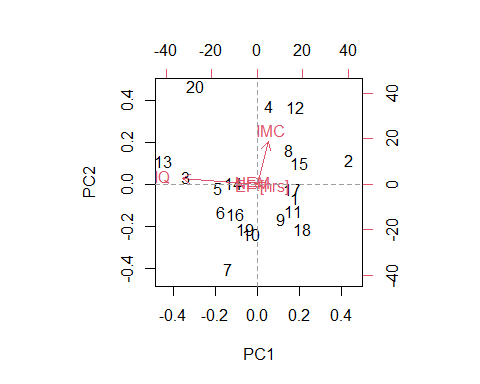
facto$loadings

##   
## Loadings:  
## RC1 RC2 RC3   
## EF [hrs] -0.100 0.988  
## IMC 0.987   
## IQ 0.879 -0.256 -0.297  
## NEM 0.923 0.283   
##   
## RC1 RC2 RC3  
## SS loadings 1.634 1.122 1.071  
## Proportion Var 0.409 0.281 0.268  
## Cumulative Var 0.409 0.689 0.957

Se tienen los siguientes factores

RC1 = IQ y NEM (Academico) RC2 = IMC (Caracteristicas del Individuo) RC3 = EF (Nivel de actividad fisica)

biplot(prcomp(data\_AFE, scale = FALSE))  
abline(h = 0, v = 0, lty = 2, col = 8)



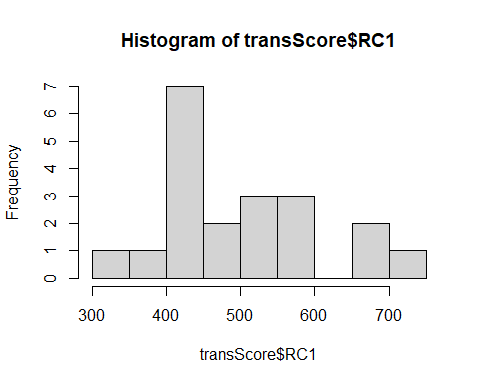
Saquemos ahora los scores

scores <- as.matrix(data\_AFE) %\*% as.matrix(facto$loadings)  
scores <- data.frame(scores) ; scores

## RC1 RC2 RC3  
## 1 82.99803 4.3416280 -26.14869  
## 2 74.23487 12.6940176 -23.87682  
## 3 103.34898 -1.5263996 -33.21591  
## 4 90.08681 12.4732635 -27.22022  
## 5 97.81104 -0.1204168 -31.09675  
## 6 96.00638 -2.7172449 -31.00943  
## 7 93.50009 -8.3734869 -28.57108  
## 8 85.44618 9.3336375 -26.05739  
## 9 85.18837 1.0958082 -27.47503  
## 10 89.61210 -3.0445936 -28.48060  
## 11 83.40019 2.9193219 -26.74265  
## 12 85.02641 14.3153197 -27.61324  
## 13 108.13035 -1.4962417 -34.78397  
## 14 95.19982 1.4903639 -30.53744  
## 15 83.54105 8.7027247 -25.46491  
## 16 93.03440 -2.0159418 -30.18438  
## 17 83.89996 5.2632489 -27.04577  
## 18 81.36580 1.6105370 -19.85049  
## 19 91.58438 -2.6406004 -28.25121  
## 20 103.23441 8.8635815 -33.68287

Realizemos una transformación manteniendo sus características

Zscores<-scale(scores)  
transScore <- Zscores\*100+500 # Proceso de baremación de PISA  
transScore <- data.frame(transScore)  
hist(transScore$RC1)



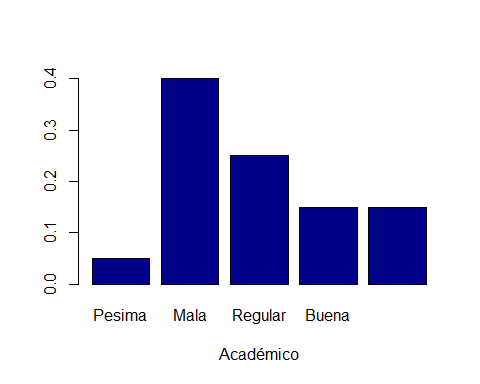
Recodifiquemos para la interpretación

#### RC1 (Academico)

transScore$RNC1[transScore$RC1<350] <-1  
transScore$RNC1[transScore$RC1>=350 & transScore$RC1<450] <-2  
transScore$RNC1[transScore$RC1>=450 & transScore$RC1<550] <-3  
transScore$RNC1[transScore$RC1>=550 & transScore$RC1<650] <-4  
transScore$RNC1[transScore$RC1>=650] <-5  
  
  
# Etiquetar  
transScore$RNC1 <- factor(transScore$RNC1,   
 labels = c("Pesima", "Mala", "Regular",  
 "Buena", "Excelente"))  
  
  
fi=table(transScore$RNC1)  
probabilidad=prop.table(table(transScore$RNC1))\*100  
cbind(fi,probabilidad)

## fi probabilidad  
## Pesima 1 5  
## Mala 8 40  
## Regular 5 25  
## Buena 3 15  
## Excelente 3 15

barplot(prop.table(table(transScore$RNC1)), col = "darkBlue", xlab = "Académico")



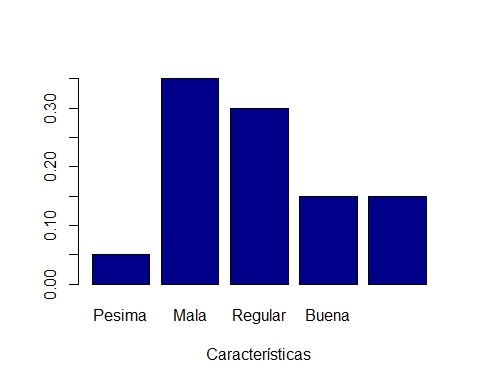
Se observa que la mayor parte se encuentra en un nivel malo y regular con respecto a la parte académica

#### RC2 (Características del individuo)

transScore$RNC2[transScore$RC2<350] <-1  
transScore$RNC2[transScore$RC2>=350 & transScore$RC2<450] <-2  
transScore$RNC2[transScore$RC2>=450 & transScore$RC2<550] <-3  
transScore$RNC2[transScore$RC2>=550 & transScore$RC2<650] <-4  
transScore$RNC2[transScore$RC2>=650] <-5  
  
  
# Etiquetar  
transScore$RNC2 <- factor(transScore$RNC2,   
 labels = c("Pesima", "Mala", "Regular",  
 "Buena", "Excelente"))  
  
  
fi=table(transScore$RNC2)  
probabilidad=prop.table(table(transScore$RNC2))\*100  
cbind(fi,probabilidad)

## fi probabilidad  
## Pesima 1 5  
## Mala 7 35  
## Regular 6 30  
## Buena 3 15  
## Excelente 3 15

barplot(prop.table(table(transScore$RNC2)), col = "darkBlue", xlab = "Características")



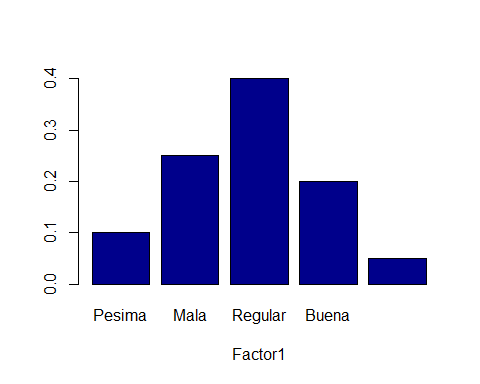
Se observa que la mayor parte se encuentra en un nivel malo y regular con respecto a las características de los individuos

#### RC3 (Nivel de actividad fisica)

transScore$RNC3[transScore$RC3<350] <-1  
transScore$RNC3[transScore$RC3>=350 & transScore$RC3<450] <-2  
transScore$RNC3[transScore$RC3>=450 & transScore$RC3<550] <-3  
transScore$RNC3[transScore$RC3>=550 & transScore$RC3<650] <-4  
transScore$RNC3[transScore$RC3>=650] <-5  
  
  
# Etiquetar  
transScore$RNC3 <- factor(transScore$RNC3,   
 labels = c("Pesima", "Mala", "Regular",  
 "Buena", "Excelente"))  
  
  
fi=table(transScore$RNC3)  
probabilidad=prop.table(table(transScore$RNC3))\*100  
cbind(fi,probabilidad)

## fi probabilidad  
## Pesima 2 10  
## Mala 5 25  
## Regular 8 40  
## Buena 4 20  
## Excelente 1 5

barplot(prop.table(table(transScore$RNC3)), col = "darkBlue", xlab = "Factor1")



Se observa que la mayor parte se encuentra en un nivel regular, bueno y malo, más balanceado con respecto a la parte del nivel de actividad física