

MAESTRIA EN MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE DATOS EN ORGANIZACIONES

E72. **01 Métodos de Análisis Multivariados**

**Trabajo Final Análisis de Componentes Principales**

Docentes: Vietri Silvia, Del Duca Silvina

Alumna: Roxana Florencia Strada

1. Introducción

En el presente trabajo se desarrolla un análisis de componentes principales sobre un datasets del plan ahora12[[1]](#footnote-1) por categorías de productos y por provincia donde busca evaluar y sacar conclusiones sobre los consumos realizados en cada provincia.

1. Presentación de la temática y desarrollo conceptual

El Programa Ahora 12 es un plan que te da la posibilidad de comprar productos nacionales en cuotas fijas con tu tarjeta de crédito.

Este Programa tiene vigencia hasta el 30 de junio del 2022.

La financiación del plan es parte de un acuerdo entre varias partes: comercios, tarjetas de crédito, bancos y el Estado. El Ministerio de Economía de la Nación es quien publicó la [resolución](http://infoleg.mecon.gov.ar/infolegInternet/anexos/235000-239999/235010/norma.htm) para generar el marco legal y ponerlo en marcha, mientras que el Banco Central de la República Argentina (BCRA) es quien define la política bancaria.

Los principales beneficiarios son los consumidores, que pueden financiar sus compras a 12 meses sin interés. Luego están los comercios, que dentro de este plan se benefician porque pagan el 10% de financiamiento a las tarjetas bancarias en lugar del 26% habitual. Los bancos son quienes absorben esa diferencia, pero a su vez el Banco Central estableció una reducción del 16% en el encaje que tienen que dejar en el Banco Central, es decir que pueden usar esos fondos que de lo contrario tendrían que estar inmovilizados.

Técnicas aplicadas:

* *MATRIZ DE CORRELACIONES:*

Se verificó que entre las variables exista correlación usando la matriz de correlaciones para observar que exista correlación entre las variables.

En este caso se partió de dicha matriz estandarizada y centrada para que las variables sean comparables.

* *TEST DE ESFERICIDAD DE BARLETT*

Adicionalmente se usó el test de esfericidad de barlett para contrastar si la matriz de correlaciones es una matriz de identidad, que indicaría que el modelo factorial es inadecuado. En este caso se obtuvo una respuesta en R de un pvalor=0 que significa que el modelo es adecuado para ser aplicado.

* *CRITERIO DEL AUTOVALOR SUPERIOR A LA UNIDAD*

Para obtener las componentes principales se usó el criterio del autovalor superior a la unidad de una matriz estandarizada y centrada, obteniendo así dos dimensiones que explican un 95% de la varianza acumulada.

* *GRAFICO DE SEDIMENTACION*

Adicionalmente se observó gráficamente a través de un gráfico de sedimentación llamado en R “scree plot” el número de componentes principales a extraer en este caso 3.

* *GRAFICO DE PUNTUACIONES*

Para la interpretación grafica se usó el grafico de puntuaciones que representa visualmente las puntuaciones del segundo componente principal versus las puntuaciones del primer componente principal.

* *GRAFICO DE INFLUENCIAS*

La gráfica de influencias grafica los coeficientes de cada variable para el primer componente versus los coeficientes para el segundo componente.

* *REPRESENTACION CONJUNTA*

Se utilizó la gráfica de doble proyección (biplot) para evaluar la estructura de los datos y las influencias de las dos primeros componentes en una gráfica.

1. Desarrollo metodológico
2. La base de datos fue extraída desde [*https://datos.gob.ar/dataset/produccion-programa-ahora-12*](https://datos.gob.ar/dataset/produccion-programa-ahora-12)
3. Representación de las variables:

|  |  |
| --- | --- |
| Variable | Tipo de dato |
| Provincia | Carácter |
| Alimentos | Numérica |
| Anteojos | Numérica |
| Artefactos de iluminación | Numérica |
| Artículos de Librería | Numérica |
| Balnearios | Numérica |
| Bicicletas | Numérica |
| Calzado y Marroquinería | Numérica |
| Colchones | Numérica |
| Computadoras | Numérica |
| Electrodomésticos | Numérica |
| Equipamiento médico | Numérica |
| Indumentaria | Numérica |
| Instrumentos musicales | Numérica |
| Juguetes | Numérica |
| Libros | Numérica |
| Máquinas y Herramientas | Numérica |
| Materiales para la construcción | Numérica |
| Medicamentos | Numérica |
| Motocicletas | Numérica |
| Muebles | Numérica |
| Neumáticos | Numérica |
| Perfumería | Numérica |
| Servicios de cuidado personal | Numérica |
| Servicios de instalación de alarmas | Numérica |
| Servicios de organización de eventos | Numérica |
| Servicios deportivos | Numérica |
| Servicios educativos | Numérica |
| Servicios técnicos | Numérica |
| Talleres de reparación | Numérica |
| Turismo | Numérica |

1. Supuestos teóricos

* *MATRIZ DE CORRELACIONES:*

Partiendo del supuesto de que exista correlación entre variables para poder aplicar el método se verificó que entre las variables exista dicha correlación usando la matriz de correlaciones.

En este caso se partió de dicha matriz estandarizada y centrada para que las variables sean comparables.

* *TEST DE ESFERICIDAD DE BARLETT*

Adicionalmente para comprobar que la matriz de correlaciones no sea la matriz identidad se usó el test de esfericidad de barlett de lo contrario indicaría que el modelo es inadecuado. En este caso se obtuvo una respuesta en R de un pvalor=0 que significa que rechazamos H0 y es posible aplicar componentes principales.

1. Sintaxis de R del análisis

**###ANALISIS DE COMPONENTES PRINCIPALES####**

#Desactivar notacion cientifica

options(scipen = 999)

**#Solicitamos descriptivas**

library(pastecs)

Descriptivas<-stat.desc(AHORA12,basic = TRUE)

**#MATRIZ DE CORRELACIONES**

library(corrplot)

R<-cor(AHORA12[2:15],method = "pearson")

testRes = cor.mtest(R,conf.level=0.95)

corrplot(R,p.mat = testRes[[1]],sig.level = 0.05, type = "lower")

**#TEST DE ESFERICIDAD DE BARLET**

library(psych)

cortest.bartlett(R,n=32)

**#NUMERO DE COMPONENTES PRINCIPALES A EXTRAER**

#Criterio del autovalor superior a la unidad

library(FactoMineR)

library(factoextra)

fit<-PCA(AHORA12[2:31],scale.unit=TRUE,ncp=10,graph=TRUE)

fit$eig

#Grafico con colores

fviz\_pca\_var(fit, col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)

**#SCREE PLOT**

library(factoextra)

fviz\_eig(fit,addlabels = TRUE, ylim = c(0, 100))+

theme\_grey()

**#INTERPRETACION DE LAS COMPONENTES PRINCIPALES**

#PCA individuales 1

library(ggplot2)

library(ggrepel)

datos.grafico2<-data.frame(fit$var$coord[,1:2])

ggplot(datos.grafico2)+

geom\_point(aes(x=Dim.1,y=Dim.2,colour="darkred"))+

geom\_text\_repel(aes(x=Dim.1,y=Dim.2),

label=rownames(datos.grafico2))+

geom\_vline(xintercept=0,colour="darkgray")+

geom\_hline(yintercept = 0,colour="darkgray")+

labs(x="Dimension 1 (78%)", y = "Dimension 2 (12,1%) ")+

theme(legend.position="none")

#PCA individuales 1

fviz\_pca\_ind(fit, pointsize = "cos2",

pointshape = 21, fill = "#E7B800",

repel = TRUE)

**#CALCULO DE AUTOVALORES Y AUTOVECTORES DE LA MATRIZ DE CORRELACIONES**

library(corrplot)

S<-cor(AHORA12[,-1])

eigen(S)

Autovalores\_Autovectores<-eigen(S)

rownames(Autovalores\_Autovectores$vectors)<-c("Alimentos","Anteojos","Artefactos de iluminación",

"Artículos de Librería", "Balnearios", "Bicicletas",

"Calzado y Marroquinería", "Colchones", "Computadoras",

"Electrodomésticos", "Equipamiento médico", "Indumentaria",

"Instrumentos musicales", "Juguetes", "Libros",

"Máquinas y Herramientas",

"Materiales para la construcción",

"Medicamentos", "Motocicletas",

"Muebles", "Neumaticos", "Perfumeria", "Servicios de cuidado personal",

"Servicios de instalacion de alarmas", "Servicios de organizacion de eventos",

"Servicios deportivos", "Servicios educativos", "Servicios tecnicos",

"Talleres de reparacion", "Turismo"

)

colnames(Autovalores\_Autovectores$vectors)<-c("PC1","PC2","PC3","PC4","PC5","PC6",

"PC7","PC8","PC9","PC10","PC11","PC12","PC13","PC14","PC15",

"PC16","PC17","PC18","PC19","PC20","PC21","PC22","PC23","PC24","PC25",

"PC26","PC27","PC28","PC29","PC30"

)

Autovalores\_Autovectores

**#CALCULO DE LAS COMPONENTES PRINCIPALES**

AHORA12.pc<-princomp(S,cor=TRUE)

AHORA12.pc

**#REPRESENTACION DE LOS OBJETOS (PROVINCIAS)**

datos.grafico<-data.frame(fit$var$coord[,1:2],AHORA12$Provincia)

colnames(datos.grafico)<-c("Dim.1","Dim.2","provincia")

ggplot(datos.grafico)+

geom\_point(aes(x=Dim.1, y=Dim.2, colour="darkred"))+

geom\_text\_repel(aes(x=Dim.1,y=Dim.2),

label=datos.grafico$provincia)+

geom\_vline(xintercept=0,colour="darkgray")+

geom\_hline(yintercept=0,colour="darkgray")+

labs(x="Dimension 1 (78%)",y="Dimension 2 (12,1%)")+

theme(legend.position="none")

**#REPRESENTACION CONJUNTA: BIPLOT**

#Grafica1

remotes::install\_github('vqv/ggbiplot', force=TRUE)

library(ggbiplot)

library(plyr)

library(scales)

library(grid)

ggbiplot(fit)+

scale\_color\_discrete(name='')+

expand\_limits(x=c(-3.5,1.5),y=c(-2,2))+

labs(x="Dimension 1 (78%)",y="Dimension 2 (12,1%)")+

geom\_text\_repel(label=AHORA12$Provincia,size=3)

#Grafica2

fviz\_pca\_biplot(fit,

geom.ind = "point",

fill.ind = AHORA12$Provincia, col.ind = "black",

pointshape = 21, pointsize = 2,

palette = "jco",

addEllipses = TRUE,

alpha.var ="contrib", col.var = "contrib",

gradient.cols = "RdYlBu",

legend.title = list(fill = "Provincia", color = "Contrib",

alpha = "Contrib"))

#exporto tabla de descriptivas

write.csv2(x = Descriptivas,

file = "Descriptivas.csv",

row.names = TRUE)

1. Presentación de resultados y conclusiones
2. Análisis estadístico completo

*Solicitud de descriptivas*

Usando la función “stat.desc” proporciona el n total, el número de valores nulos, el número de valores na, mín., máx., rango, suma, mediana, media, SE de la media, IC del 95 % de la media, var, desviación estándar y coeficiente de var.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Estadistica** | **Alimentos** | **Anteojos** | **Artefactos de iluminación** | **Artículos de Librería** | **Balnearios** |
| **nbr.val** | 24,00 | 24,00 | 24,00 | 24,00 | 24,00 |
| **nbr.null** | - | - | - | - | 4,00 |
| **nbr.na** | - | - | - | - | - |
| **min** | 1.532.952,00 | 9.619.490,00 | 1.697.339,00 | 237.033,00 | - |
| **max** | 1.752.655.342,00 | 3.523.790.947,00 | 1.527.268.942,00 | 527.885.672,00 | 116.818.983,00 |
| **range** | 1.751.122.390,00 | 3.514.171.457,00 | 1.525.571.603,00 | 527.648.639,00 | 116.818.983,00 |
| **sum** | 5.855.904.792,00 | 11.777.327.606,00 | 3.519.954.271,00 | 1.876.068.611,00 | 132.000.914,00 |
| **median** | 33.301.755,00 | 158.546.139,50 | 32.380.451,50 | 23.502.111,00 | 110.954,50 |
| **mean** | 243.996.033,00 | 490.721.983,58 | 146.664.761,29 | 78.169.525,46 | 5.500.038,08 |
| **SE.mean** | 94.090.999,07 | 178.476.891,76 | 71.590.373,63 | 27.768.291,69 | 4.858.340,13 |
| **CI.mean** | 194.642.061,29 | 369.207.580,42 | 148.095.971,24 | 57.443.087,93 | 10.050.242,29 |
| **var** | 2,12475E+17 | 7,64496E+17 | 1,23004E+17 | 1,85059E+16 | 5,66483E+14 |
| **std.dev** | 460.949.874,20 | 874.354.631,38 | 350.719.771,78 | 136.036.291,34 | 23.800.908,65 |
| **coef.var** | 1,89 | 1,78 | 2,39 | 1,74 | 4,33 |
|  | **Bicicletas** | **Calzado y Marroquinería** | **Colchones** | **Computadoras** | **Electrodomésticos** |
| **nbr.val** | 24,00 | 24,00 | 24,00 | 24,00 | 24,00 |
| **nbr.null** | - | - | - | - | - |
| **nbr.na** | - | - | - | - | - |
| **min** | 10.483.257,00 | 36.030.939,00 | 15.823.297,00 | 2.470.832,00 | 480.252.738,00 |
| **max** | 3.370.429.454,00 | 12.125.749.621,00 | 6.948.803.372,00 | 4.779.204.220,00 | 133.278.125.554,00 |
| **range** | 3.359.946.197,00 | 12.089.718.682,00 | 6.932.980.075,00 | 4.776.733.388,00 | 132.797.872.816,00 |
| **sum** | 13.084.443.985,00 | 42.324.435.927,00 | 17.394.717.570,00 | 9.087.476.775,00 | 288.042.351.189,00 |
| **median** | 194.979.511,00 | 371.277.322,50 | 82.717.250,50 | 46.920.344,00 | 2.014.109.300,00 |
| **mean** | 545.185.166,04 | 1.763.518.163,63 | 724.779.898,75 | 378.644.865,63 | 12.001.764.632,88 |
| **SE.mean** | 166.579.302,13 | 685.142.530,20 | 373.307.361,66 | 204.322.712,04 | 6.041.389.057,39 |
| **CI.mean** | 344.595.541,08 | 1.417.325.309,33 | 772.245.114,72 | 422.673.733,25 | 12.497.565.451,07 |
| **var** | 6,65968E+17 | 1,12661E+19 | 3,3446E+18 | 1,00195E+18 | 8,75961E+20 |
| **std.dev** | 816.068.583,84 | 3.356.499.200,15 | 1.828.825.106,57 | 1.000.972.774,73 | 29.596.641.056,48 |
| **coef.var** | 1,50 | 1,90 | 2,52 | 2,64 | 2,47 |
|  | **Equipamiento médico** | **Indumentaria** | **Instrumentos musicales** | **Juguetes** | **Libros** |
| **nbr.val** | 24,00 | 24,00 | 24,00 | 24,00 | 24,00 |
| **nbr.null** | 6,00 | - | 1,00 | - | - |
| **nbr.na** | - | - | - | - | - |
| **min** | - | 258.325.348,00 | - | 9.055.750,00 | 150.777,00 |
| **max** | 384.949.091,00 | 86.996.518.988,00 | 406.601.965,00 | 4.100.539.259,00 | 1.638.726.668,00 |
| **range** | 384.949.091,00 | 86.738.193.640,00 | 406.601.965,00 | 4.091.483.509,00 | 1.638.575.891,00 |
| **sum** | 776.580.611,00 | 232.001.012.991,00 | 1.061.696.469,00 | 10.757.472.755,00 | 3.911.746.361,00 |
| **median** | 471.181,50 | 2.287.253.268,50 | 8.050.853,50 | 97.976.307,50 | 14.316.591,50 |
| **mean** | 32.357.525,46 | 9.666.708.874,63 | 44.237.352,88 | 448.228.031,46 | 162.989.431,71 |
| **SE.mean** | 17.824.240,66 | 4.382.902.294,94 | 19.562.469,97 | 222.290.525,69 | 79.226.562,34 |
| **CI.mean** | 36.872.251,09 | 9.066.724.188,14 | 40.468.052,39 | 459.842.987,69 | 163.892.631,13 |
| **var** | 7,62489E+15 | 4,61036E+20 | 9,18457E+15 | 1,18591E+18 | 1,50644E+17 |
| **std.dev** | 87.320.589,33 | 21.471.748.430,14 | 95.836.139,08 | 1.088.996.725,19 | 388.129.303,61 |
| **coef.var** | 2,70 | 2,22 | 2,17 | 2,43 | 2,38 |
|  | **Máquinas y Herramientas** | **Materiales para la construcción** | **Medicamentos** | **Motocicletas** | **Muebles** |
| **nbr.val** | 24,00 | 24,00 | 24,00 | 24,00 | 24,00 |
| **nbr.null** | - | - | - | - | - |
| **nbr.na** | - | - | - | - | - |
| **min** | 2.248.675,00 | 114.965.652,00 | 158.389,00 | 2.962.720,00 | 28.641.943,00 |
| **max** | 2.721.641.666,00 | 35.988.846.235,00 | 4.261.479.275,00 | 3.241.796.843,00 | 13.115.870.722,00 |
| **range** | 2.719.392.991,00 | 35.873.880.583,00 | 4.261.320.886,00 | 3.238.834.123,00 | 13.087.228.779,00 |
| **sum** | 6.480.022.535,00 | 93.455.081.557,00 | 10.691.551.035,00 | 10.087.597.985,00 | 39.378.884.316,00 |
| **median** | 44.590.333,50 | 1.057.624.604,00 | 20.892.491,00 | 52.089.976,50 | 412.338.300,50 |
| **mean** | 270.000.938,96 | 3.893.961.731,54 | 445.481.293,13 | 420.316.582,71 | 1.640.786.846,50 |
| **SE.mean** | 124.393.676,12 | 1.535.051.666,63 | 238.742.626,44 | 178.892.253,96 | 651.955.686,75 |
| **CI.mean** | 257.327.924,79 | 3.175.496.312,55 | 493.876.751,12 | 370.066.822,60 | 1.348.673.093,05 |
| **var** | 3,71371E+17 | 5,65532E+19 | 1,36795E+18 | 7,68059E+17 | 1,02011E+19 |
| **std.dev** | 609.402.067,43 | 7.520.186.624,08 | 1.169.595.229,27 | 876.389.482,28 | 3.193.917.534,88 |
| **coef.var** | 2,26 | 1,93 | 2,63 | 2,09 | 1,95 |
|  | **Neumaticos** | **Perfumeria** | **Servicios de cuidado personal** | **Servicios de instalacion de alarmas** | **Servicios de organizacion de eventos** |
| **nbr.val** | 24,00 | 24,00 | 24,00 | 24,00 | 24,00 |
| **nbr.null** | - | - | - | 20,00 | 3,00 |
| **nbr.na** | - | - | - | - | - |
| **min** | 68.042.362,00 | 5.853.367,00 | 43.417,00 | - | - |
| **max** | 17.810.095.997,00 | 2.049.789.998,00 | 187.609.438,00 | 371.848,00 | 6.453.740,00 |
| **range** | 17.742.053.635,00 | 2.043.936.631,00 | 187.566.021,00 | 371.848,00 | 6.453.740,00 |
| **sum** | 45.719.557.008,00 | 6.801.244.125,00 | 414.734.593,00 | 571.205,00 | 18.147.675,00 |
| **median** | 389.409.548,50 | 75.050.159,50 | 3.319.736,00 | - | 96.878,50 |
| **mean** | 1.904.981.542,00 | 283.385.171,88 | 17.280.608,04 | 23.800,21 | 756.153,13 |
| **SE.mean** | 791.214.808,30 | 112.469.505,35 | 8.915.703,01 | 16.984,71 | 313.156,74 |
| **CI.mean** | 1.636.752.534,66 | 232.660.898,18 | 18.443.536,89 | 35.135,55 | 647.814,06 |
| **var** | 1,50245E+19 | 3,03585E+17 | 1,90775E+15 | 6923528207 | 2,35361E+12 |
| **std.dev** | 3.876.145.114,53 | 550.985.799,45 | 43.677.846,17 | 83.207,74 | 1.534.148,42 |
| **coef.var** | 2,03 | 1,94 | 2,53 | 3,50 | 2,03 |
|  | **Servicios deportivos** | **Servicios educativos** | **Servicios tecnicos** | **Talleres de reparacion** | **Turismo** |
| **nbr.val** | 24,00 | 24,00 | 24,00 | 24,00 | 24,00 |
| **nbr.null** | 4,00 | 13,00 | 2,00 | 13,00 | - |
| **nbr.na** | - | - | - | - | - |
| **min** | - | - | - | - | 1.978.684,00 |
| **max** | 3.402.822.030,00 | 19.362.234,00 | 1.106.496.234,00 | 6.111.479,00 | 36.447.813.136,00 |
| **range** | 3.402.822.030,00 | 19.362.234,00 | 1.106.496.234,00 | 6.111.479,00 | 36.445.834.452,00 |
| **sum** | 5.113.840.498,00 | 44.937.310,00 | 1.471.025.006,00 | 8.114.996,00 | 56.384.176.786,00 |
| **median** | 1.211.226,50 | - | 5.111.850,00 | - | 389.680.593,50 |
| **mean** | 213.076.687,42 | 1.872.387,92 | 61.292.708,58 | 338.124,83 | 2.349.340.699,42 |
| **SE.mean** | 144.948.865,39 | 981.713,80 | 45.837.097,38 | 254.184,76 | 1.507.950.293,15 |
| **CI.mean** | 299.849.573,52 | 2.030.829,72 | 94.821.260,34 | 525.821,25 | 3.119.432.850,05 |
| **var** | 5,04244E+17 | 2,31303E+13 | 5,04249E+16 | 1,55064E+12 | 5,45739E+19 |
| **std.dev** | 710.101.518,02 | 4.809.395,76 | 224.554.999,75 | 1.245.245,95 | 7.387.417.551,39 |
| **coef.var** | 3,33 | 2,57 | 3,66 | 3,68 | 3,14 |

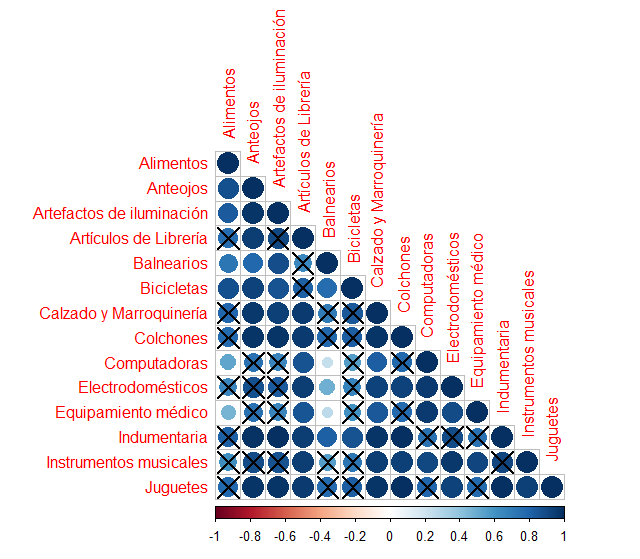
Se puede observar que los máximos valores correspondiente a facturación o consumo en el plan ahora12 de bienes se encuentran en las variables: “Indumentaria”, “Materiales para la construcción” y “Neumáticos”.

En el caso de servicios podemos ver que se encuentran en “Turismo”,” Servicios deportivos” y “Perfumería”.

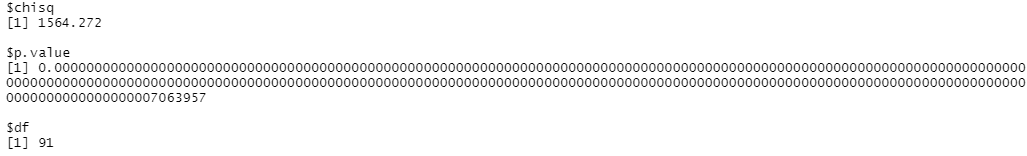
Además, se puede ver que el mayor promedio se encuentra en los “Electrodomésticos” en cuanto a los bienes y para el caso de los servicios “Turismo”.

Los dos mayores desvíos estándar se pueden ver en “Electrodomésticos” y “Turismo”.

*Matriz de correlaciones*

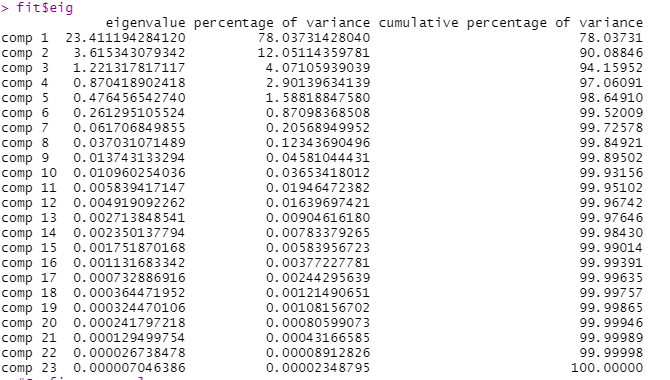


En esta matriz de correlaciones vemos a nivel general que existe una correlación positiva entre todas las variables y la menor correlación entre Electrodomésticos y Balnearios.

*Test de esfericidad de Barlett*

Se Puede observar que como resultado en el test de barlett tenemos un Pvalor 0 que significa que podemos rechazar H0 donde nuestra hipótesis fue que la matriz de correlación es la identidad, eso nos deja constancia de que es aplicable el método y existe intercorrelaciones.

*Numero de componentes principales a extraer*

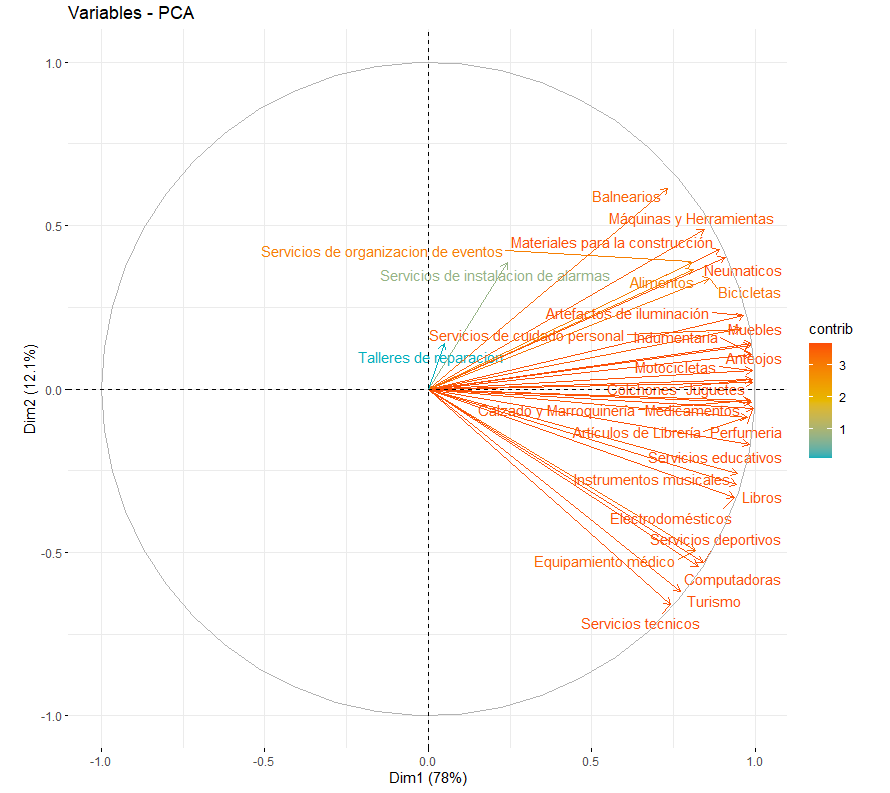


A partir de la obtención de los autovalores vemos que la cantidad de componentes principales a extraer son aquellas que tienen el autovalor mayor a 1, es decir en este caso son 3.

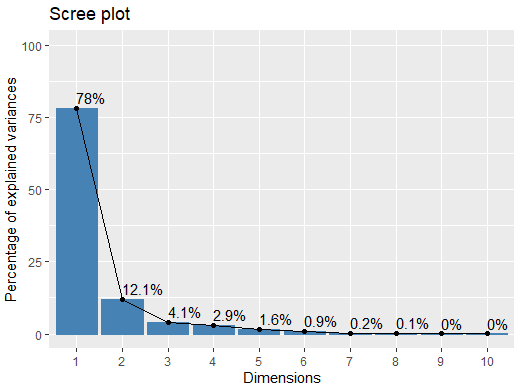
*Las variables mas importantes o contribuyentes se pueden ver en el siguiente gráfico de correlaciones*

Muestra las relaciones entre todas las variables, es decir su carga factorial. Se puede interpretar de la siguiente manera:

* Las variables correlacionadas positivamente se agrupan.
* Las variables correlacionadas negativamente se colocan en lados opuestos del origen del gráfico (cuadrantes opuestos).

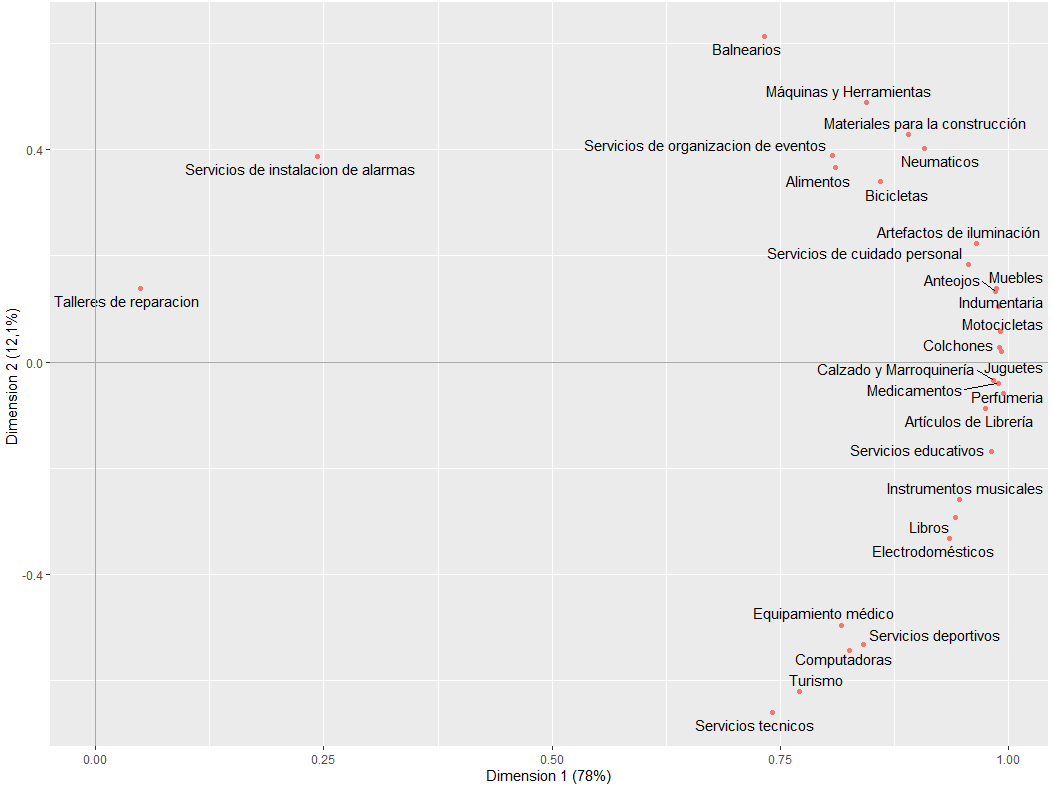


*Gráfico de sedimentación*



Este grafico muestra la cantidad de componentes principales a extraer de forma gráfica, es decir cuando la curva se comienza a aplanar significa que hasta ese punto debemos extraer componentes principales.

*Grafica de coordenadas de las variables*

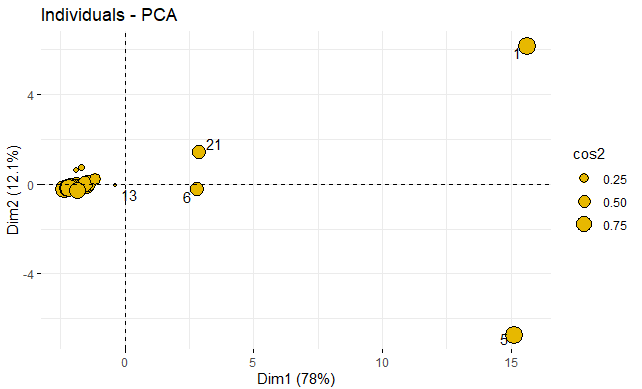


En este grafico se utilizaron las coordenadas de las variables para crear un diagrama de dispersion

*Gráfico calidad de representación*

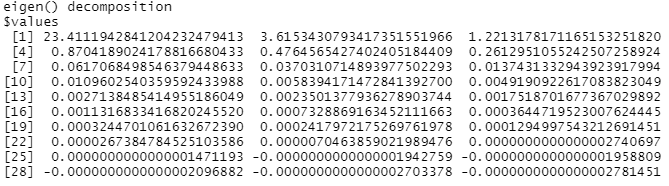
La calidad de representación de las variables en el mapa de factores se llama **cos2** (coseno cuadrado, coordenadas cuadradas).

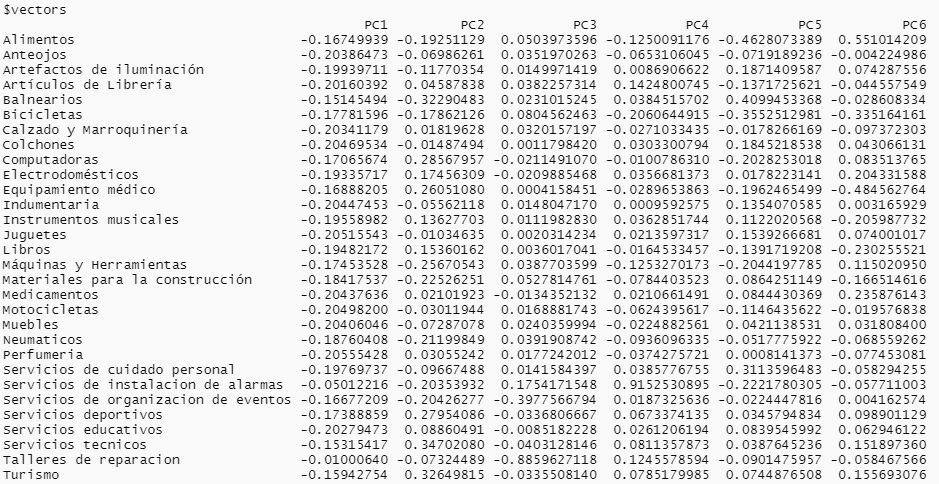
* Un cos2 alto indica una buena representación de la variable en el componente principal. En este caso la variable se posiciona cerca de la circunferencia del círculo de correlación.
* Un cos2 bajo indica que la variable no está perfectamente representada por las PC. En este caso la variable está cerca del centro del círculo.



*Cálculo de autovalores y autovectores a partir de la matriz de correlaciones*

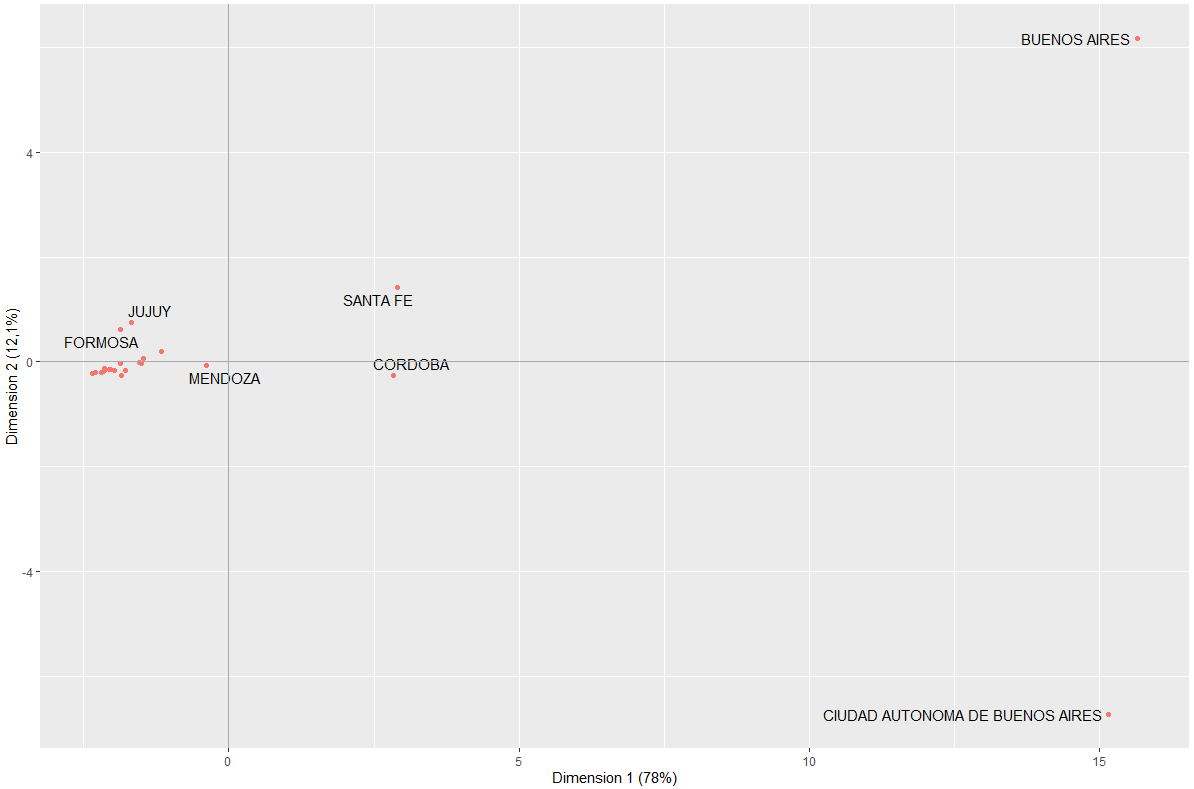
Se extraen los autovalores y los autovectores correspondientes a esos autovalores, anteriormente se había sacado los autovectores para ver la cantidad de componentes principales, donde se eligen aquellas mayores a 1.





En estos resultados, la primera componente principal tiene asociaciones negativas con Anteojos, Calzado, Colchones, Indumentaria, Juguetes, Medicamentos, Motocicletas, Perfumería. Se puede interpretar esta componente principalmente como una medición de los gastos en bienes que se realizaron con el plan ahora12 por cada provincia. El segundo componente tiene asociaciones positivas grandes con Computadoras, Equipamientos médico, Instrumentos musicales, Servicios técnicos, Servicios deportivos, Turismo así que esta componente mide principalmente el consumo o gasto en servicio con el plan ahora12 por cada provincia. La tercera componente tiene una asociación positiva en Servicios de instalación de alarmas y la componente mide principalmente la seguridad adquirida por cada provincia con el plan ahora12.

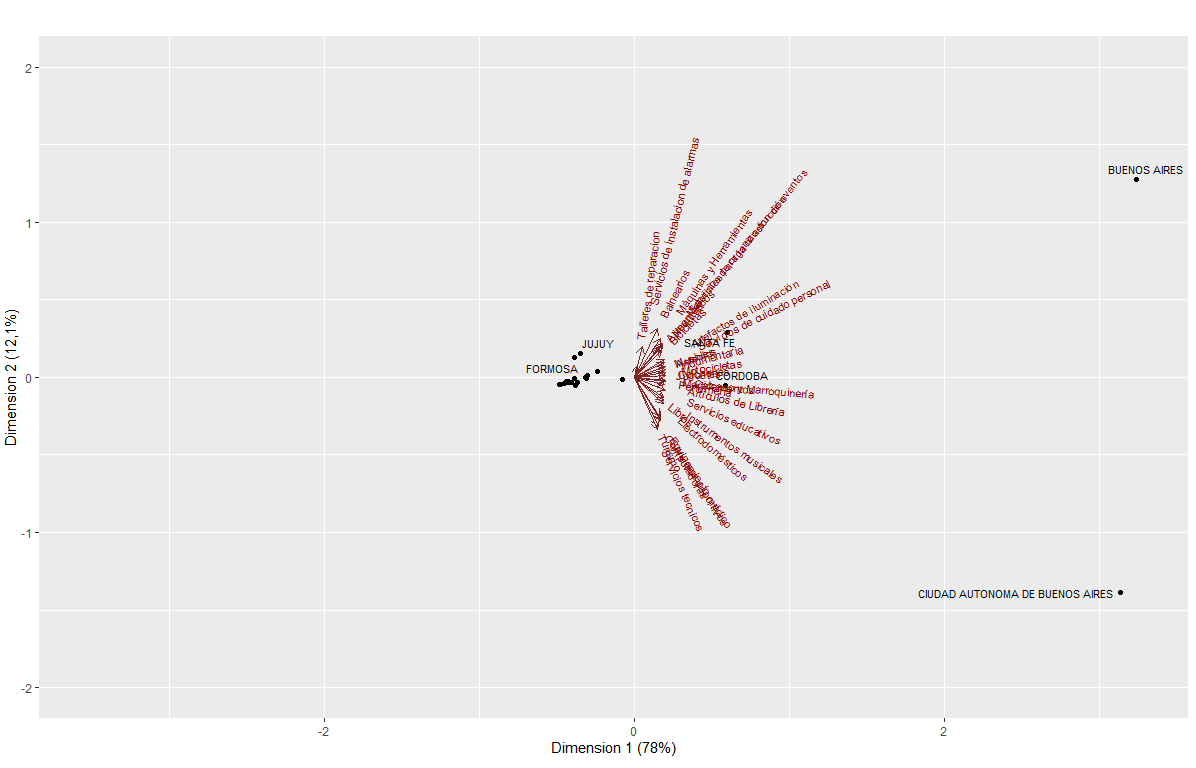
*Representación gráfica de los objetos (provincias)*



Este grafico representa las puntuaciones representa visualmente las puntuaciones del segundo componente principal versus las puntuaciones del primer componente principal.

Si las dos primeras componentes explican la mayor parte de la varianza en los datos, se utiliza la gráfica de puntuaciones para evaluar la estructura de los datos y detectar conglomerados, valores atípicos y tendencias. Si los datos siguen una distribución normal y no hay valores atípicos presentes, los puntos están distribuidos aleatoriamente alrededor de cero. En este caso se ve como los objetos “Buenos Aires” y “Ciudad Autónoma de Buenos Aires” son aquellas con mas peso y es lógico pensar que así sea debido a que existe en éstos mayor cantidad de habitantes[[2]](#footnote-2) , pero también podrían ser analizados como outliers.

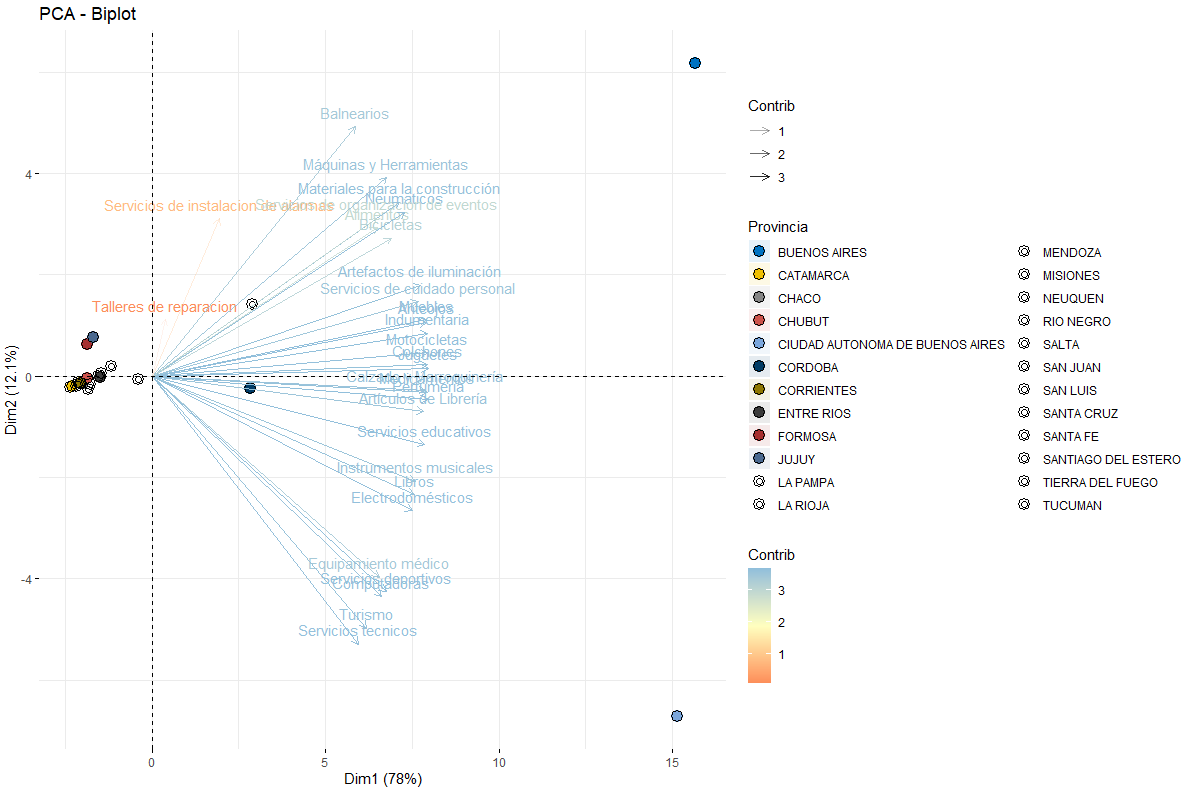
*Representación conjunta (BIPLOT)*



La gráfica de doble proyección evalúa la estructura de los datos y las influencias de los dos primeros componentes en una gráfica. Se grafica las puntuaciones del segundo componente principal versus las puntuaciones del primer componente principal, así como las influencias de ambos componentes.

Aquí podemos observar como todas las variables tienen influencias positivas sobre sobre la primera componente por lo tanto esta componente.

*Gráfico adicional biplot*



* Conclusiones

Para concluir con el análisis del presente trabajo podemos decir que en Buenos Aires y Santa Fe existe una fuerte correlación entre la primera componente que anteriormente habíamos denominado “bienes” y la segunda componente “servicios”, lo que quiere decir que el gasto en bienes esta correlacionado positivamente con servicios.

Y para los otros casos Córdoba y Ciudad Autónoma de Buenos Aires existe una imagen positiva en bienes y negativa en servicios.

* Fuente de datos y Bibliográfica utilizada

# Bibliografía

Análisis de datos multivariantes. Daniel Peña. Editorial McGraw Hill, Madrid, 2002.

# Referencias

<https://www.argentina.gob.ar/pais/poblacion/proyecciones>

[*https://datos.gob.ar/dataset/produccion-programa-ahora-12*](https://datos.gob.ar/dataset/produccion-programa-ahora-12)

1. https://datos.gob.ar/dataset/produccion-programa-ahora-12 [↑](#footnote-ref-1)
2. https://www.argentina.gob.ar/pais/poblacion/proyecciones [↑](#footnote-ref-2)