Ejemplo (datos reales: 34 Estados sobre los que se observan variables socio-demográficas y económicas)

Objetivos:

1. **Ayudar a comprender los conceptos relacionados con un análisis de componentes principales. Interpretación de resultados.**
2. **Aprender a ejecutar con R el ACP. Familiarizarse con los términos y funciones ligadas a un ACP.**

Objetivo del ACP:
Datos
Exploración previa al análisis ACP

Grafico caja
Correlaciones observadas entre pares de variables:
Determinante de la matriz de correlaciones

ACP
Función prcomp()
Función princomp()

Puntuaciones de las observaciones en las componentes
Gráfico de dispersión de las componentes Y1 e Y2:

Análisis mediante extracción de unas pocas componentes:

Matriz de componentes C
Gráficos de componentes
La matriz de correlaciones reproducidas y la matriz residual
Matriz de comunalidades

Rotaciones de la solución
La función varimax()
Uso de la matriz de componentes de la solución no rotada, C, como input en la función varimax
Matriz de componentes rotados (C^R):
Tabla de coeficientes para el cálculo de las puntuaciones en las componentes:

Ejemplo 2: (archivo estdo3)

En el conjunto constituido por 34 Estados del mundo se han observado las 11 variables siguientes.

El archivo **estdo3** contiene dichas variables (las cuales se han estandarizado previo uso):

Variables estandarizadas (Z)

Ztlibrop: número de libros publicados
Ztejerci: Cociente entre el número de individuos en ejército de tierra y población total del estado.
Ztpobact: cociente entre población activa y total
Ztenergi: tasa de consumo energético
Zpservi: Población del sector servicios
Zpagricu: Población del sector agrícola
Ztmedico: Tasa de médicos por habitante
Zespvida: Esperanza de vida
Ztminfan: tasa de mortalidad infantil
Zpobdens:Densidad de población
Zpoburb: Porcentaje de población urbana

Tomadas para 34 Estados del mundo.

Objetivo del análisis:

1. Seleccionar un número pequeño de componentes que resuma las 11 variables observadas en unas pocas dimensiones latentes, procurando que la información perdida no sea de mucha importancia.

2. Intentar interpretar las componentes principales derivadas del análisis

Datos

```
>#  11 variables Ztlibrop a Zpoburb observadas en 34 Estados del mundo. No olvide usar separador decimal ,
#ACP estados3
a=read.table("estdo3.DAT", header=T, sep="\t",dec=",")
x=a[,-1]
row.names(x)=as.character(a$pais)
```

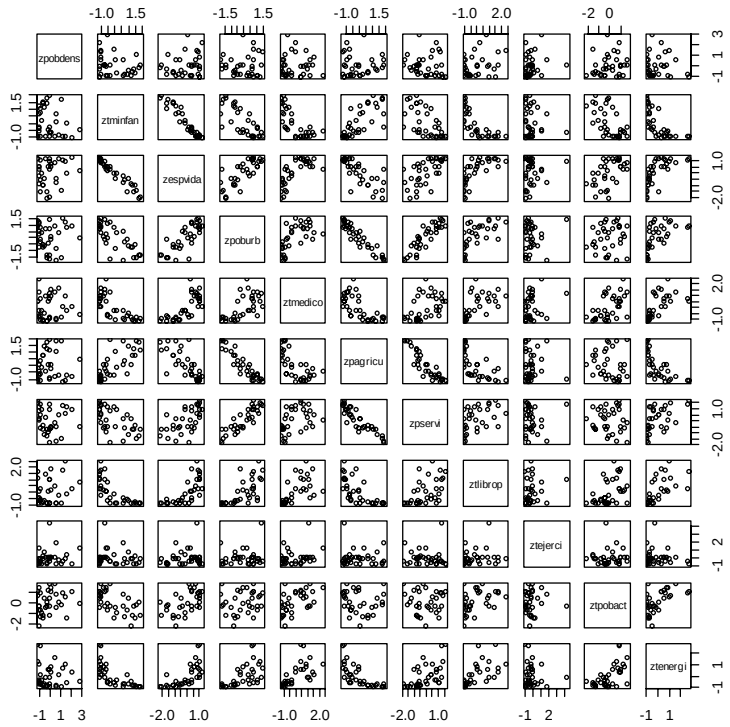
Dado que las variables presentan unidades de medida muy diversas así como varianzas muy distintas, se han estandarizado previamente para evitar que algunas puedan anular o minimizar los efectos de otras. El análisis ACP se basará, por tanto, en dichas variables. O lo que es igual, ACP de matriz de correlaciones de variables sin estandarizar.

Exploración previa al análisis ACP

La matriz de correlaciones observadas entre las variables, R, presenta en sus elementos fuera de la diagonal, lo que podríamos llamar información redundante o compartida por pares de variables originales.

Otro estadístico que sirve de indicador de la adaptación de los datos a la aplicación de la técnica es el determinante de la matriz de correlaciones. Su valor es muy bajo.

>#Exploración de los datos
>plot(x)



>cor(x,use = "pairwise") #Hay algún dato missing, por lo usamos la opción “pairwise”
> round(cor(x,use = "pairwise"),3)

```
          zpobdens ztminfan zespvida zpoburb ztmedico zpagricu zpservi ztlibrop
zpobdens     1.000    -0.224     0.161   0.040    -0.014     0.019  -0.095     0.271
ztminfan    -0.224     1.000    -0.967  -0.757    -0.751     0.752  -0.590    -0.736
zespvida     0.161    -0.967     1.000   0.787     0.736    -0.754   0.612     0.712
zpoburb      0.040    -0.757     0.787   1.000     0.635    -0.938   0.890     0.667
ztmedico    -0.014    -0.751     0.736   0.635     1.000    -0.675   0.445     0.621
zpagricu     0.019     0.752    -0.754  -0.938    -0.675     1.000  -0.907    -0.672
zpservi     -0.095    -0.590     0.612   0.890     0.445    -0.907   1.000     0.509
ztlibrop     0.271    -0.736     0.712   0.667     0.621    -0.672   0.509     1.000
ztejerci     0.144    -0.108     0.126   0.104     0.231    -0.035  -0.009     0.156
ztpobact     0.189    -0.603     0.541   0.155     0.534    -0.147  -0.056     0.426
ztenergi    -0.090    -0.695     0.667   0.582     0.651    -0.697   0.559     0.647
          ztejerci ztpobact ztenergi
zpobdens     0.144     0.189    -0.090
ztminfan    -0.108    -0.603    -0.695
zespvida     0.126     0.541     0.667
zpoburb      0.104     0.155     0.582
ztmedico     0.231     0.534     0.651
zpagricu    -0.035    -0.147    -0.697
zpservi     -0.009    -0.056     0.559
ztlibrop     0.156     0.426     0.647
ztejerci     1.000     0.026    -0.105
ztpobact     0.026     1.000     0.598
ztenergi    -0.105     0.598     1.000
```

Determinante de la matriz de correlaciones

```
>det(cor(x,use = "pairwise"))#valores bajos son indicio de existencia de correlaciones entre variables

 [1] 1.191599e-06
```

Análisis de componentes principales princomp(): ACP

Función prcomp()

```
> acp=princomp(~.,data=x,na.action=na.exclude,cor=T)#Usamos la opción con fórmula. Hay algún NAs.
> summary(acp)
```

Se desea obtener las componentes principales (**Y**) o combinaciones lineales de coeficientes, $\mathbf{A}_{11x11}$, de las variables originales (**Z**), tal que **Y = ZA** con $\mathbf{Y}_{34x11}$ (componentes principales) y $\mathbf{Z}_{34x11}$ (variables originales).
Así, por ejemplo, $y_1 = Za_1$ se obtiene de modo que alcanza la mayor varianza posible; es decir

$$V(Y_1) = \frac{1}{N-1} Y'_1 Y_1 = \frac{1}{N-1} a'_1 Z'Za_1 = a'_1 Ra_1 = \lambda_1 = 6{,}103$$

(véase el autovalor correspondiente a la componente **1** en tabla **Varianza total explicada**).
La varianza del conjunto de variables observadas (Z) proyectada sobre el vector a_1 es λ_1 = **6,103**
La primera componente o combinación lineal, con el mayor valor propio, es la que mejor resume la información contenida en los datos.

La segunda componente, que debe maximizar la varianza después de la extracción de Y_1, se obtendría calculando el vector a_2 tal que:
$y_2 = Za_2$ e Y_2 está incorrelacionado con Y_1

$$V(Y_2) = (1/N)Y'_2 Y_2 = (1/N)\, a'_2 Z'Za_2 = a'_2 Ra_2 = \lambda_2 = \mathbf{1{,}617}$$

$$a'_2 a_2 = 1$$

De modo similar, la tercera componente, Y_3, presenta varianza igual a **1,202**.

```
> acp=princomp(~.,data=x,na.action=na.exclude,cor=T)#equivale a estandarizar
```

> summary(acp)

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4    Comp.5
Standard deviation     2.4703503 1.2715183 1.0963014 0.96165926 0.6295387
Proportion of Variance 0.5547846 0.1469781 0.1092615 0.08407168 0.0360290
Cumulative Proportion  0.5547846 0.7017627 0.8110242 0.89509589 0.9311249
                           Comp.6     Comp.7     Comp.8      Comp.9     Comp.10
Standard deviation     0.5364019 0.49001802 0.34597391 0.230304892 0.199295299
Proportion of Variance 0.0261570 0.02182888 0.01088163 0.004821849 0.003610783
Cumulative Proportion  0.9572819 0.97911077 0.98999240 0.994814247 0.998425031
                          Comp.11
Standard deviation     0.131623180
Proportion of Variance 0.001574969
Cumulative Proportion  1.000000000
```

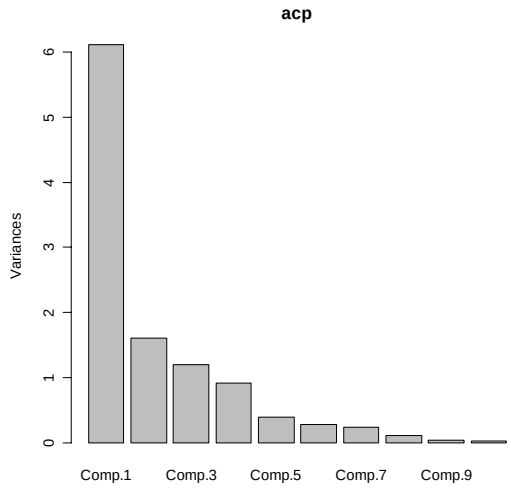
Las 3 primeras componentes tienen varianza superior a 1, tal como muestra el resultado. La cuarta componente presenta un valor muy próximo a 1 (con un valor propio o varianza igual a 0.92):

Valores propios o varianzas de las componentes:

> summary(acp)$sd^2

```
    Comp.1     Comp.2     Comp.3     Comp.4     Comp.5     Comp.6     Comp.7
6.10263079 1.61675875 1.20187676 0.92478853 0.39631896 0.28772697 0.24011766
    Comp.8     Comp.9    Comp.10    Comp.11
0.11969795 0.05304034 0.03971862 0.01732466
```

>plot(acp)



Cada k-ésimo valor propio o **autovalor** (k=1, ..,11) se interpreta como la parte de la varianza que el k-ésimo **eje principal** (o sea, la correspondiente componente principal) explica. Y el cociente **autovalor / p**, como la proporción correspondiente a dicha componente; muestra, en consecuencia, la importancia de esta componente en el conjunto.

En particular, para la primera componente tenemos: 6,103 / 11 = 0,55478 (véase **% de la varianza** en **Varianza total explicada**)

Matriz de vectores propios:

Los vectores propios de la matriz de correlaciones de X con la función **princomp** son las columnas de la matriz Loadings

> loadings(acp)# Vectores propios: coeficientes de combinaciones lineales que proporcionan componentes

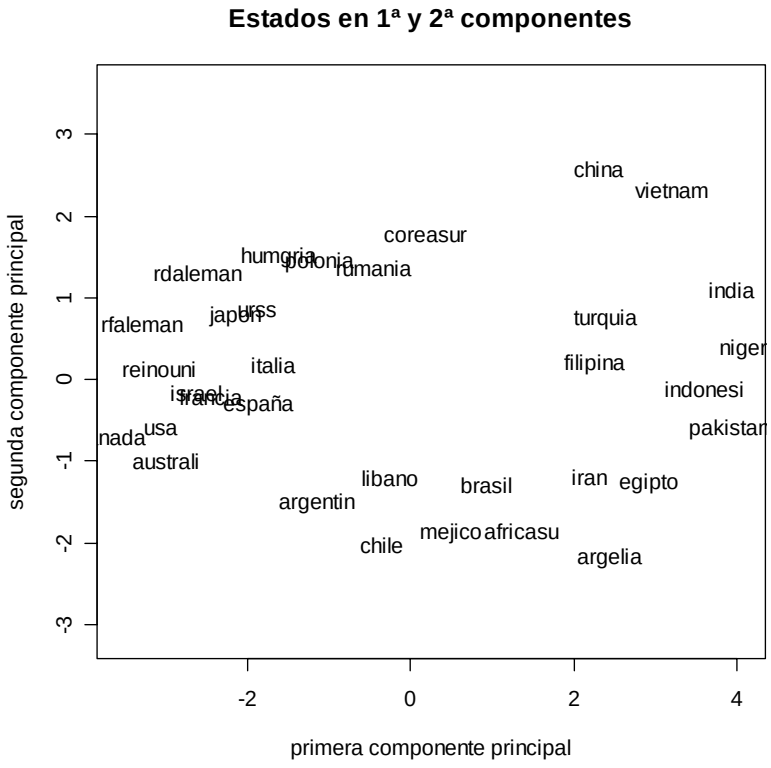
```
Loadings:
         Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
zpobdens         0.422  0.489  0.639                -0.396
ztminfan  0.376 -0.162                -0.371 -0.136 -0.184  0.287  0.165 -0.440
zespvida -0.373  0.111                 0.443  0.167  0.296 -0.315  0.127 -0.425
zpoburb  -0.359 -0.265  0.172         0.128                0.534  0.534 -0.243
ztmedico -0.328  0.135        -0.350  0.121 -0.762 -0.300         -0.179 -0.166
zpagricu  0.366  0.293                                0.211         -0.229 -0.658
zpservi  -0.306 -0.462  0.112  0.104         0.260 -0.181  0.159 -0.708 -0.200
ztlibrop -0.332  0.147  0.120  0.132 -0.607 -0.226  0.631         -0.108
ztejerci         0.198  0.643 -0.649 -0.148  0.303
ztpobact -0.193  0.576 -0.368                0.270 -0.106  0.579 -0.137  0.124
ztenergi -0.325        -0.384        -0.486  0.281 -0.371 -0.402  0.233 -0.211
         Comp.11
zpobdens
ztminfan  0.579
zespvida  0.490
zpoburb  -0.349
ztmedico
zpagricu -0.485
zpservi
ztlibrop
ztejerci
ztpobact  0.183
ztenergi -0.174

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.091  0.091  0.091  0.091  0.091  0.091  0.091  0.091  0.091
Cumulative Var  0.091  0.182  0.273  0.364  0.455  0.545  0.636  0.727  0.818
               Comp.10 Comp.11
SS loadings      1.000   1.000
Proportion Var   0.091   0.091
Cumulative Var   0.909   1.000
```

Gráfico de dispersión de las componentes Y1 e Y2:

Pueden realizarse gráficos con las puntuaciones en las componentes principales y visualizar qué casos destacan en determinadas componentes:

```
> require(MASS)
> eqscplot(predict(acp)[,1:2],type="n",xlab="primera componente principal"
+ ,ylab="segunda componente principal")#Gráfico que usa la misma escala en los ejes
> text(predict(acp) [,1:2],labels=row.names(x))
> title(main="Estados en 1ª y 2ª componentes")
```



Análisis con extracción de 3 componentes

Sólo tres componentes capturan una variabilidad total del 81% (columna **% acumulado**). Esto supone que se puede reducir la dimensionalidad de los datos al pasar de 11 variables observadas a trabajar con sólo 3, sin distorsionar demasiado la información inicial (habrá 19% de variabilidad en los datos originales del que las tres componentes extraídas no pueden dar cuenta). En sólo 3 dimensiones puede registrarse el 81% de la variabilidad original, de modo que los tres factores o componentes explican el 81% de la variabilidad total.

En principio, se prescinde de las componentes asociadas a los valores propios con autovalores inferiores a 1. No obstante, dado que el cuarto autovalor está próximo a 1, convendrá examinar también las posibles ventajas e inconvenientes de su inclusión[1].

La matriz de componentes o matriz factorial

Un output usual en ACP desde la perspectiva del análisis factorial, que es más ilustrativo que la matriz de vectores propios, es la **matriz de componentes** (o **matriz factorial**). La

[1] Aumentar el número de componentes supone aumentar la dimensionalidad de la información resumida en las componentes. No obstante, a veces, un subgrupo de variables importantes podría no quedar bien representado si se omite una componente que recoge la variabilidad del mismo.

relación entre una y otra es simplemente para cada columna un factor de proporcionalidad igual a la raíz del correspondiente valor propio.

Sus valores representan **correlaciones** entre cada una de las variables y cada una de las componentes. Las filas representan las variables y las columnas, las componentes.
Estos coeficientes reciben también el nombre de **pesos**, ponderaciones, **saturaciones**[2] o **cargas factoriales (factor loading)**.

Siguiendo el criterio de Kaiser seleccionaremos las componentes cuyos autovalores superen el valor 1. En este caso, hay 3.

La **matriz de componentes** contiene las **correlaciones** entre las **variables originales (X)** y las **componentes** (Y). Se puede obtener, por tanto con la función cor, cor(X,Y)

Tomamos las 3 primeras componentes (81% de variabilidad)

```
> C=cor(x,predict(acp)[,1:3], use = "pairwise") #Matriz de componentes
> C
              Comp.1      Comp.2       Comp.3
zpobdens -0.08466927  0.53649427  0.535579714
ztminfan  0.92773936 -0.20556883 -0.002878376
zespvida -0.92103278  0.14078899  0.021249026
zpoburb  -0.88727602 -0.33731703  0.188105631
ztmedico -0.80960016  0.17208428 -0.063954881
zpagricu  0.90305814  0.37281485 -0.075169981
zpservi  -0.75511812 -0.58783056  0.123199006
ztlibrop -0.82082940  0.18745282  0.131091878
ztejerci -0.11265240  0.25221115  0.704759043
ztpobact -0.47659619  0.73300090 -0.403427946
ztenergi -0.80276787  0.05368212 -0.421478772
```

Concretamente la **Matriz de componentes C** muestra, por ejemplo, que la tasa de mortalidad infantil (TMINFAN) tiene un coeficiente de correlación con la primera componente igual a 0,928. Las variables TMINFAN y PAGRICUL tienen correlaciones altas con la primera componente de signo positivo, mientras que las variables ESPVIDA, POBURB, TLIBROPU, TMEDICOS, TENERGIA y PSERVI presentan también correlaciones altas pero de distinto signo.
Lógicamente, si alguna de las variables presentara un valor igual a 1, su variabilidad sería explicada totalmente por el factor[3].

Puede considerarse que la suma de los cuadrados de los pesos factoriales por **columna** es una medida de la varianza de la matriz R que viene **explicada** por esa **componente principal**.

[2] Especialmente en el contexto del Análisis Factorial.

[3] Existe solapamiento perfecto o variabilidad total compartida entre ellas.

```
> apply(C*C,2,sum)#Varianzas explicadas por las componentes
  Comp.1   Comp.2   Comp.3
6.102631 1.616759 1.201877
```

$$(-0.08466927)^2 + (0.92773936)^2 + (-0.92103278)^2 + \ldots + (-0.80276787)^2 = \mathbf{6,103}$$

(Recuerde que cada columna j de C es un vector propio multiplicado por la raíz de λ_j).

Esto indica que el valor más alto que puede alcanzar un valor propio o autovalor es **p** (número de variables)[4]

Normalmente, es preciso efectuar una rotación de los ejes en la solución inicial para intentar mejorar la interpretación de las componentes.

Los **gráficos de componentes** visualizan la tabla anterior. Los presentamos en planos, de modo que puedan interpretarse más claramente.

Comunalidad

```
> apply(C*C,1,sum)#Comunalidades
 zpobdens  ztminfan  zespvida   zpoburb  ztmedico  zpagricu   zpservi  ztlibrop
0.5818406 0.9029672 0.8685744 0.9364252 0.6891557 0.9601554 0.9309261 0.7260845
 ztejerci  ztpobact  ztenergi
0.5729863 0.9271884 0.8249624
```

Otro resultado útil es la comunalidad. La **comunalidad** de una variable observada es la **proporción de varianza explicada** por los **factores comunes**. (O coeficiente R^2 suponiendo las variables X_i combinaciones lineales de las componentes principales extraídas)
En ACP la comunalidad inicial es siempre 1. Se supone un número de componentes igual al de variables. Tras la **Extracción** de un número k menor de componentes, la comunalidad de cada variable es la proporción de varianza explicada por las k componentes extraídas; refleja el coeficiente de correlación múltiple al cuadrado de cada variable como si fuera predicha por los k factores o componentes extraídos. Puede observarse que con 3 componentes extraídos, las variables mejor representadas son PAGRICUL, PSERVI y TPOBACTI con valores iguales a 0,960, 0,931 y 0,927, respectivamente. Las peor representadas son TEJERCIT y POBDENS.

Se obtiene a partir de la matriz factorial o matriz de componentes (que notaremos con C). Es igual a la suma[5] de los cuadrados de las ponderaciones factoriales de cada variable (fila). (Véase **Matriz de componentes C**)

---

[4] Correspondería al hipotético caso en que todas las variables tuviesen ponderación igual a uno en el factor o componente.

[5] Dado que las componentes principales no están correlacionadas.

Por ejemplo, la variable TPOBACTI tiene una comunalidad igual a:

$$0{,}927 = 0{,}477^2 + 0{,}733^2 + (-0{,}403)^2$$

Por tanto el 92,7 % de la variabilidad de TPOBACTI viene explicado por las 3 componentes principales. Representa, pues, la varianza que comparte con los factores.
Cuanto más se aproxima a cero la comunalidad de una variable, peor representada es por las componentes. Valores próximos a 1, por el contrario, indican que la variabilidad de la variable queda totalmente representada por las componentes.

La matriz de correlaciones reproducidas y la matriz residual

La matriz R (observada) y la matriz R* (reproducida):

En general, si tomamos las 11 componentes, se obtiene:

RA = AD con A'A = I; o bien, las expresiones equivalentes:
A'RA = D y R = ADA'= C C'

Donde D representa una matriz diagonal (constituida por los autovalores asociados a las distintas componentes). La suma de los elementos diagonales vale 11 (cumpliéndose que traza de R = traza de D = 11). A es la matriz de vectores propios[6]. C es la matriz de componentes que expresa las correlaciones entre variables y componentes.

Otro modo de expresar R = ADA' es mediante:

$R = \lambda_1 a_1 a'_1 + \lambda_2 a_2 a'_2 + ... + \lambda_{11} a_{11} a'_{11}$. (Según (5.1_4)).

Donde a_i es el vector propio px1.

Si despreciamos todas las componentes desde la 4 en adelante, tendremos una aproximación de la matriz R, sumando las tres matrices siguientes:

$R^* = \lambda_1 a_1 a'_1 + \lambda_2 a_2 a'_2 + \lambda_3 a_3 a'_3$
$R^* = 6{,}103$ (matriz $a_1 a'_1$) + 1,617 (matriz $a_2 a'_2$) + 1,202 (matriz $a_3 a'_3$)
O bien multiplicando la matriz de componentes rotados por su traspuesta (CC'):

[5]

---

[6] Puede obtenerse la matriz de vectores propios a partir de la matriz de componentes dividiendo cada columna de dicha matriz por la raíz cuadrada del correspondiente autovalor.

R* = CC’ = ADA’

(véase matriz de **correlación reproducida** en tabla de **correlaciones reproducidas**)

La matriz de residuos, elementos no diagonales de R-R*, nos indica en cierto modo la posible distorsión en la que se hubiera podido incurrir al describir las variables observadas mediante tres componentes Y_1, Y_2 e Y_3.

Matriz residual = R-R* (elementos no diagonales)

(véase matriz **residual** en tabla de **correlaciones reproducidas**)

```
> Rrep=C%*%t(C)#Correlaciones reproducidas
> Rrep
             zpobdens   ztminfan   zespvida      zpoburb   ztmedico    zpagricu
zpobdens  0.581840613 -0.1903791  0.1648962 -0.005098081  0.1266175  0.08329224
ztminfan -0.190379110  0.9029672 -0.8834814 -0.754360463 -0.7862890  0.76137984
zespvida  0.164896206 -0.8834814  0.8685744  0.773716837  0.7685369 -0.78085521
zpoburb  -0.005098081 -0.7543605  0.7737168  0.936425249  0.6482616 -0.94115853
ztmedico  0.126617547 -0.7862890  0.7685369  0.648261582  0.6891557 -0.66215296
zpagricu  0.083292240  0.7613798 -0.7808552 -0.941158532 -0.6621530  0.96015545
zpservi  -0.185449535 -0.5800678  0.6153463  0.891457890  0.5023082 -0.91032840
ztlibrop  0.240276538 -0.8004275  0.7851876  0.689730336  0.6884173 -0.68122565
ztejerci  0.522302678 -0.1583874  0.1542405  0.147447802  0.0895322 -0.06068033
ztpobact  0.217536006 -0.5916780  0.5335867  0.099731616  0.5377915 -0.12679478
ztenergi -0.128965564 -0.7545815  0.7379774  0.614886261  0.6861145 -0.67325002
              zpservi   ztlibrop    ztejerci    ztpobact   ztenergi
zpobdens -0.18544954  0.2402765  0.52230268  0.21753601 -0.1289656
ztminfan -0.58006778 -0.8004275 -0.15838738 -0.59167796 -0.7545815
zespvida  0.61534633  0.7851876  0.15424055  0.53358672  0.7379774
zpoburb   0.89145789  0.6897303  0.14744780  0.09973162  0.6148863
ztmedico  0.50230818  0.6884173  0.08953219  0.53779147  0.6861145
zpagricu -0.91032840 -0.6812257 -0.06068033 -0.12679478 -0.6732500
zpservi   0.93092614  0.5257830  0.02363406 -0.12069583  0.5227028
ztlibrop  0.52578305  0.7260845  0.23213428  0.47572112  0.6137459
ztejerci  0.02363406  0.2321343  0.57298633 -0.04575879 -0.1930680
ztpobact -0.12069583  0.4757211 -0.04575879  0.92718835  0.5919815
ztenergi  0.52270281  0.6137459 -0.19306802  0.59198146  0.8249624
```

#matriz de residuos

>Resi=cor(x,use = "pairwise")-Rrep #diferencias entre la matriz R de correlaciones de los datos y matriz reproducida

```
> Resi
             zpobdens     ztminfan     zespvida      zpoburb     ztmedico
zpobdens  0.418159387 -0.033871059 -0.003661443  0.045341399 -0.140788767
ztminfan -0.033871059  0.097032848 -0.083402044 -0.003134972  0.035370105
zespvida -0.003661443 -0.083402044  0.131425559  0.013429204 -0.032408697
zpoburb   0.045341399 -0.003134972  0.013429204  0.063574751 -0.012899529
ztmedico -0.140788767  0.035370105 -0.032408697 -0.012899529  0.310844346
zpagricu -0.064552521 -0.009218802  0.027317791  0.003136173 -0.012360188
zpservi   0.090303471 -0.010097473 -0.003219363 -0.001442695 -0.057682691
ztlibrop  0.030671070  0.064032998 -0.073197819 -0.022610719 -0.067061136
ztejerci -0.378779240  0.050296321 -0.028360826 -0.043357878  0.141827653
ztpobact -0.028217628 -0.011569016  0.007612791  0.055156858 -0.004143106
ztenergi  0.038515625  0.060026706 -0.070838157 -0.032860577 -0.035444366
             zpagricu      zpservi    ztlibrop    ztejerci     ztpobact
zpobdens -0.064552521  0.090303471  0.03067107 -0.37877924 -0.028217628
ztminfan -0.009218802 -0.010097473  0.06403300  0.05029632 -0.011569016
zespvida  0.027317791 -0.003219363 -0.07319782 -0.02836083  0.007612791
zpoburb   0.003136173 -0.001442695 -0.02261072 -0.04335788  0.055156858
ztmedico -0.012360188 -0.057682691 -0.06706114  0.14182765 -0.004143106
zpagricu  0.039844554  0.003109947  0.00909341  0.02582779 -0.020355893
zpservi   0.003109947  0.069073862 -0.01661175 -0.03270851  0.064534414
ztlibrop  0.009093410 -0.016611749  0.27391546 -0.07636061 -0.049238639
ztejerci  0.025827791 -0.032708509 -0.07636061  0.42701367  0.071893185
ztpobact -0.020355893  0.064534414 -0.04923864  0.07189318  0.072811647
ztenergi -0.024009374  0.035943782  0.03309585  0.08758459  0.006090014
             ztenergi
zpobdens  0.038515625
ztminfan  0.060026706
zespvida -0.070838157
zpoburb  -0.032860577
```

```
ztmedico -0.035444366
zpagricu -0.024009374
zpservi   0.035943782
ztlibrop  0.033095849
ztejerci  0.087584585
ztpobact  0.006090014
ztenergi  0.175037620
```

Hay 16 residuos no redundantes con valor absoluto mayor a 0.05

Puntuaciones

Nota: Se pueden obtener las puntuaciones en las componentes **como variables** disponibles usar en otras aplicaciones. Las puntuaciones se dan **estandarizadas**. Observe que si multiplicamos la matriz de datos Z por la de valores propios, A, se obtienen las puntuaciones centradas pero con varianza igual a cada correspondiente valor propio, tal como se ha expresado en párrafos anteriores. Para obtener las puntuaciones estandarizadas es preciso dividir cada columna Y_i por la raíz cuadrada del autovalor i-ésimo; una opción para obtener las puntuaciones tipificadas es mediante:

$$Y_{nxk} D_{kxk}^{-1/2} = Z_{nxp} A_{pxk} D_{kxk}^{-1/2}$$

Donde $D^{-1/2} = Diag\left[\frac{1}{\sqrt{\lambda_i}}\right]$ es la matriz diagonal que contiene como elementos diagonales los inversos de las raíces cuadradas de los k primeros valores propios.

Puede proporcionarse además la matriz de coeficientes para el cálculo de las puntuaciones en las componentes (Véase más adelante el cálculo de **las puntuaciones en las componentes**).

Rotaciones de la solución

R permite rotar la solución de modo que sea más interpretable.

La función varimax()

La función **varimax** proporciona los **loadings** rotados (obtenidos como el producto de los loadings no rotados por la matriz de rotación) y la **matriz de rotación**.

El input lo constituyen los loadings o la matriz de componentes de la solución no rotada.

Uso de la matriz de componentes de la solución no rotada, C, como input en la función varimax

>rota=varimax(C, normalize = T) # C es la matriz de componentes (correlaciones entre X e Y) en solución no rotada, ya obtenida en párrafos anteriores.

> rota
$loadings

```
Loadings:
         Comp.1 Comp.2 Comp.3
zpobdens  0.101  0.177  0.735
ztminfan  0.641 -0.677 -0.184
zespvida -0.675  0.621  0.167
zpoburb  -0.944  0.201
ztmedico -0.548  0.615
zpagricu  0.950 -0.236
zpservi  -0.958         -0.117
ztlibrop -0.593  0.546  0.277
ztejerci -0.114         0.744
ztpobact  0.102  0.956
ztenergi -0.526  0.690 -0.269

                Comp.1 Comp.2 Comp.3
SS loadings      4.539  3.043  1.339
Proportion Var   0.413  0.277  0.122
Cumulative Var   0.413  0.689  0.811

$rotmat
           [,1]       [,2]        [,3]
[1,]  0.8100587 -0.5799115 -0.08664594
[2,]  0.5409860  0.6821916  0.49188293
[3,] -0.2261394 -0.4453283  0.86633923
```

En un intento de encontrar una solución más interpretable efectuaremos la rotación o giro de los ejes coordenados (que representan a las componentes) tal que las distintas variables, representadas por puntos cuyas coordenadas constituyen los pesos o elementos de la matriz de componentes, "caigan" o se sitúen de forma que se organicen subgrupos claramente definidos y próximos a diferentes ejes[7]

Matriz de componentes rotados (C^R):

Esta matriz se obtiene a partir de la matriz de componentes sin rotar C mediante el producto:

$$C^R = C\,T$$

Donde la matriz T es la constituida por los coeficientes que definen la rotación ejercida (rota$rotmat).

Varimax proporciona directamente la **matriz de componentes rotadas** (rota$loadings).

Loadings:

```
         Comp.1 Comp.2 Comp.3
zpobdens  0.101  0.177  0.735
ztminfan  0.641 -0.677 -0.184
zespvida -0.675  0.621  0.167
zpoburb  -0.944  0.201
ztmedico -0.548  0.615
zpagricu  0.950 -0.236
zpservi  -0.958         -0.117
ztlibrop -0.593  0.546  0.277
ztejerci -0.114         0.744
ztpobact  0.102  0.956
ztenergi -0.526  0.690 -0.269
```

La tabla **Matriz de componentes rotados** ofrece un esquema ligeramente distinto de la matriz sin rotar. Es una combinación lineal de la primera y **explica** la **misma** cantidad de la

[7] Los distintos ejes se sitúan próximos a las variables en que están **saturadas**; es decir, junto a aquellas que presentan pesos máximos en dichos ejes.

varianza inicial. Las comunalidades [8]son las mismas. No obstante, cambia la variabilidad capturada por cada componente. (Véase tabla de varianza total explicada. Compare suma de las saturaciones al cuadrado de la extracción y la de la rotación).

```
> apply(rota$loadings^2,1,sum)#Comunalidades

 zpobdens  ztminfan  zespvida   zpoburb  ztmedico  zpagricu   zpservi  ztlibrop
0.5818406 0.9029672 0.8685744 0.9364252 0.6891557 0.9601554 0.9309261 0.7260845
 ztejerci  ztpobact  ztenergi
0.5729863 0.9271884 0.8249624
```

La suma de las saturaciones al cuadrado de las 3 componentes es la misma, pero la variabilidad capturada por cada componente difiere:

```
> apply(rota$loadings^2,2,sum) #Varianzas explicadas por las componentes

  Comp.1   Comp.2   Comp.3
4.539149 3.043067 1.339049
```

Efectivamente, cada componente no explica lo mismo. Por ejemplo, la suma de los valores al cuadrado de la columna primera de la tabla Matriz de componentes rotados es:

$$(0.101)^2 + (0.641)^2 + ... + (-0.526)^2 = \mathbf{4,53}$$

Observe que la suma obtenida al acumular los valores propios es la misma. Lo cual es lógico, porque la dimensión (igual a 3) es la misma en ambas soluciones. Los puntos están ubicados en tres dimensiones en ambos casos.

Interpretación:

Las variables más importantes para definir la variabilidad registrada en la primera componente son PSERVI, PAGRICUL, POBURB, ESPVIDA, TMINFA.

Las variables más correlacionadas con la segunda componente son TPOBACTI, TENERGIA,TMINFAN y TMEDICOS

Las variables más correlacionadas con la tercera componente son TEJERCIT y POBDENS

No estamos ante la situación ideal descrita en párrafos anteriores. Ni siquiera la rotación de ejes ha permitido una estructura simple que permita interpretar las componentes de modo más nítido.

Por ejemplo, las variables ESPVIDA, TLIBROPUB, TENERGIA, TMINFAN y TMEDICOS presentan correlaciones altas con las componentes 1 y 2.

En un esfuerzo por caracterizar las componentes según las relaciones observadas con las variables originales podríamos decir lo siguiente:

Es difícil delinear una composición específica para las componentes 1 y 2. ambas parecen contraponer los países más desarrollados a los menos desarrollados. Esperanza de vida alta,

[8] Geométricamente, podemos imaginar cada variable representada en el espacio tridimensional constituido por las componentes. Sus coordenadas son sus correspondientes correlaciones con ellas. La comunalidad de dicha variable representa la norma o longitud del vector que representa. El giro de los ejes no supone una modificación de la longitud del vector, sólo de las coordenadas del punto respecto a los ejes.

nivel cultural y bienestar social altos frente a bajos (libros publicados, tasa de médicos por habitantes alta frente a baja, tasa de mortalidad infantil baja frente a alta)

Primera componente: Destaca la comparación de países con fuerte sector de servicios y población urbana frente a países con importante sector agrícola. Es decir, cierta estructura que compara sociedades modernas con más primitivas.

Segunda componente: Destaca el potencial humano desde una perspectiva laboral (población activa) y tal vez industrial (consumo energético).

Tercera componente: Indica fundamentalmente **potencial humano**, desde una perspectiva demográfica, reflejado en la densidad de población y el efectivo registrado en su ejército.

Estas componentes podrían usarse en otros estudios como indicadores globales de tipo socioeconómico.

Los gráficos muestran visualmente la información recogida en la tabla **Matriz de componentes rotados** y son de gran ayuda en la interpretación de las componentes:

Representación gráfica de las columnas de la matriz de componentes rotados:

Gráficos de componentes en espacio rotado

Las coordenadas de cada variable vienen dadas en las filas de la tabla **Matriz de componentes rotada** (correlaciones de cada variable con cada componente).

```
plot(rota$loadings[,1],rota$loadings[,2],type="n",xlim=c(-1,1),ylim=c(-1,1))
text(rota$loadings[,1],rota$loadings[,2],labels=row.names(rota$loadings))
title(main="Gráfico de componentes en espacio rotado")
abline(h=0, v=0, col = "gray60")

plot(rota$loadings[,1],rota$loadings[,3],type="n",xlim=c(-1,1),ylim=c(-1,1))
text(rota$loadings[,1],rota$loadings[,3],labels=row.names(rota$loadings))
title(main="Gráfico de componentes en espacio rotado")
abline(h=0, v=0, col = "gray60")

plot(rota$loadings[,3],rota$loadings[,2],type="n",xlim=c(-1,1),ylim=c(-1,1))
text(rota$loadings[,3],rota$loadings[,2],labels=row.names(rota$loadings))
title(main="Gráfico de componentes en espacio rotado")
abline(h=0, v=0, col = "gray60")
```

El gráfico muestra las componentes 1 y 2 rotadas. Las coordenadas de cada variable, tal como hemos dicho, son las correlaciones entre variable y componente. Una variable próxima a un eje y alejada del origen indica alta correlación con la componente.

Por ejemplo, los puntos representados por poburb y pserv (a la izquierda) y pagricul (a la derecha) están altamente correlacionados con la primera componente. Los puntos con coordenadas más altas en la componente 2 son tpobacti y tenergia. Los puntos de correlaciones más bajas en dichas componentes son los más próximos al origen (tejercit y pobdens)

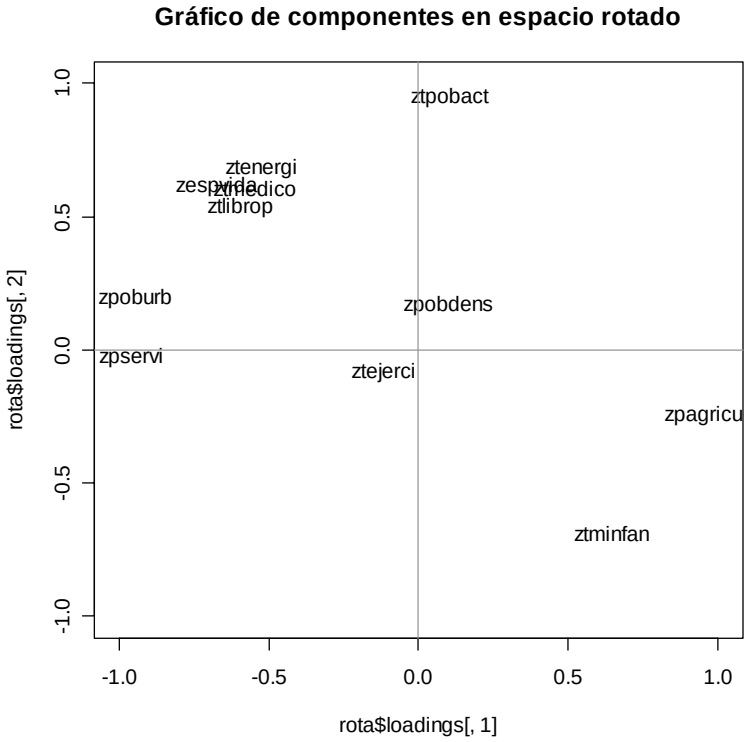


Gráfico de componentes en espacio rotado
ztpobact
ztenergi
zespvida
ztmedico
ztlibrop
zpoburb
zpobdens
zpservi
ztejerci
zpagricu
ztminfan
rota$loadings[, 2]
rota$loadings[, 1]
1.0
0.5
0.0
-0.5
-1.0

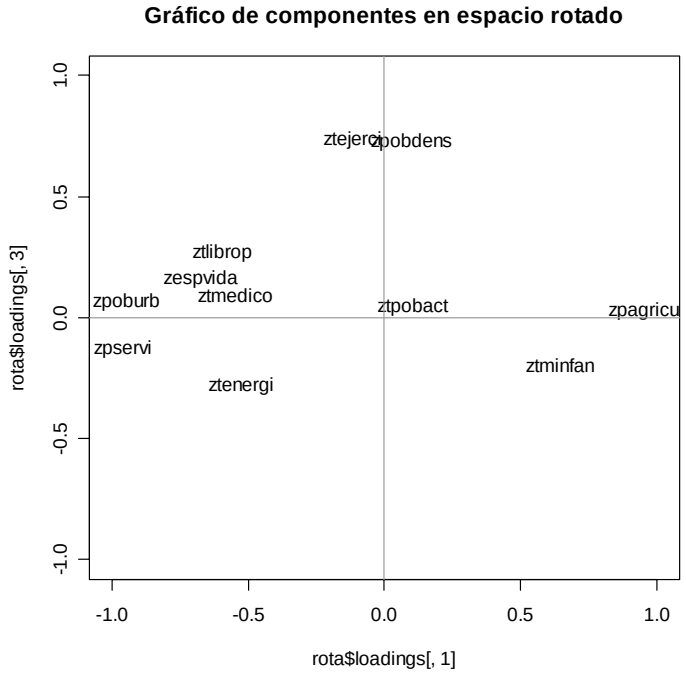
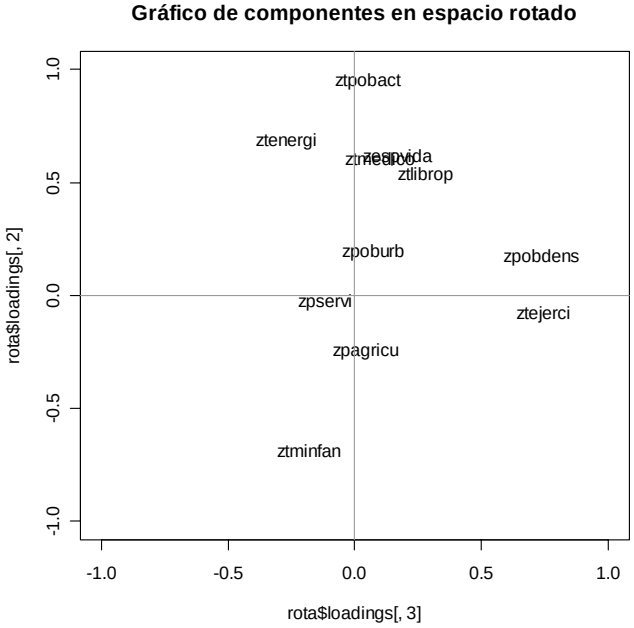


Gráfico de componentes en espacio rotado
ztejerci
zpobdens
ztlibrop
zespvida
zpoburb
ztmedico
ztpobact
zpagricu
zpservi
ztminfan
ztenergi
rota$loadings[, 3]
rota$loadings[, 1]
1.0
0.5
0.0
-0.5
-1.0



La m**atriz de transformación de las componentes** muestra la matriz de rotación (T=`$rotmat`) efectuada, mediante la cual transformamos la solución inicial (tabla Matriz de componentes) en la solución final (tabla Matriz de componentes rotados):

```
$rotmat
           [,1]       [,2]        [,3]
[1,]  0.8100587 -0.5799115 -0.08664594
[2,]  0.5409860  0.6821916  0.49188293
[3,] -0.2261394 -0.4453283  0.86633923
```

Si multiplicamos la matriz C (matriz de componentes) por la matriz T (**matriz de transformación de las componentes**) obtenemos la solución rotada o matriz C^{R} (**matriz de componentes rotados**

Obtención de las puntuaciones en las componentes:

A veces interesa conocer las puntuaciones que tienen los individuos en las componentes.

Se obtienen a través de la matriz C (de componentes rotados si se ha efectuado una rotación) mediante

$Y_{Nxk} = Z_{Nxp} R^{-1}_{pxp} C_{pxk}$ (donde Y está formada por la matriz de puntuaciones estandarizadas en las componentes)

Los coeficientes que proporcionan la combinación lineal de las variables observadas para obtener las componentes mediante $R^{-1}C$ son:

Tabla de coeficientes para el cálculo de las puntuaciones en las componentes:

Las puntuaciones sin rotar por la matriz de rotación

El resultado es la matriz de puntuaciones estandarizadas en componentes rotadas, dada en la tabla siguiente:

```
> scale(predict(acp)[,1:3])%*% rota$rotmat
```

```
>  scale(predict(acp)[,1:3])%*% rota$rotmat
              [,1]        [,2]         [,3]
africasu -0.2211897 -1.06463482 -1.154286257
argelia  -0.1525921 -1.76869940 -0.772868263
argentin -0.9402002 -0.41655095 -0.740976950
australi -1.1623287  0.59009574 -1.076543473
brasil   -0.1406334 -0.70828261 -0.862229951
canada   -1.0743802  1.26913416 -1.622798879
chile    -0.9983720 -1.07206959 -0.551038109
china     2.0915914  1.35577123 -0.098258846
coreasur  0.2998699 -0.08953580  2.591908625
egipto    0.3309670 -1.51186342 -0.225494092
españa   -0.8251721  0.08061296  0.346246029
filipina  0.8258322 -0.39314421 -0.014671518
francia  -0.8622797  0.46867818  0.011287623
humgria   0.1869883  1.31701958  0.375700010
india     1.7800302 -0.24168012  0.150418915
indonesi  1.2643138 -0.61052970 -0.686574006
iran      0.1931894 -1.16868322 -0.428876240
israel   -1.6404195 -0.90718024  2.847447672
italia   -0.5149699  0.38332097  0.354176615
japon    -0.4520497  0.72869859  0.739876905
libano   -1.0024816 -1.41054246  1.196015603
marrueco         NA          NA           NA
mejico   -0.5644022 -0.99616702 -0.920742480
nigeria   1.6713361 -0.45281389 -0.591199206
pakistan  0.9750060 -1.28236262 -0.215456823
polonia   0.4317143  1.38932765 -0.091387996
rdaleman -0.1314808  1.61358883  0.003363623
reinouni -1.0138967  0.61803914  0.521507489
rfaleman -0.8987302  0.86455207  0.915107903
rumania   0.6125882  1.18324490 -0.140760414
turquia   1.1040511 -0.12348888  0.171792246
urss     -0.1167349  1.14895463 -0.108758660
usa      -0.8841353  1.07843767 -1.431846891
vietnam   1.8289708  0.12875266  1.509919796
>
```