Evidencia 1. Examen Multivariante

Veronica Marquez

22/11/2021

“El Informe Mundial de la Felicidad es una encuesta histórica sobre el estado de la felicidad global. El informe continúa ganando reconocimiento mundial a medida que los gobiernos, las organizaciones y la sociedad civil utilizan cada vez más los indicadores de felicidad para informar sus decisiones de formulación de políticas. Los principales expertos en todos los campos (economía, psicología, análisis de encuestas, estadísticas nacionales, salud, políticas públicas y más) describen cómo las mediciones de bienestar se pueden utilizar de manera efectiva para evaluar el progreso de las naciones. Los informes revisan el estado de la felicidad en el mundo de hoy y muestran cómo la nueva ciencia de la felicidad explica las variaciones personales y nacionales en la felicidad.”

Fuente de la base de datos Helliwell, John F., Richard Layard, Jeffrey Sachs, and Jan-Emmanuel De Neve, eds. 2021. World Happiness Report 2021. New York: Sustainable Development Solutions Network.

Aplicaremos la técnica de análisis de componentes principales para estimar uno o varios índices de felicidad con las siguientes variables:

| Clave | Descripción |
| --- | --- |
| Bienestar | Puntuación promedio de bienestar subjetivo. |
| Riqueza | Logaritmo del Producto Interno Bruto per capita en dólares a precios 2011 |
| Apoyo\_social | ¿tiene parientes o amigos con los que puedes contar para pedir ayuda siempre que los necesites |
| Esperanza\_vida | Edad promedio de vida |
| Libertad\_Elegir | Está satisfecho o insatisfecho con tu libertad para elegir qué es lo que haces con tu vida? Porcentaje de satisfechos. |
| Generosidad | La generosidad es el residual de la regresión del promedio nacional de respuesta a la Pregunta “¿Ha donado dinero a una organización benéfica en el último mes |
| Corrupción | “¿Está la corrupción generalizada en todo el mundo en el gobierno o no? |
| Afecto\_Positivo | Felicidad, risa y disfrute |
| Afecto\_Negativo | Preocupación, tristeza y enojo |

# Cargar la base de datos

library(readxl)  
Datos<-read\_excel("DatosEx.xlsx", sheet="Datos")

# Selección de las variables a utilizar

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

D1 <- Datos %>%  
 select("Bienestar","Riqueza","Apoyo\_social", "Esperanza\_vida", "Libertad\_Elegir", "Generosidad", "Corrupción", "Afecto\_Positivo", "Afecto\_Negativo" )  
  
#Nos queamos con las variables que solo necesitamos

# Verificación de observaciones faltantes

A continuación verificaremos si hay observaciones faltantes en nuestra base de datos

colSums(is.na(D1))

## Bienestar Riqueza Apoyo\_social Esperanza\_vida Libertad\_Elegir   
## 0 0 0 0 0   
## Generosidad Corrupción Afecto\_Positivo Afecto\_Negativo   
## 0 0 0 0

No hay observaciones faltantes en nuestra base de datos, por lo que no es necesario su reemplazo

# Analisis base de datos

Antes de empezar con la estimación de los índices de felicidad importante asegurarse de analizar los datos en cuanto a:

1. Analizar la centralidad y dispersión de los datos
2. Identificar observaciones extremas
3. Conocer si la distribución de las variables se asemeja a la normal
4. Tipo de relación
5. Ver si hay cuestiones que pueden impactar en las correlaciones entre las variables (como relaciones curvas, observaciones extremas)

library(psych)  
describe(D1)

## vars n mean sd median trimmed mad min max range skew  
## Bienestar 1 81 5.87 1.00 5.88 5.88 0.94 3.16 7.89 4.73 -0.11  
## Riqueza 2 81 9.75 0.95 9.91 9.81 1.02 7.68 11.32 3.64 -0.48  
## Apoyo\_social 3 81 0.84 0.11 0.89 0.86 0.10 0.51 0.98 0.48 -0.88  
## Esperanza\_vida 4 81 66.98 6.11 68.30 67.57 6.23 50.50 75.20 24.70 -0.78  
## Libertad\_Elegir 5 81 0.82 0.09 0.82 0.82 0.09 0.51 0.96 0.45 -0.61  
## Generosidad 6 81 0.00 0.14 -0.02 -0.01 0.14 -0.26 0.47 0.73 0.61  
## Corrupción 7 81 0.71 0.20 0.78 0.74 0.14 0.16 0.96 0.80 -1.09  
## Afecto\_Positivo 8 81 0.72 0.08 0.73 0.73 0.09 0.38 0.88 0.49 -0.77  
## Afecto\_Negativo 9 81 0.29 0.07 0.29 0.29 0.07 0.14 0.53 0.39 0.43  
## kurtosis se  
## Bienestar -0.53 0.11  
## Riqueza -0.87 0.11  
## Apoyo\_social -0.01 0.01  
## Esperanza\_vida -0.20 0.68  
## Libertad\_Elegir 0.36 0.01  
## Generosidad 0.37 0.02  
## Corrupción 0.17 0.02  
## Afecto\_Positivo 1.35 0.01  
## Afecto\_Negativo 0.09 0.01

Notas: 1) Contamos con 81 observaciones en la base de datos, 4) De acuerdo al coeficiente de asimetria (skew) y a la curtosis (kurtosis) todas las variables tienen distribuciones que se alejan de la normalidad pues el coeficiente de simetría (skew) no es muy cercano a cero y la curtosis (kurtosis) es mas baja que la normal con valor de 3. Sin embargo no vamos a hacer inferencias con los factores estimados.

#Graficos de dispersion e histogramas

También es conveniente analizar las relaciones entre las variables con métodos gráficos para ver si hay observaciones extremas o curvas que puedan afectar las correlaciones

library(car)

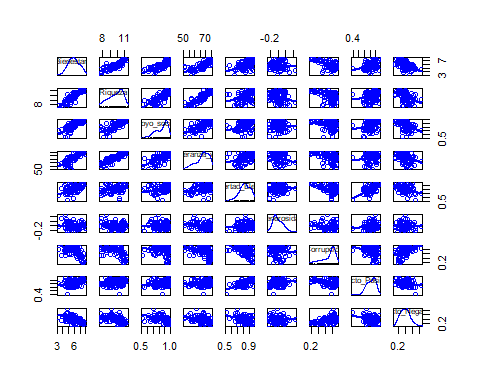
## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

## The following object is masked from 'package:dplyr':  
##   
## recode

scatterplotMatrix(D1,smooth=FALSE)



Resultados observados: 1) En la diagonal se puede validar con los histogramas que las distribuciones no son normales pues no coinciden con el diagrama de “campana. En este caso asumiremos normalidad a través del Teorema de Limite Central pues nuestra muestra es mayor a 30 observaciones. 2) No se observan relaciones curvas 3 En algunas variables se observan observaciones extremas, esperando que no afecten la signficania estadística de las correlaciones de las variables.

# Métodos para validar que podamos aplicar un análisis de factores

Para determinar si es factible aplicar un análisis de componentes principales a un conjunto de datos debemos evaluar:

1. Matriz de correlación
2. El determinante de la matriz de correlaciones
3. La prueba de contraste de esfericidad de Bartlett
4. El análisis de suficiencia general o Kaiser-Meyer-Olkin
   1. Matriz de correlación

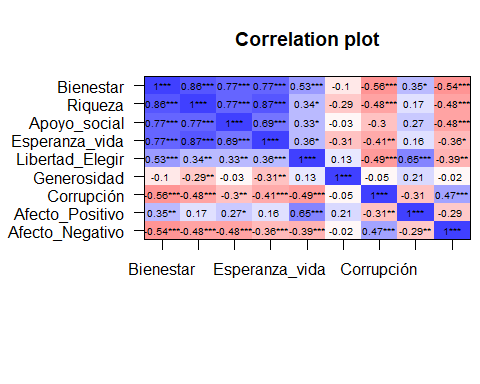
* La matriz nos da una idea de qué variables no se correlacionan con la totalidad, esas podrían no ser buenas variables para el analisis.

library(psych)  
corPlot(D1,cex = .6,stars = TRUE, show.legend=FALSE)

## Warning in abbreviate(rownames(r), minlength = minlength): abreviatura utilizada  
## con caracteres no ASCII

## Warning in abbreviate(colnames(r), minlength = minlength): abreviatura utilizada  
## con caracteres no ASCII

## Warning in abbreviate(dimnames(ans)[[2L]], minlength = abbr.colnames):  
## abreviatura utilizada con caracteres no ASCII  
  
## Warning in abbreviate(dimnames(ans)[[2L]], minlength = abbr.colnames):  
## abreviatura utilizada con caracteres no ASCII

 Prueba de hipótesis: H0: correlación = 0 H1: correlación =! 0

1. Es posible observar que no todas las variables tienen correlaciones significativas, es decir, diferentes de cero.
2. Algunas variables tienen correlación pero es muy débil
3. La correlación más alta es de 0.87 y la más baja 0.27. Es necesario validar si realmente las correlaciones existentes entre las variables permiten realizar el análisis de factores o debemos borrar variables que no se correlacionan
   1. Determinante de la matriz de correlaciones

Para poder validar el uso de la técnica se obtendrá el determinante de una matriz de correlación, la cual oscila entre 0 y.

0 ≤ |R| ≤ 1

#Estimar la matriz de correlaciones  
R<-cor(D1)  
#Calcular el determinante de la matriz  
det(R)

## [1] 0.002125726

La prueba indica que si el determinante es cercano a cero significa que las variables están altamente correlacionados y sí se puede realizar el análisis de factores. Como podemos observar el resultado del determinante de la matriz de correlaciones es 0.0021, cercano a cero. Sin embargo, la cercanía al cero es relativo.

3. Prueba de contraste de esfericidad de Bartlett

Ahora comprobaremos si la matriz de correlaciones es diferente a una matriz identidad, esto para descartar que no sea posible aplicar el análisis. La prueba de Bartlett plantea las siguientes hipótesis:

library(psych)  
cortest.bartlett(R,n=81)

## $chisq  
## [1] 468.7024  
##   
## $p.value  
## [1] 9.812831e-77  
##   
## $df  
## [1] 36

La pruba de hipótesis indica:

H0: R = I (no se debe de utilizar la técnica de análisis de factores) Ha: R diferente I (sí se puede aplicar)

Conclusión: - Estadístico de prueba, Chi-cuadrado = 468.7024 - A través del método del valor P, rechazamos H0 a favor de Ha ya que el valor de alpga =(0.05) es mayor al valor p registrado (9.81 e-77), por lo tanto sí se puede aplicar el análisis puesto que la matriz de correlación es diferente a la matriz identidad.

4. El análisis de suficiencia general o índice Kaiser-Meyer-Olkin

Una vez que ser rechazó la hipótesis nula de la prueba de Bartlett, se comprobará el nivel de suficiencia a la posible solución que se encuentre con el análisis de factores.

| Criterio | Evaluación |
| --- | --- |
| MSA ≥ 0.9 | Excelente |
| 0.8 ≤ MSA < 0.9 | Bueno |
| 0.7 ≤ MSA < 0.8 | Aceptable |
| 0.6 ≤ MSA < 0.7 | Regular |
| 0.5 ≤ MSA < 0.6 | Bajo |
| MSA < 0.5 | Inaceptable |

library(psych)  
KMO(R)

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = R)  
## Overall MSA = 0.82  
## MSA for each item =   
## Bienestar Riqueza Apoyo\_social Esperanza\_vida Libertad\_Elegir   
## 0.88 0.79 0.84 0.85 0.75   
## Generosidad Corrupción Afecto\_Positivo Afecto\_Negativo   
## 0.54 0.82 0.73 0.90

El índice glboal, MSA, es de 0.82 el cual tiene un nivel BUENO.

En el análisis individual de las variables es posible observar que ninguna tiene un MSA <0.5, por lo tanto no es necesario eliminar varibles en nuestro modelo.

#Estandarización de las variables

Con el fin de que todas nuestra variables se encuentren en la misma esacala, se procederá a estandarizarlas:

Dz <- data.frame(scale(D1)) #Estandarizar las variables para que estén en la misma escala

# Estimar el modelo de análisis de componentes principales

Una vez estandarizadas, pocedemos a la estimación de nuestro modelo obteniendo los componentes, los cuales son combinaciones lineales de las 9 variables. Lo ideal es sacar solo un índice, veremos con las pruebas cuántos componentes son necesarios.

PRUEBA 1. Estimación del modelo con todas las variables

library(psych)  
  
modelo1<-principal(Dz,nfactors = 9, rotate= "none")  
modelo1

## Principal Components Analysis  
## Call: principal(r = Dz, nfactors = 9, rotate = "none")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 h2 u2  
## Bienestar 0.93 -0.08 0.06 0.07 0.07 0.00 -0.10 0.29 -0.13 1 0.0e+00  
## Riqueza 0.89 -0.36 0.05 0.05 0.08 -0.05 0.07 0.09 0.23 1 1.1e-16  
## Apoyo\_social 0.81 -0.17 0.30 0.34 -0.10 -0.06 -0.25 -0.20 -0.01 1 5.6e-16  
## Esperanza\_vida 0.83 -0.38 -0.06 0.12 0.17 0.08 0.30 -0.13 -0.10 1 1.2e-15  
## Libertad\_Elegir 0.64 0.54 -0.35 0.05 0.06 0.40 -0.09 -0.04 0.04 1 8.9e-16  
## Generosidad -0.12 0.72 0.61 0.16 0.24 0.05 0.11 0.02 0.02 1 1.8e-15  
## Corrupción -0.66 -0.25 0.03 0.58 -0.35 0.18 0.08 0.09 0.01 1 1.3e-15  
## Afecto\_Positivo 0.46 0.67 -0.35 0.28 -0.17 -0.32 0.09 0.00 0.00 1 8.9e-16  
## Afecto\_Negativo -0.66 -0.16 -0.26 0.38 0.55 -0.07 -0.09 0.00 0.01 1 7.8e-16  
## com  
## Bienestar 1.3  
## Riqueza 1.6  
## Apoyo\_social 2.2  
## Esperanza\_vida 2.0  
## Libertad\_Elegir 3.4  
## Generosidad 2.5  
## Corrupción 3.2  
## Afecto\_Positivo 3.6  
## Afecto\_Negativo 3.2  
##   
## PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9  
## SS loadings 4.5 1.65 0.79 0.73 0.57 0.32 0.21 0.16 0.08  
## Proportion Var 0.5 0.18 0.09 0.08 0.06 0.04 0.02 0.02 0.01  
## Cumulative Var 0.5 0.68 0.77 0.85 0.91 0.95 0.97 0.99 1.00  
## Proportion Explained 0.5 0.18 0.09 0.08 0.06 0.04 0.02 0.02 0.01  
## Cumulative Proportion 0.5 0.68 0.77 0.85 0.91 0.95 0.97 0.99 1.00  
##   
## Mean item complexity = 2.5  
## Test of the hypothesis that 9 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

# Interpretación de los resultados

Se tienen nueve variables, entonces hay nueve soluciones, esto es, seis autovalores/ eigenvalores (SS loadings).

PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9

SS loadings 4.5 1.65 0.79 0.73 0.57 0.32 0.21 0.16 0.08 Proportion Var 0.5 0.18 0.09 0.08 0.06 0.04 0.02 0.02 0.01 Cumulative Var 0.5 0.68 0.77 0.85 0.91 0.95 0.97 0.99 1.00 Proportion Explained 0.5 0.18 0.09 0.08 0.06 0.04 0.02 0.02 0.01 Cumulative Proportion 0.5 0.68 0.77 0.85 0.91 0.95 0.97 0.99 1.00

* Serían necesarios solo los primeros dos componentes explican el 0.68 % de la varianza
* De acuerdo con el mean item complexity, necesitamos 3 componentes solamente

Una vez probando esto, es necesario ahondar en criterios para determinar el número de factores necesarios para nuestro modelo final.

#Criterios para determinar el número de factores

Los siguientes criterios permiten determinar el número de factores como posible solución inicial:

1. Criterio a priori
2. Criterio del índice de complejidad
3. Criterio de la raíz latente
4. Criterio del porcentaje de la variación explicada acumulada
5. Criterio del gráfico de sedimentación
   1. Indice de complejidad (Mean item complexity) Fijarse en el mean item complexity, el cual sugiere cuántos componenetes son necesarios para explicar las variables observadas.

Mean item complexity = 2.5 El modelo con las 9 variables sugiere 3 componentes.

2. Criterio de la raiz latente (eigen valores o SS loadings)

Con este criterio se incluye en la solución solamente a los componentes con un autovalor/eigenvalor mayor o igual a 1.

PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9

SS loadings 4.5 1.65 0.79 0.73 0.57 0.32 0.21 0.16 0.08

Por lo que solo PC1 y PC2 tiene un eigen valor mayor a 1, y solo son necesarios 2 compontentes.

3. Criterio del porcentaje de la variación explicada acumulada  
  
 PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9

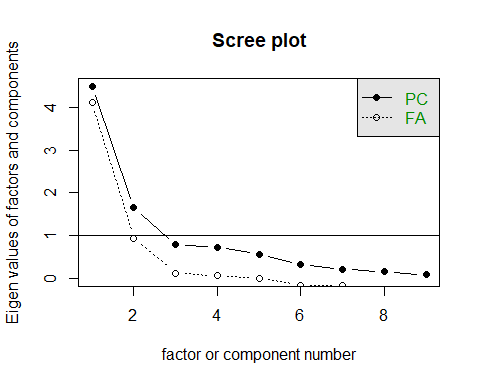
Cumulative Var 0.5 0.68 0.77 0.85 0.91 0.95 0.97 0.99 1.00

Este criterio sugiere que se requieren dos componentes, PC1 y PC2, pues entre los dos explican el 68% de la varianza (mayor al menos 60%)

4. Criterio del gráfico de sedimentación (gráfico del codo).

La idea de este gráfico está enfocada en que si un componente es importante tendrá una varianza grande. En este gráfico en el eje de las Y se grafican los autovalores/eigenvalores y en el eje de las X los componentes. El gráfico tiene forma de precipicio, se retendrá en al solución inicial solo los componentes que estén antes de la zona de sedimentación (antes del cambio drástico en pendiente).

library(psych)  
R<-cor(D1) #Matriz de correlación  
scree(R) #Gráfico de sedimentación

 Podríamos decir que este método nos sugiere 2 compoenetes.

Para concluir cuántos componentes necesitamos, en 3 de 4 criterios concluimos que se necesitan 2, por lo que el modelo final contará con 2 componentes

#Estimación del modelo final con el # de componentes adecuado

library(psych)  
modelo2<-principal(Dz,nfactors = 2, rotate='none')  
modelo2

## Principal Components Analysis  
## Call: principal(r = Dz, nfactors = 2, rotate = "none")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## PC1 PC2 h2 u2 com  
## Bienestar 0.93 -0.08 0.87 0.13 1.0  
## Riqueza 0.89 -0.36 0.92 0.08 1.3  
## Apoyo\_social 0.81 -0.17 0.68 0.32 1.1  
## Esperanza\_vida 0.83 -0.38 0.83 0.17 1.4  
## Libertad\_Elegir 0.64 0.54 0.70 0.30 2.0  
## Generosidad -0.12 0.72 0.53 0.47 1.1  
## Corrupción -0.66 -0.25 0.49 0.51 1.3  
## Afecto\_Positivo 0.46 0.67 0.66 0.34 1.8  
## Afecto\_Negativo -0.66 -0.16 0.46 0.54 1.1  
##   
## PC1 PC2  
## SS loadings 4.50 1.65  
## Proportion Var 0.50 0.18  
## Cumulative Var 0.50 0.68  
## Proportion Explained 0.73 0.27  
## Cumulative Proportion 0.73 1.00  
##   
## Mean item complexity = 1.3  
## Test of the hypothesis that 2 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.09   
## with the empirical chi square 42.38 with prob < 0.0016   
##   
## Fit based upon off diagonal values = 0.97

Resultados:

Los dos componentes explican un 68% de la varianza

PC1 PC2

SS loadings 4.50 1.65 Proportion Var 0.50 0.18 Cumulative Var 0.50 0.68 Proportion Explained 0.73 0.27 Cumulative Proportion 0.73 1.00

De acuerdo con la comunalidad (h2) la variable que mejor explican los dos componentes es la Riqueza, con un 91%, seguida del bienestar con un 87% y la que menor la explica es el afecto negativo con un 46%,

# Interpretación de componentes

## Matriz de factores rotada

La matriz de factores rotada (al igual que la no rotada) proporciona información de la relación de las variables con los factores (las cargas de los factores), sin embargo ahora serán las definitivas las que permitirán agrupar los factores e interpretarlos, darán la solución final.

Utilizaremos el método de “Varimax” Es el método más utilizado, minimiza el número de variables que tienen saturaciones altas en cada factor

library(psych)  
modelo2<-principal(Dz,nfactors = 2, rotate="varimax")  
modelo2

## Principal Components Analysis  
## Call: principal(r = Dz, nfactors = 2, rotate = "varimax")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## RC1 RC2 h2 u2 com  
## Bienestar 0.88 0.31 0.87 0.13 1.2  
## Riqueza 0.96 0.04 0.92 0.08 1.0  
## Apoyo\_social 0.81 0.18 0.68 0.32 1.1  
## Esperanza\_vida 0.91 0.00 0.83 0.17 1.0  
## Libertad\_Elegir 0.36 0.76 0.70 0.30 1.4  
## Generosidad -0.40 0.60 0.53 0.47 1.7  
## Corrupción -0.49 -0.50 0.49 0.51 2.0  
## Afecto\_Positivo 0.14 0.80 0.66 0.34 1.1  
## Afecto\_Negativo -0.53 -0.42 0.46 0.54 1.9  
##   
## RC1 RC2  
## SS loadings 4.01 2.14  
## Proportion Var 0.45 0.24  
## Cumulative Var 0.45 0.68  
## Proportion Explained 0.65 0.35  
## Cumulative Proportion 0.65 1.00  
##   
## Mean item complexity = 1.4  
## Test of the hypothesis that 2 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.09   
## with the empirical chi square 42.38 with prob < 0.0016   
##   
## Fit based upon off diagonal values = 0.97

Al rotar los factores las cargas cambian dada la nueva posición de los ejes, esto implica que también cambia el autovalor/eigenvalor, pero no cambia la variación explicada total ni las comunalidades

* La varianza acumulada sigue siendo 60%
* Lo que cambia son las correlaciones, lo que permitirá poner nombre a cada componente

# Criterios para identificar cargas significativas

Los criterios para determinar cargas significativas permiten identificar la variable con el factor.

Prioridad 1: |Carga| ≥ 0.5 Prioridad 2: 0.4 ≤|Carga|< 0.5 Prioridad 3: 0.3 ≤|Carga|< 0.4

Criterio 1. Para determinar qué variables definen al componente:

print(modelo2$loadings,cut=0.5, sort=TRUE)

##   
## Loadings:  
## RC1 RC2   
## Bienestar 0.882   
## Riqueza 0.958   
## Apoyo\_social 0.806   
## Esperanza\_vida 0.910   
## Afecto\_Negativo -0.534   
## Libertad\_Elegir 0.759  
## Generosidad 0.605  
## Afecto\_Positivo 0.797  
## Corrupción   
##   
## RC1 RC2  
## SS loadings 4.010 2.137  
## Proportion Var 0.446 0.237  
## Cumulative Var 0.446 0.683

#Del modelo 2 saca las cargas y desaparece las que tenga menos de o.5 (Propiedad 1 =0.5)

Debido a que corrupción se quedó fuera de un componente, es necesario bajar al criterio 2. .

#Criterio 2. Para todas las cargas que tengan entre 0.4 y 0.5 y que no expliquen ningun componente con el criterio 1.

print(modelo2$loadings,cut=0.4, sort=TRUE) #(Prioridad 2 = 0.4)

##   
## Loadings:  
## RC1 RC2   
## Bienestar 0.882   
## Riqueza 0.958   
## Apoyo\_social 0.806   
## Esperanza\_vida 0.910   
## Afecto\_Negativo -0.534 -0.423  
## Libertad\_Elegir 0.759  
## Generosidad -0.401 0.605  
## Afecto\_Positivo 0.797  
## Corrupción -0.495 -0.500  
##   
## RC1 RC2  
## SS loadings 4.010 2.137  
## Proportion Var 0.446 0.237  
## Cumulative Var 0.446 0.683

Interpretación de las cargas de la matriz rotada

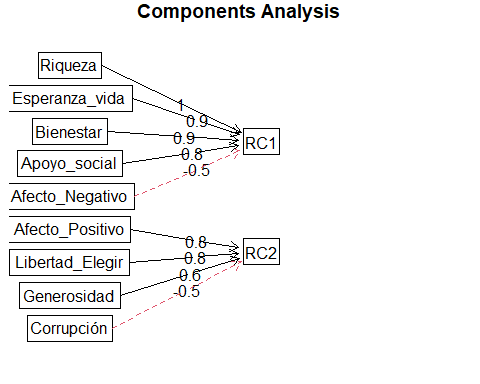
Componente 1 Representa a las variables: Bienestar, Riqueza, Apoyo\_social, Esperanza de vida y Afecto\_Negativo. Esta última se queda en este componente debido a que tiene la carga más alta a pesar de que explica ambos.

Podríamos llamarlo “índice objetivo de la felicidad que toma en cuenta el bienestar subjetivo”

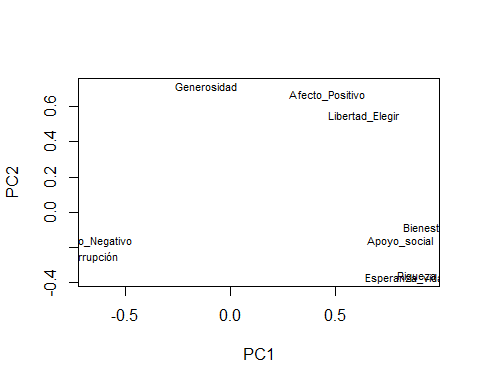
Componente 2 Representa a las variables: Libertad\_Elegir, Generosidad, Afecto positivo y corrupción Podríamos llamarlo “índice subjetivo de la felicidad”

# Alternativas gráficas para interpretar componentes

library(psych)  
fa.diagram(modelo2)

 Aquí podemos notar que la corrupción y el afecto negativo tienen correlación negativa con sus respectivos componentes

load <- modelo1$loadings[,1:2] #obtiene las cargas de las variables  
plot(load,type="n") # Dibujar el área de visualización  
text(load,labels=names(Dz),cex=.7) # Añadir las variables



# Obtención de puntajes

Obtendremos los puntajes con el método de regresión.

library(psych)  
modelo2<-principal(Dz,nfactors = 2, rotate="varimax",  
 scores=TRUE,method="regression") #con el método que queramos calcular los puntajes  
  
Dz$PC1<-modelo2$scores[,1]  
Dz$PC2<-modelo2$scores[,2] #salen en la base de datos en otra columna

Al dar clic en el objeto Dz apareceran las nuevas variables PC1 y PC2

# Matriz de coeficientes estandarizados para el cálculo de los factores

La ecuación se obtiene con el siguiente código:

modelo2$weights #

## RC1 RC2  
## Bienestar 0.209268399 0.03976444  
## Riqueza 0.270206529 -0.11716781  
## Apoyo\_social 0.205598960 -0.01757627  
## Esperanza\_vida 0.262069549 -0.13170639  
## Libertad\_Elegir -0.007262906 0.35907219  
## Generosidad -0.202558026 0.38442367  
## Corrupción -0.070351931 -0.19863534  
## Afecto\_Positivo -0.073710621 0.41012393  
## Afecto\_Negativo -0.092770992 -0.15135457

RC1 RC2

Bienestar 0.209268399 0.03976444 Riqueza 0.270206529 -0.11716781 Apoyo\_social 0.205598960 -0.01757627 Esperanza\_vida 0.262069549 -0.13170639 Libertad\_Elegir -0.007262906 0.35907219 Generosidad -0.202558026 0.38442367 Corrupción -0.070351931 -0.19863534 Afecto\_Positivo -0.073710621 0.41012393 Afecto\_Negativo -0.092770992 -0.15135457

La ecuación de los componentes para obtener los índices está dada por:

Componente 1 RC1i = 0.209*Bienestar + 0.270*Riqueza + 0.205*Apoyo social - 0.262*Esperanza\_vida -0.007*Libertad\_Elegir - 0.202* Generosidad - 0.070*Corrupción -0.073* Afecto\_Positivo - 0.092\* Afecto\_Negativo

Componente 2 RC2i = 0.039*Bienestar - 0.1171*Riqueza - 0.017*Apoyo social - 0.131*Esperanza\_vida +0.3597*Libertad\_Elegir + 0.3844* Generosidad - 0.19860*Corrupción +0.41* Afecto\_Positivo - 0.151\* Afecto\_Negativo

# Alfa de Cronbach

Una alternativa para validar si las variables permiten hacer un índice es utilizar una medida de consistencia interna para verificar que estas se muevan en el mismo sentido.

El alfa de Cronbach toma un rango de 0 a 1, un valor confiable es de al menos 0.6.

Para el componente 1 formado por las variables Bienestar, Riqueza, Apoyo\_social, Esperanza de vida y Afecto\_Negativo:

library(psych)  
alpha(Dz[,c(1,2,3,4,9)],check.keys=TRUE) #variables a utilizar

## Number of categories should be increased in order to count frequencies.

## Warning in alpha(Dz[, c(1, 2, 3, 4, 9)], check.keys = TRUE): Some items were negatively correlated with total scale and were automatically reversed.  
## This is indicated by a negative sign for the variable name.

##   
## Reliability analysis   
## Call: alpha(x = Dz[, c(1, 2, 3, 4, 9)], check.keys = TRUE)  
##   
## raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r  
## 0.91 0.91 0.91 0.66 9.7 0.017 0.032 0.85 0.73  
##   
## lower alpha upper 95% confidence boundaries  
## 0.87 0.91 0.94   
##   
## Reliability if an item is dropped:  
## raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r  
## Bienestar 0.86 0.86 0.86 0.61 6.2 0.026 0.0387  
## Riqueza 0.86 0.86 0.85 0.60 6.1 0.026 0.0276  
## Apoyo\_social 0.88 0.88 0.89 0.65 7.3 0.023 0.0459  
## Esperanza\_vida 0.88 0.88 0.87 0.65 7.4 0.022 0.0286  
## Afecto\_Negativo- 0.94 0.94 0.93 0.79 14.8 0.012 0.0046  
## med.r  
## Bienestar 0.59  
## Riqueza 0.61  
## Apoyo\_social 0.65  
## Esperanza\_vida 0.65  
## Afecto\_Negativo- 0.77  
##   
## Item statistics   
## n raw.r std.r r.cor r.drop mean sd  
## Bienestar 81 0.92 0.92 0.92 0.87 3.8e-17 1  
## Riqueza 81 0.93 0.93 0.94 0.89 -4.2e-16 1  
## Apoyo\_social 81 0.87 0.87 0.82 0.79 -2.2e-17 1  
## Esperanza\_vida 81 0.87 0.87 0.85 0.78 -8.3e-16 1  
## Afecto\_Negativo- 81 0.67 0.67 0.54 0.51 1.6e-01 1

El valor de alfa es de 0.91, esto significa que si se puede hacer un índice a partir de las varialbesBienestar, Riqueza, Apoyo\_social, Esperanza de vida y Afecto\_Negativo

También nos da el análisis de si podemos mejorar quitando variables con el raw-alpha. Si se quitara la variable Afecto\_Negativo el alfa aumentaría a 0.93, pero como no hay cambios sustanciales la evidencia indica quedaranos con todas las variables.

Para el componente 2 formado por Libertad\_Elegir, Generosidad, Afecto positivo y corrupción

library(psych)  
alpha(Dz[,c(5,6,7,8)],check.keys=TRUE)

## Number of categories should be increased in order to count frequencies.

## Warning in alpha(Dz[, c(5, 6, 7, 8)], check.keys = TRUE): Some items were negatively correlated with total scale and were automatically reversed.  
## This is indicated by a negative sign for the variable name.

##   
## Reliability analysis   
## Call: alpha(x = Dz[, c(5, 6, 7, 8)], check.keys = TRUE)  
##   
## raw\_alpha std.alpha G6(smc) average\_r S/N ase mean sd median\_r  
## 0.64 0.64 0.64 0.31 1.8 0.065 -0.14 0.69 0.26  
##   
## lower alpha upper 95% confidence boundaries  
## 0.51 0.64 0.77   
##   
## Reliability if an item is dropped:  
## raw\_alpha std.alpha G6(smc) average\_r S/N alpha se var.r med.r  
## Libertad\_Elegir 0.41 0.41 0.34 0.19 0.71 0.113 0.016 0.21  
## Generosidad 0.74 0.74 0.69 0.48 2.79 0.052 0.029 0.49  
## Corrupción- 0.60 0.60 0.58 0.33 1.48 0.079 0.076 0.21  
## Afecto\_Positivo 0.47 0.47 0.43 0.23 0.88 0.103 0.054 0.13  
##   
## Item statistics   
## n raw.r std.r r.cor r.drop mean sd  
## Libertad\_Elegir 81 0.82 0.82 0.80 0.63 3.1e-16 1  
## Generosidad 81 0.51 0.51 0.20 0.17 1.3e-17 1  
## Corrupción- 81 0.67 0.67 0.49 0.38 -5.8e-01 1  
## Afecto\_Positivo 81 0.78 0.78 0.72 0.56 3.6e-16 1

Nos indica un alfa de 0.64 entonces es confiable hacer un índice con Libertad\_Elegir, Generosidad, Afecto positivo y corrupció

El análisis por variable indica que se mejora sustancialmente el alfa si se quitara la variable generosidad del índice, pues aumentaría el alfa a 0.73.

De aquí podemos observar que es necesario reescalar las variables corrupción y afecto negativo, ya que tienen una correlación negativa. Además se mejoraría sustencialmente el índice 2 si se quitara la varaible generosidad

# Rescalamiento de variables con correlación negativa

Dz$Corrupciónr<-with(Dz,max(Corrupción)-Corrupción)  
Dz$Afecto\_Negativor<-with(Dz,max(Afecto\_Negativo)-Afecto\_Negativo)

Una vez que se cambio la escala se tiene que volver a estimar el modelo y los componentes.

library(psych)  
modelo3<-principal(Dz[,c(1,2,3,4,5,7,8,9)],nfactors = 2, rotate="varimax", scores=TRUE,method="regression")  
#sin la variable generosidad pues se decidió quitarla (columna 6)  
  
Dz$PC1<-modelo3$scores[,1]  
Dz$PC2<-modelo3$scores[,2]  
  
modelo3

## Principal Components Analysis  
## Call: principal(r = Dz[, c(1, 2, 3, 4, 5, 7, 8, 9)], nfactors = 2,   
## rotate = "varimax", scores = TRUE, method = "regression")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## RC1 RC2 h2 u2 com  
## Bienestar 0.86 0.38 0.88 0.119 1.4  
## Riqueza 0.95 0.15 0.92 0.081 1.0  
## Apoyo\_social 0.84 0.18 0.74 0.261 1.1  
## Esperanza\_vida 0.89 0.12 0.81 0.187 1.0  
## Libertad\_Elegir 0.23 0.86 0.79 0.215 1.1  
## Corrupción -0.43 -0.56 0.50 0.499 1.9  
## Afecto\_Positivo 0.01 0.87 0.75 0.247 1.0  
## Afecto\_Negativo -0.50 -0.46 0.46 0.542 2.0  
##   
## RC1 RC2  
## SS loadings 3.62 2.23  
## Proportion Var 0.45 0.28  
## Cumulative Var 0.45 0.73  
## Proportion Explained 0.62 0.38  
## Cumulative Proportion 0.62 1.00  
##   
## Mean item complexity = 1.3  
## Test of the hypothesis that 2 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.08   
## with the empirical chi square 26.24 with prob < 0.016   
##   
## Fit based upon off diagonal values = 0.98

En la base de datos aparecerán los componentes estimados ya corregidos.

Los componentes estimados están en la escala de la normal estándar de -2 a 2. Pueden expresarse en escala 0 a 100 con el siguiente código.

#Poner el índice de 0 a 100

Dz$Indice1<-with(Dz,100\*(PC1-min(PC1))/(max(PC1)-min(PC1))) #Componente 1  
Dz$Indice2<-with(Dz,100\*(PC2-min(PC2))/(max(PC2)-min(PC2))) #Componente 2

#Punto final

En este ejercicio se tienen dos índices, pero queremos tener solo 1, por lo que obtendremos un promedio ponderado:  
Indice = w1*Indice1+w1*Indice2

Donde:

w1 = VE\_PC1/(VE\_PC1+VE\_PC2) w1 = VE\_PC2/(VE\_PC1+VE\_PC2)

VE = Varianza explicada

Resultados del modelo 3: RC1 RC2 SS loadings 3.62 2.23 Proportion Var 0.45 0.28 Cumulative Var 0.45 0.73 Proportion Explained 0.62 0.38 Cumulative Proportion 0.62 1.00

#Ponderadores obtenidos del modelo estimado, renglon "Proportion Var".  
w1=0.45/(0.45+0.28)  
w2=0.28/(0.28+0.45)  
#Indice global de Marginación de la base de datos  
  
Dz$Indice<-with(Dz,w1\*Indice1+w2\*Indice2) #ponderando los dos índices, nuestro final se llama columna Índice

La columna “índice” muestra el índice global de los dos componentes estimados de felicidad.

—- HASTA AQUI EL EXAMEN ——–