**Borrador: interpretación del Análisis Factorial para Datos Mixtos (FAMD)**

1. **EJEMPLO DE PÁGES (2002): BASE DE DATOS PEQUEÑA**
2. En el presente borrador, sigo el ejemplo de Páges (2002, p. 72) derivado de la siguiente base de datos pequeña con información biométrica de seis individuos. Se consideran dos variables continuas no-correlacionadas y una variable categórica con tres categorías (color de cabello negro, marrón o rubio).

Tabla

Descripción generada automáticamente

1. Se presenta la matriz de relaciones, la cual muestra o bien el coeficiente de correlación (de Pearson) o el eta-2 (razón de correlación entre variables categóricas y variables continuas).

Tabla

Descripción generada automáticamente

1. Se implementa el análisis factorial de datos mixtos (FAMD) sobre las variables continúas estandarizadas, usando la función FAMD de la librería FactoMineR. **En primer lugar,** se presenta la proporción de inercia explicada por cada uno de los factores. Se observa que los primeros dos factores (el primer plano) expresa casi toda la inercia (95.56%).

Texto

Descripción generada automáticamente

1. **En segundo lugar**, la contribución absoluta de una variable a la inercia de un eje es interpretada como una medida de la relación entre la variable y el eje (squared correlation, ratio or coefficient according to the nature of the variable):

Texto

Descripción generada automáticamente

1. **En tercer lugar**, estas inercias pueden ser representadas mediante el siguiente gráfico. Este gráfico muestra tres (3) resultados principales: (1) el **primer factor** resulta de las variables **altura** y **color del cabello**; (2) el **segundo factor** resulta de las variables **peso** y **color del cabello**. En cada caso, el factor está relacionado cercanamente (closely linked) a alguna de las dos variables.

Gráfico

Descripción generada automáticamente

1. **En cuarto lugar**, considerando el marco de interpretación de los puntos anteriores, se puede interpretar específicamente la relación con cada una de las categorías de la variable categórica. En el siguiente gráfico se puede ver que el primer factor opone principalmente **small blondes** y **tall people with black hair**.

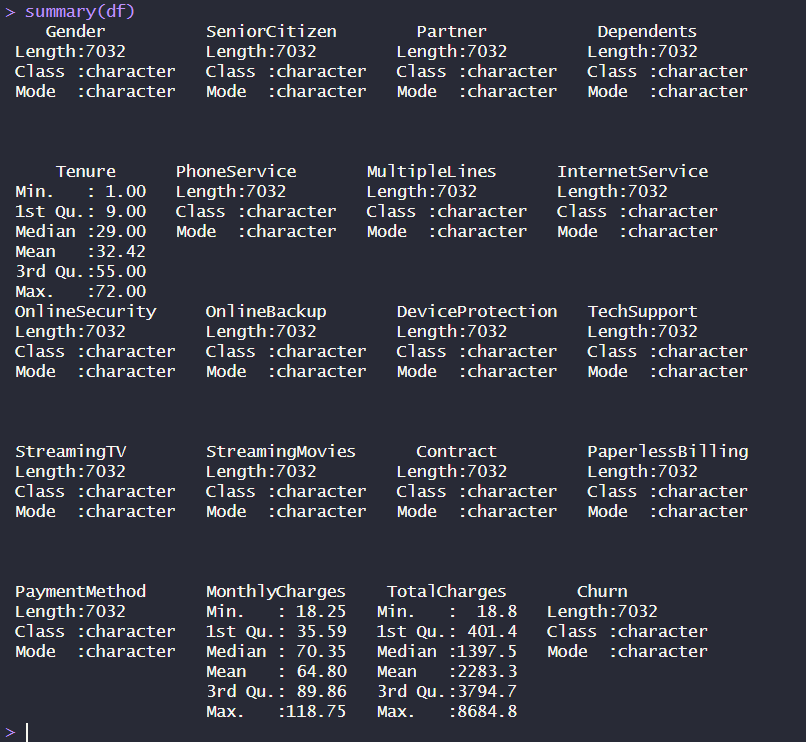
Gráfico, Diagrama

Descripción generada automáticamente

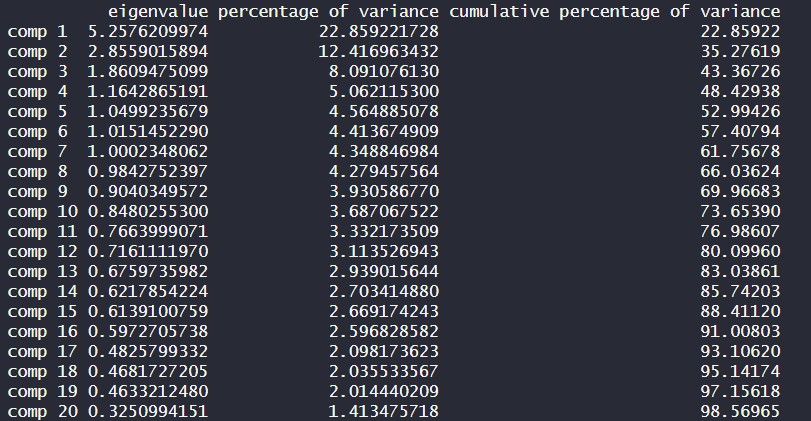
Gráfico, Diagrama

Descripción generada automáticamente

1. **EJEMPLO 2: BASE DE DATOS IBM CON ROTACIÓN DE FACTORES**
2. **En primer lugar,** se presenta un resumen sobre la base de datos de IBM:



1. **En segundo lugar,** se examinan las dimensiones principales (PC), las cuales son combinaciones lineales de las variables originales. Se inspeccionan los valores propios y el porcentaje de inercia (%) explicada:



1. **En tercer lugar,** se grafican las observaciones individuales en el nuevo *feature space*. Las distribuciones se superponen, lo cual indica que la base de datos no suficiente para capturar la diferencia entre los dos tipos de clientes (**churn** & **no-churn**).

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

1. **En cuarto lugar**, se examina la relación entre las variables cuantitativas. La **carga factorial** de una variable describe la correlación entre la variable y la PD. La carga factorial al cuadrado corresponde a cos2 que es el porcentaje de varianza en esa variable explicada por ese factor. Nótese que esto, en términos de la función FAMD de FactoMineR, corresponde a las coordenadas para las variables cuantitativas.

Diagrama

Descripción generada automáticamente

1. **En quinto lugar,** se examina el **cuadrado de las cargas factoriales**, las cuales permiten visualizar las variables cualitativas y cuantitativas en el nuevo *feature space*. De acuerdo con los autores del paquete FAM, “the coordinates are to be interpreted as measuring the links (signless) between variables and principal dimensions” (Páges, 2002). **This may be interpreted as the coordinates of each variable being the absolute value its squared loading (the strength of the relationship between variables and components or factors, disregarding whether the relationship is positive or negative)**. Así, se verifica que:

Imagen que contiene Aplicación

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Texto

Descripción generada automáticamente

1. **En sexto lugar**, mientras que la **carga factorial** y su **cuadrado** miden qué tan bien un PD describe la variación capturada en una variable; la **contribución** describe lo converso: qué tanto una variable explica la variación total capturada en el PD.

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

1. **En séptimo lugar**, se examina la rotación de factores. Para facilitar la interpretación de la relación entre las variables y los PDs, una rotación puede ser aplicada a los PDs, lo cual resulta en cargas factoriales más altas para unas pocas variables (y cargas bajas para las demás). Es decir, un número pequeño de variables estarán altamente correlacionadas con cada PD. En este caso, se usa la rotación VARIMAX, una forma generalizada implementada en la librería PCAmixdata.

Gráfico

Descripción generada automáticamente

**FINAL REMARK:** Nótese que aquí, por definición, Páges (2002, p. 74) está haciendo referencia al **cuadrado de las cargas factoriales.** Estos son los valores que se deben interpretar:

Texto

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente