

3. Componentes principales y análisis factorial exploratorio

En esta sección se analiza la personalidad de un grupo de 228 estudiantes de licenciatura de una universidad de los Estados Unidos a partir de una encuesta resumida en la base de datos `Dat3Ex.csv`. Las respuestas van del 1 al 5 de acuerdo con un cuestionario, donde 1 corresponde a “muy en desacuerdo”, 2 a “un poco en desacuerdo”, 3 a “ni de acuerdo ni en desacuerdo”, 4 a “un poco de acuerdo” y 5 a “muy de acuerdo”, para un grupo de 44 preguntas, de las cuales solamente tomaremos las siguientes 15: $V1, V2, V4, V6, V9, V12, V14, V16, V17, V26, V27, V29, V31, V34, V37$ (para mayor detalle ver el cuestionario).

Los objetivos son los de obtener los componentes principales y hacer un análisis exploratorio factorial, para identificar dimensiones interesantes de los datos en su escala original y transformada.

Con la ayuda de la librería `factoextra` se obtuvieron los *Componentes Principales* con la función `prcomp` (ver `Chunk factorCP` en la línea 64).

Posteriormente se usa la función `fviz_eig` como un índice para el número a considerar de estas componentes y en `@ref(fig:Grafica13)` se muestran los resultados, para los datos escalados y no escalados, se sugieren entre 4 o 5 componentes pues después de estos ya no hay mucho cambio en la varianza que aportan. Además se acumula en los tres casos un aproximado de 62 % a 63 % de la varianza total cuando consideramos 4 componentes, mientras que estos son entre 66 % y 67 % para 5 componentes (`Chunk Grafica13`, línea 79).

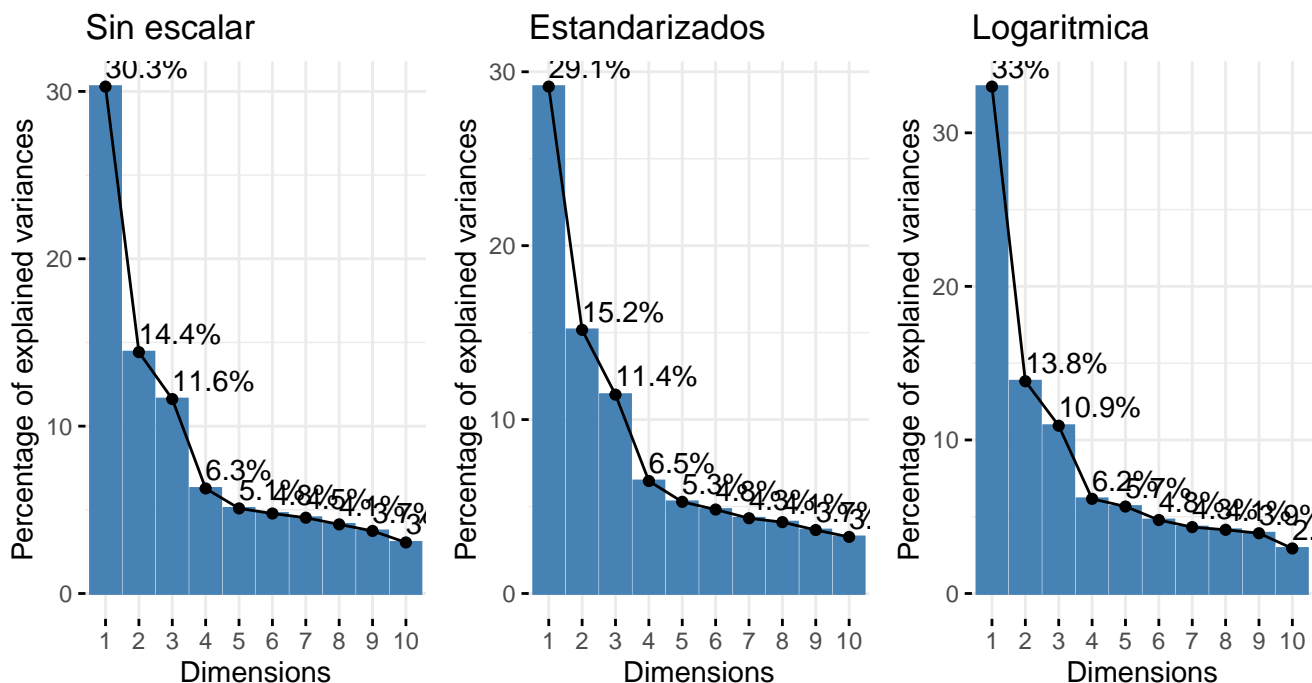


Figura 1: Índices para número de componentes principales

Analizamos las correlaciones de las primeras cuatro componentes con las variables originales (`Chunk Correlation`, línea 87). Se describen los siguientes resultados generales, considerando correlaciones mayores a 0.5 en valor absoluto, para dar una mayor comprensión y contexto de las variables y componentes principales.

1.— Para los datos sin escalar, las variables *Deprimido*, *Tenso*, *Malhumorado* y *Grosero* son las que tienen mayor asociación positiva en el componente 1, y por otro lado *Relajado*, *Calmado* y *Entusiasta* son las que tienen mayor asociación negativa con el componente 1. Las variables *Parlanchin*, *Asertivo* y *Entusiasta* son las de mayor asociación positiva para el componente 2, y *Tímido* y *Reservado* son las de mayor asociación negativa para el componente 2. Para el componente 3 las de mayor relación positiva son *Relajado*, *Frío* y *Calmado*, mientras que para la relación negativa con el componente 3 no hay valores mayores a 0.5 en valor absoluto. Y para el componente 4 no hay valores mayores a 0.5 en valor absoluto (sin embargo, mencionaremos que las de mayor relación positiva son *Tímido*, *Indulgente* y *Entusiasta*, mientras que las únicas con relación negativa son *Frío*, *Peleonero* y *Victimista*).

2.— Para los datos estandarizados, las variables Deprimido, Tenso, Malhumorado, Grosero, Victimista, Peleonero y Frío son las que tienen mayor asociación positiva en el componente 1, y por otro lado Relajado, Calmado y Entusiasta son las que tienen mayor asociación negativa con el componente 1. Las variables Parlanchin y Asertivo son las de mayor asociación positiva para el componente 2, y Tímido y Reservado son las de mayor asociación negativa para el componente 2. Para el componente 3 las de mayor relación positiva son Relajado, Frío y Calmado, mientras que para la relación negativa con el componente 3 no hay valores mayores a 0.5 en valor absoluto. Y para el componente 4 la única variable con asociación negativa importante es indulgente, y todas las demás asociaciones son negativas, menores 0.5 en valor absoluto.

3.— Para los datos en escala logarítmica, las variables Deprimido, Tenso, Malhumorado y Grosero son las que tienen mayor asociación positiva en el componente 1, y por otro lado Relajado, Calmado y Entusiasta son las que tienen mayor asociación negativa con el componente 1.

Las variables Parlanchin, Asertivo y Entusiasta son las de mayor asociación positiva para el componente 2, y Tímido y Reservado son las de mayor asociación negativa para el componente 2. Para el componente 3 las de mayor relación positiva son Relajado, Frío y Calmado, mientras que para la relación negativa con el componente 3 no hay valores mayores a 0.5 en valor absoluto.

Y para el componente 4 no hay valores mayores a 0.5 en valor absoluto, sin embargo mencionamos que las de mayor relación positiva son Tímido, Indulgente y Entusiasta, mientras que las únicas con relación negativa son Frío, Peleonero y Victimista.

Para mayor interpretabilidad tenemos la Gráfica @ref(fig:Grafica23), en esta sólo se presentan los datos originales y los de escala logarítmica, son las proyecciones de las variables de mayor peso en los primeros 2 componentes principales pues estos rescatan la mayor varianza, podemos guiarnos por el sentido y magnitud de las flechas para intuir la influencia de cada variables en cada componente, por ejemplo entre más Peleonero te consideres, mayor valor tendrá el primero componente. Algo a notar es que las direcciones en ambas graficas son muy similares pero el primer componente en escala logarítmica rescata más varianza.

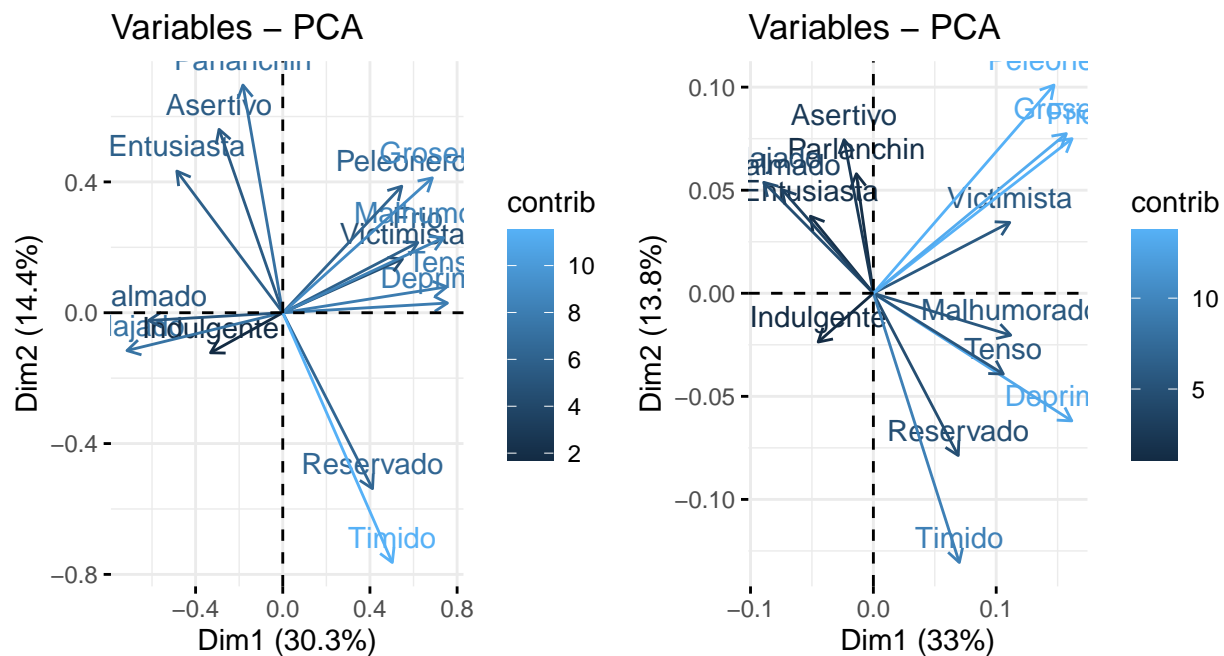
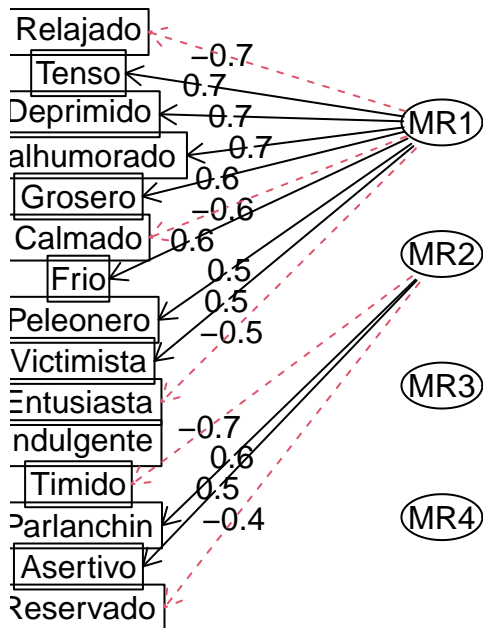


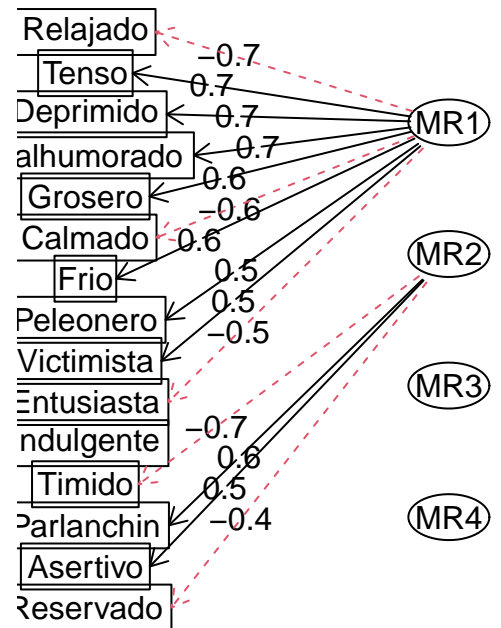
Figura 2: Proyeccion en componentes

Para continuar con el análisis consideramos el enfoque de *Análisis Factorial Exploratorio*, para ello nos apoyamos de la librería psych y la función fa. De nuevo consideramos datos sin escalar, estandarizados y con escala logarítmica, optamos por seguir la recomendación de parallel y considerar 4 factores (Chunk AFE).

Sin escala



Estandarizados



De las gráficas anteriores podemos notar, 3 componentes parecen ser suficiente para resumir la información, a diferencia de componentes principales hemos reducido un poco más la dimensionalidad, además los resultados son muy similares a CP pues las variables de mayor peso se repiten casi todas. Aquí tenemos que el MR1 está asociado con una personalidad “brusca” donde ser frío o rudo la aumentan pero considerarse entusiasta la disminuye, el MR2 lo relacionamos con habilidades sociales, entre más callado menor es este valor y por último el MR3 lo asociamos con el desempeño y dedicación que le ponemos a las cosas, es una relación monótona creciente.

Para decidirnos por un modelo se probaron varias rotaciones como varimax y simplimax, también se consideraron a las variables como ordinales y de nuevo con ayuda de fa se obtuvieron las variables latentes mientras que con principal las componentes principales, más detalles en Chunk RotacionesCP, RotacionesAFE y Ordinales. Optamos por un modelo de Componente principales pues estos recuperan más varianza y dentro de estos el que usa la rotación “cluster” y maneja las variables como ordinales es el mejor rankeado pues recupera un 66 % de varianza total, además nos restringimos a considerar sólo 3 componentes pues el cuarto sólo está relacionado con una variable.

Components Analysis

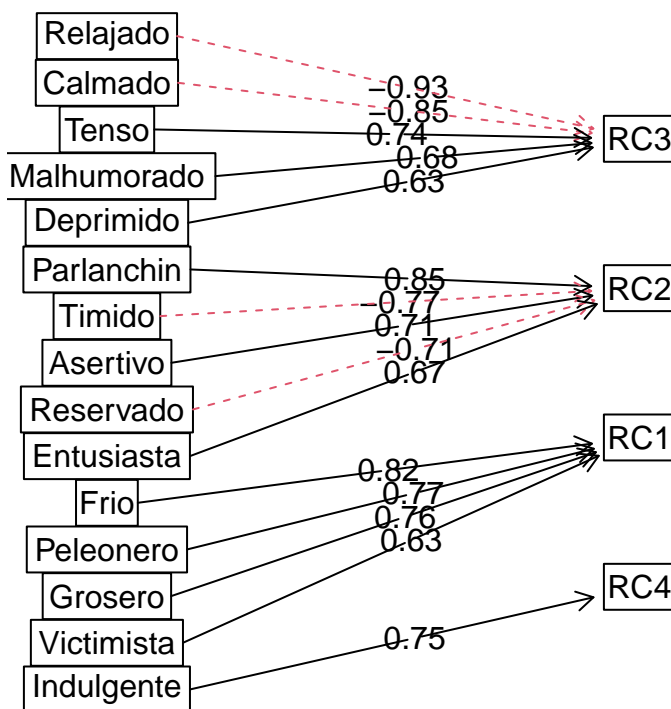


Figura 3: Componentes principales modelo seleccionado

Ya con nuestro modelo seleccionado pasamos a la interpretación, según las Gráfica @ref(fig:Grafica43):

Entre más Victimista, Frío y Rudo te consideras mayor es el valor en el componente 1 mientras que considerarse Indulgente lo reduce y esto tiene sentido pues las características marcan una personalidad hosca. Frío y Peleonero son las que más pesan.

Cuando observamos el componente 2 este es más fácil de interpretar y parece referirse a personas sociables pues pesan variables como Asertivo, Parlanchín, Tímido y Reservado.

Sobre el componente 3 podemos notar todas las variables que pesan tienen correlación positiva, a mayor valor de cualquiera mayor será el valor del componente, ya que contempla variables como Exhaustivo y Perseverante podemos verlo como una persona optimista y/o dedicada.