

3. Componentes principales y análisis factorial exploratorio

Se analiza la personalidad de 228 estudiantes de una universidad de los Estados Unidos a partir de una encuesta resumida en `Dat3Ex.csv`. Las respuestas de 1 “muy en desacuerdo”, 2 “un poco en desacuerdo”, 3 “ni de acuerdo ni en desacuerdo”, 4 “un poco de acuerdo” y 5 “muy de acuerdo”, para un grupo de 44 preguntas, de las cuales tomaremos 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` para el número de componentes a considerar según varianzas y en la siguiente Figura se muestran 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 (Chunk Grafica13, línea 79).

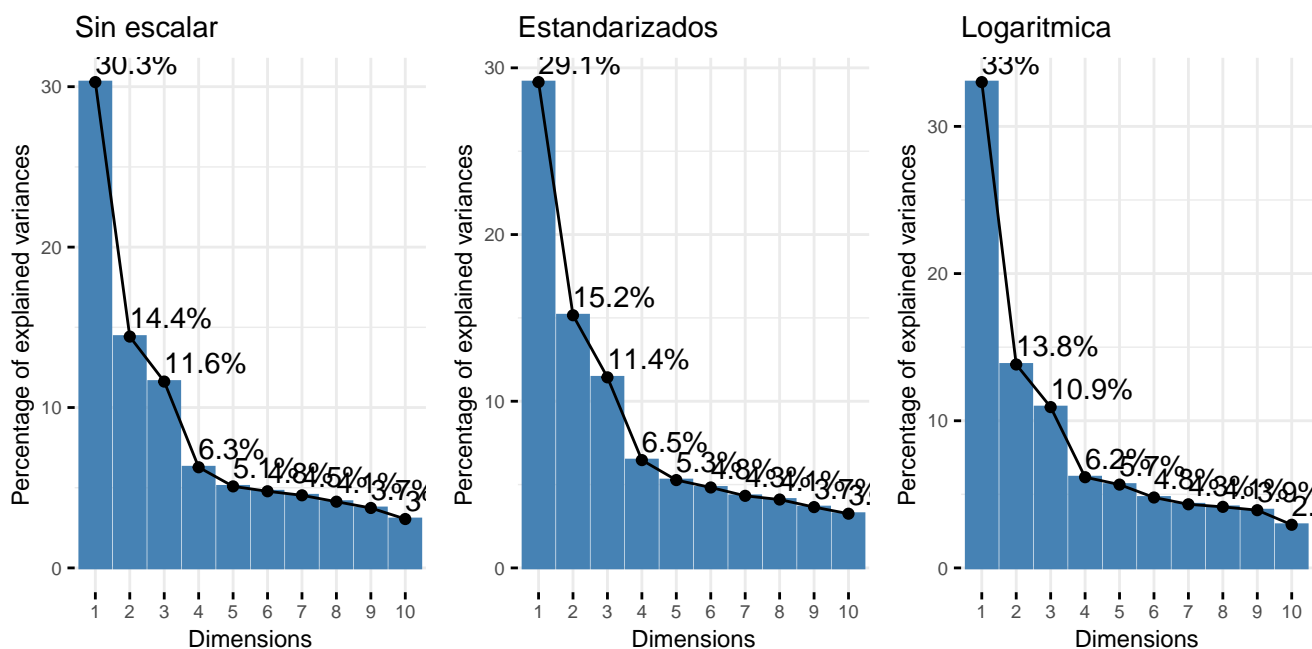


Figura 1: Índices para número de componentes principales

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. La siguiente descripción solamente se presenta para los valores originales, también se hace el ejercicio para datos estandarizados y en logaritmos, sin embargo los resultados son similares por lo que no se describen.

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).

Para mayor interpretabilidad visual tenemos la siguiente Gráfica (Chunk Grafica23, línea 118), sólo se presentan los datos originales y los de escala logaritmica, las estandarizadas son iguales a las originales. Estas son las proyecciones

de las variables de mayor peso en los primeros 2 componentes principales, rescatan la mayor varianza, podemos observar el sentido y magnitud de las fechas para visualizar la influencia de cada variables en cada componente.

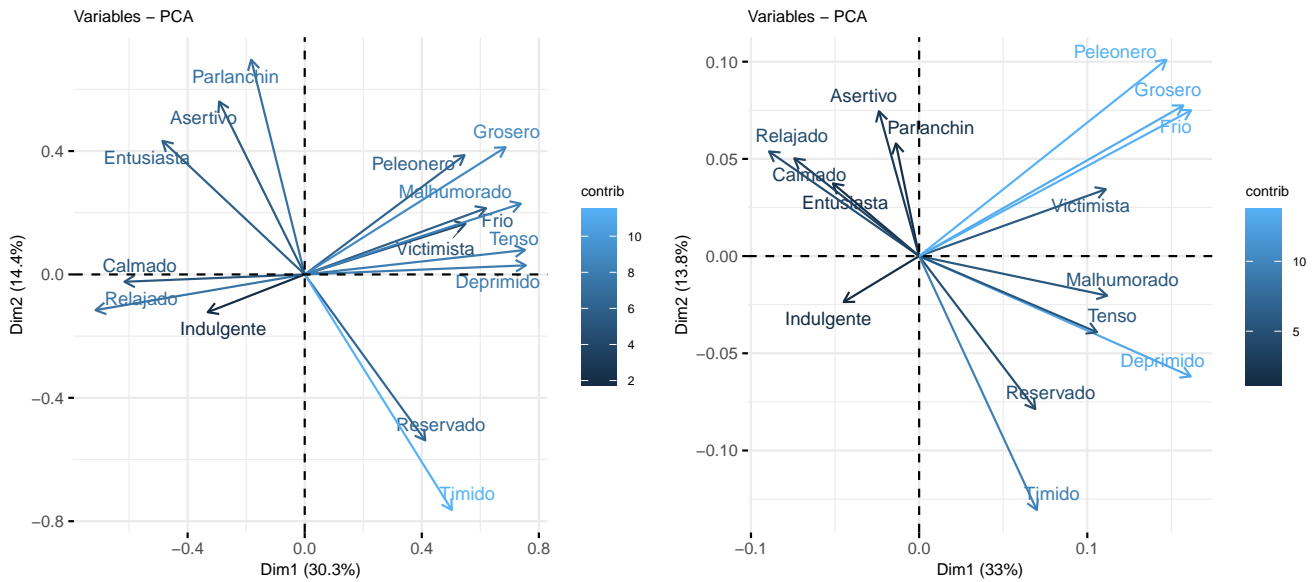
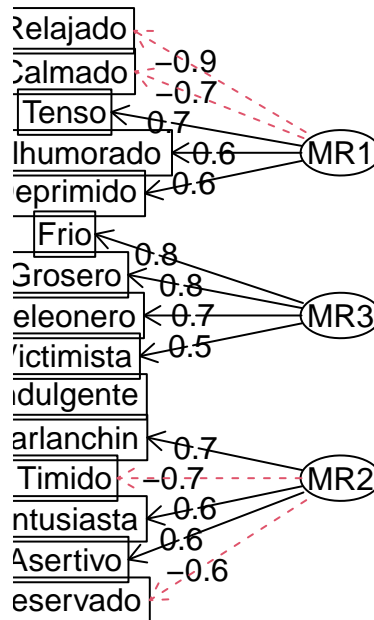
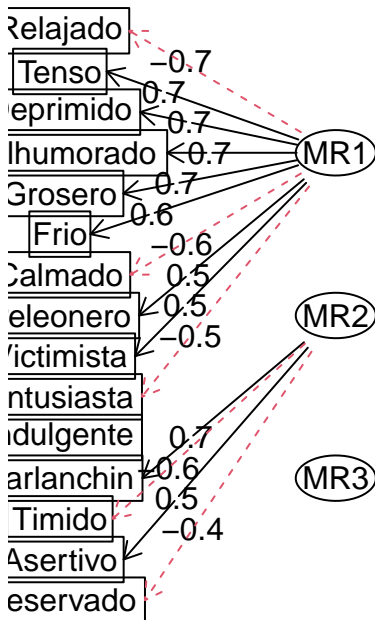


Figura 2: Proyeccion en componentes (izquierda: escala original; derecha:escala log)

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 y estandarizados, optamos por considerar 3 factores, en los 2 casos Indulgente no queda en ninguno (Chunk Factorial, línea 131).

Sin escala

Estandarizados



De las gráficas anteriores podemos notar, 3 componentes parecen ser suficiente para resumir la información, en contraste con componentes principales hemos reducido un poco más la dimensionalidad, además los resultados son muy similares a los componentes principales pues las variables de mayor peso se repiten casi todos los casos.

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 (ver Chunks RotacionesCP, RotacionesAFE y Ordinales; líneas 170, 203 y 228). 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 (Indulgente).

Components Analysis

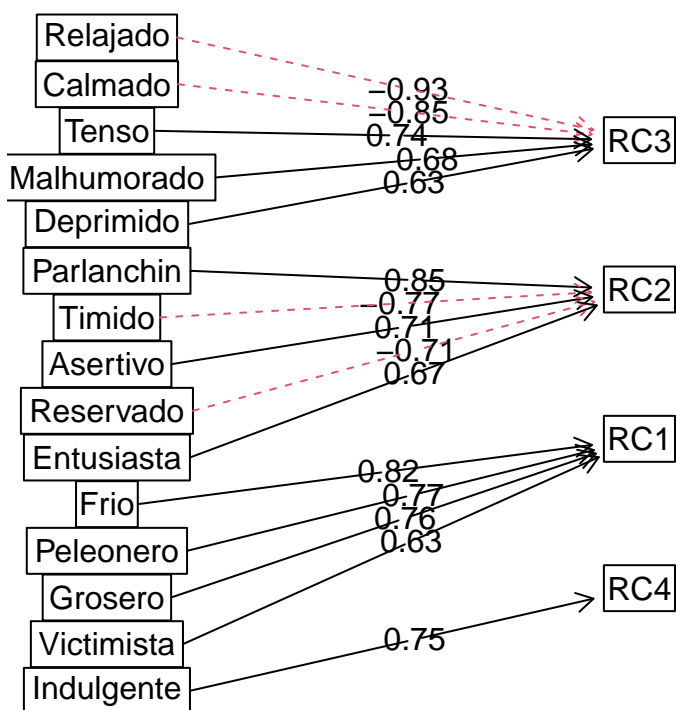


Figura 3: Componentes principales modelo seleccionado

Ya con nuestro modelo seleccionado pasamos a la interpretación, según las Gráfica anterior. El componente 1 corresponde a alumnos victimistas, fríos, groseros y peleoneros. En el componente 2, tenemos alumnos para los que ser asertivo, parlanchín y entusiasta se tiene un mayor relación positiva con el componente y ser tímidos y reservados los mayores valores negativos. Finalmente, en el componente 3 podemos notar mayores relaciones de alumnos deprimidos, malhumorados y tensos, mientras tenemos negativamente a alumnos calmados y relajados.