



Instituto de Estadística  
Ingeniería en Estadística y Ciencia de Datos

# **Análisis factorial sobre datos de rasgos de personalidad del modelo de los cinco grandes obtenidos desde la librería “*Psych*” de R**

Ricardo Cristian Menares Eyzaguirre  
ricardomenares1@gmail.com

Valparaíso, Chile  
9 de Julio de 2022

# Índice general

<b>1. Resumen</b>	<b>2</b>
<b>2. Introducción</b>	<b>3</b>
2.1. Objetivo general . . . . .	6
2.2. Objetivos específicos . . . . .	6
<b>3. Formulación teórica</b>	<b>7</b>
3.1. Estudios anteriores . . . . .	7
3.2. Correlación $\tau$ de Kendall . . . . .	8
3.3. Análisis factorial . . . . .	9
3.4. Análisis de componentes principales . . . . .	11
3.5. Ventaja principal de AF sobre ACP . . . . .	12
<b>4. Metodología</b>	<b>13</b>
4.1. Base de datos . . . . .	13
4.2. Análisis de datos . . . . .	15
4.3. Aplicación análisis factorial . . . . .	22
4.4. Aplicación análisis de componentes principales . . . . .	30
<b>5. Conclusiones</b>	<b>32</b>
<b>6. Referencias</b>	<b>33</b>

# Capítulo 1

## Resumen

El presente estudio busca explorar los resultados de una encuesta sobre el modelo de los cinco grandes de la librería *Psych* de *R*, específicamente en los datos llamados *bfi* que cuenta con 2800 observaciones y 28 variables. Para ello, se realiza un análisis factorial para reducir la dimensionalidad de las variables de cada uno de los 5 rasgos del método en cuestión, logrando obtener una interpretabilidad perfecta. Además, se compara el análisis factorial con el análisis de componentes principales, donde si bien el segundo obtuvo una mayor proporción de variabilidad acumulada explicada, es el primer procedimiento el cual consigue una mayor interpretabilidad según la naturaleza de los datos.

## Capítulo 2

# Introducción

Es sabido que para poder describir la personalidad de una persona de manera perfecta, habría que señalar incontables características de ésta, lo cual sería un trabajo arduo. Además, no se puede saber con exactitud si se están tomando todas las variables que pueden diferenciar a un sujeto de otro. Es por lo anterior, que el investigar cuáles características son las más relevantes de una persona, se vuelve indispensable para poder diferenciar y catalogar diferentes personalidades.

A lo largo de los años, han sido creados diferentes modelos que seleccionan distintos rasgos, donde el más conocido es el modelo de los cinco grandes (model Big five, en inglés) que fue descubierto experimentalmente en una investigación sobre las descripciones de personalidad de unas personas hacían de otras (Goldberg, 1993).

Los 5 grandes rasgos de personalidad según Goldberg son los siguientes:

- **Apertura a la experiencia (O):** inventiva/curiosa vs. consistente/cautelosa
- **Escrupulosidad (G):** eficiente/organizado frente a extravagante/descuidado
- **Extraversión (E):** extrovertida/enérgica vs. solitaria/reservada
- **Amabilidad (A):** amigable/compasivo vs. crítico/racional
- **Neuroticismo (N):** sensible/nervioso vs. resistente/seguro

Más específicamente, se destaca las siguientes características de cada factor:

- Apertura a la experiencia (O): El primer rasgo que se identifica en este modelo es la apertura a la experiencia, acuñado como el Factor O. Este rasgo está totalmente relacionado con la capacidad humana de buscar nuevas experiencias en nuestra vida, así mismo también tiene que ver con la habilidad de visualizar un futuro de forma creativa.

Las personas con un nivel elevado de apertura a la experiencia son perfiles imaginativos, que aprecian la cultura y que consiguen establecer relaciones de equipo con los demás. Este tipo de personas persiguen el cambio continuo, ya que están seguras de que si se aferran a ideas fijas significa aferrarse al inmovilismo y a la quietud.

- Responsabilidad (C): El rasgo de la responsabilidad tiene que ver con la habilidad del autocontrol y la capacidad de diseñar métodos de acción eficaces. Las personas que tienen un alto grado de responsabilidad son grandes planificadores y organizadores, además de tener un fuerte compromiso con los objetivos y metas.

A su vez, este tipo de perfiles son vistos por los demás como personas confiables y escrupulosas. En el caso de encontrar individuos con una puntuación extrema en este rasgo de personalidad se puede observar comportamientos demasiados perfeccionistas e incluso obsesivos, por ese motivo las personas de factor C requieren un cierto equilibrio para no caer en el extremo.

- Extraversión (E): La extraversión tiene que ver con el grado en el que el sujeto está abierto con los demás, es decir, el factor E analiza cuánto le agrada a un sujeto estar rodeado de los demás.

Lógicamente, el perfil opuesto es el individuo introvertido, estas personas se caracterizan por tener una personalidad reservada, lo que los lleva a que en muchas ocasiones puedan ser juzgados como antipáticos. Los perfiles introvertidos son más reflexivos que los extrovertidos y les gusta menos formar parte de grupos elevados de personas, prefieren establecerse en una rutina y pasar tiempo con la familia.

- Amabilidad (A): La amabilidad es el rasgo que muestra el grado de tolerancia y respeto de una persona. Una persona amable será aquella que confía en la honestidad de la palabra, su vocación es prestar la ayuda a aquellos que lo necesiten. La humildad, la sencillez y la empatía son los atributos básicos de las personas amables.

- Neuroticismo (N): El neuroticismo es la resiliencia con la que una persona afronta las situaciones problemáticas en la vida, los individuos tranquilos no suelen sentir rabia y huyen del enfado, su estado es animado y saben gestionar correctamente las crisis personales. En el polo opuesto nos encontramos a las personas que se caracterizan por tener un comportamiento impredecible, ya que sus reacciones varían sin que sea muy claro por qué.

Es de vital importancia el poder investigar los rasgos de personalidad de las personas, ya sea para labores de contratación de personal en una empresa o entidad gubernamental, selección de estudiantes en una universidad e incluso para labores de análisis criminalístico. Es por lo anterior, que se vuelve indispensable estudiar el método de los cinco grandes, de manera que se pueda establecer conexiones entre sus variables. Es por ello, que el presente estudio busca realizar un análisis factorial para identificar que variables comparten información parecida, a través de los datos de la librería `Psych` de `R`.

## 2.1. Objetivo general

Examinar la estructura interna del *Big Five Inventory* (BFI) mediante un análisis factorial, buscando reducir la dimensionalidad de los datos y lograr una interpretabilidad satisfactoria.

## 2.2. Objetivos específicos

1. Examinar el modelo teórico elegido para los datos obtenidos.
2. Aplicar experimentalmente diversos modelos de análisis factorial para las variables propuestas.
3. Comparar el método análisis factorial y de análisis de componentes principales.
4. Seleccionar el mejor método tomando en consideración la proporción de variabilidad acumulada explicada y la interpretabilidad obtenida.

## Capítulo 3

# Formulación teórica

### 3.1. Estudios anteriores

Existen muchas investigaciones acerca de las variables del *Big five*, debido a la popularidad que este modelo tiene. Comenzando, se observa la investigación de Barrio & et al (2006) en donde explora sobre las diferencias por edad y sexo en los cinco factores de personalidad en una población infantil en España, administrado a 852 escolares (501 varones y 351 mujeres) de edades comprendidas entre 8 y 15 años. Como conclusión, destacan que encontraron patrones de personalidad en los niños, donde los sujetos de más edad presentaron significativamente más características de neuroticismo y extraversión y menores niveles de conciencia, apertura y agradabilidad, mientras que las chicas se caracterizaron por mayores rasgos de conciencia y agradabilidad.

Sumado a lo anterior, Lovik & et al. (2017) examinaron una versión en holandés del modelo de los cinco grandes, usando una muestra procedente del estudio Divorce in Flanders. Se demostró que el BFI holandés tiene buenas propiedades psicométricas y una estructura factorial clara comparable a una muestra holandesa anterior y de la investigación internacional. Además, comparan el análisis factorial con rotación VARIMAX con el análisis de componentes principales con rotación VARIMAX, no difiriendo sustancialmente, sin embargo, el análisis factorial explicó de mejor manera la encuesta.

Finalizando, Domínguez & César (2018) realizaron un estudio de las variables del *Big five* en 332 universitarios peruanos con (82.83% mujeres) entre 16 y 48 años. En cuanto a los resultados. En conclusión, las versiones estudiadas evidencian indicadores psicométricos favorables que posibilitarían su uso en evaluaciones de grupo e investigación básica.



### 3.2. Correlación $\tau$ de Kendall

En consiguiente, se busca un método de correlación para las variables estudiadas, tomando en cuenta su naturalidad, a lo que se llega a un coeficiente de correlación particular. En estudios exploratorios y confirmatorios, las pruebas de correlaciones  $\tau$  de Kendall, propuesto por Kendall (1938), son utilizadas para el tratamiento de variables puramente ordinales. Es justamente esta prueba que se ocupará para tratar las variables del Big five, pues éstas son ordinales.

Esta clase de correlación es una prueba de hipótesis no paramétrica para la dependencia estadística basada en el coeficiente  $\tau$ . Para calcularla, primero se define un conjunto de observaciones de las variables aleatorias conjuntas  $X$  e  $Y$ , de modo que todos los valores de  $x_i, y_i$  son únicos (los vínculos se ignoran por simplicidad). Cualquier par de observaciones  $(x_i, y_i)$  y  $(x_j, y_j)$  con  $i < j$  se dice que son un par concordante si el orden de clasificación de  $x_i, x_j$  e  $y_i, y_j$  está de acuerdo: es decir, si ambos  $x_i > x_j$  e  $y_i > y_j$  o al revés, si ocurre lo contrario se dice que son discordantes. Entonces, el coeficiente de Kendall se define como:

$$\tau = \frac{(\text{n}^\circ \text{ de pares concordantes}) - (\text{n}^\circ \text{ de pares discordantes})}{\binom{n}{2}}, \quad -1 < \tau < 1 \quad (3.1)$$

### 3.3. Análisis factorial

En cuanto al análisis a realizar en este estudio, se ocupa un análisis factorial, en el cual según Rencher (2019) se representan las variables  $y_1, y_2, \dots, y_p$  como combinaciones lineales de variables aleatorias  $f_1, f_2, \dots, f_m (m < p)$  llamadas factores. Como las variables originales, los factores varían de un individuo a otro; pero a diferencia de las variables, los factores no se pueden medir ni observar. Si las variables  $y_1, y_2, \dots, y_p$  están medianamente correlacionadas, la dimensionalidad efectiva es menor que  $p$ . Entonces, el objetivo del análisis de factor es reducir la redundancia entre variables utilizando un menor número de factores.

Se define primero suponiendo una muestra  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p$  de una población homogénea con vectores de media  $\bar{\boldsymbol{\mu}}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ .

Luego, el modelo de análisis de factor expresa cada variable como una combinación lineal de factores comunes subyacentes  $f_1, f_2, \dots, f_m$ , con un término de error que lo acompaña a dar cuenta de la parte de la variable que es única. El modelo para cualquier vector  $\mathbf{y}$  es el siguiente:

$$\begin{aligned} y_1 - \mu_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1m}f_m + \epsilon_1 \\ y_2 - \mu_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2m}f_m + \epsilon_2 \\ &\vdots \\ y_p - \mu_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pm}f_m + \epsilon_p \end{aligned} \quad (3.2)$$

Donde los  $\lambda_{ij}$  son llamados *loadings*, además,  $m < p$ . Luego, se asume  $E(f_j) = 0$ ,  $Var(f_j) = 1$  y  $Cov(f_j, f_k) = 0$  para todo  $j$  y  $k = 1, 2, \dots, m$ . Además, lo supuestos para  $\epsilon_i, i=1, 2, \dots, p$ , son similares, pero se debe permitir que cada  $\epsilon_i$  tenga varianza diferente, sea  $\sigma^2$  pues representa la parte residual de  $y_i$  que no es común con las otras variables. Notar entonces, que la varianza de  $y_i^* = y_i - \bar{\mu}$  es:

$$Var(y_i^*) = \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2 + \sigma_i^2 \quad (3.3)$$

En donde las sumas de los  $\lambda_{im}^2$  se le dice comunalidad, que es la parte de variabilidad común entre factores, por otro lado, la parte de  $\sigma_i^2$  es llamada varianza específica.

Existen diversos métodos para estimar los factores en el análisis factorial, en el caso de esta investigación se usa el método de máxima verosimilitud por sus potentes propiedades. En cuanto el número de factores a usar, se busca en principio ocupar 5 (por el big five), sin embargo, se estará atento al % de varianza

que se está ocupando, además de cuantos factores son mayores a 1.

Cabe destacar, que a veces es difícil obtener una interpretación a los factores resultantes al aplicar un análisis factorial, sin embargo, según Rencher (2019) los *loadings* de los factores rotados conservan las propiedades esenciales de los originales, esto es, satisfacer todos los supuestos básicos y reproducir la matriz de covarianza. Es por eso que se pueden aplicar diversos tipos de rotaciones, en esta investigación, de ser necesario, se aplicará una rotación VARIMAX propuesta por Kaiser (1958) que resulta ser un método de rotación ortogonal que minimiza el número de variables que tienen altas cargas en cada factor. Este modelo se representa de la siguiente manera:

$$R_{\text{VARIMAX}} = \underset{R}{\operatorname{argm\acute{a}x}} \left( \frac{1}{p} \sum_{j=1}^k \sum_{i=1}^p (\Lambda R)_{ij}^4 - \sum_{j=1}^k \left( \frac{1}{p} \sum_{i=1}^p (\Lambda R)_{ij}^2 \right) \right) \quad (3.4)$$

Donde  $\Lambda$  es la matriz de las comunales,  $R$  la matriz de *loadings* y  $p$  el número total de variables.

### 3.4. Análisis de componentes principales

En adición y luego de analizar la investigación de Lovik & et al (2017), se busca comparar los resultados del análisis factorial mediante otra técnica, la cual resulta ser el análisis de componentes principales (ACP) creado por Pearson (1908), que es otra técnica para reducir dimensionalidad, de las más conocidas y usadas en la actualidad. Este método busca un número de dimensiones que sean lo más interesantes posible medido por la variación de las observaciones en cada una de las dimensiones. Éstas dimensiones encontradas por ACP son combinaciones lineales de las  $p$  variables, donde el primer componente principal de un conjunto de variables  $X_1, X_2, \dots, X_p$  está dado por la combinación lineal normalizada:

$$\mathbf{z}_1 = \Phi_{11}\mathbf{X}_1 + \Phi_{21}\mathbf{X}_2 + \dots + \Phi_{p1}\mathbf{X}_p \quad (3.5)$$

Con:  $\sum_{j=1}^p \Phi_{j1}^2 = 1$ , donde los  $\Phi_{j1}, j=1, 2, \dots, p$ , son comúnmente llamados “loadings”. Así, éstos conforman el vector de *loadings* del primer componente principal:  $\Phi_1 = (\Phi_{11}, \Phi_{21}, \dots, \Phi_{p1})'$ .

Para poder calcular el primer componente principal, se asume que cada variable de  $\mathbf{X}$  ha sido centrada en 0. Luego, se busca la combinación lineal de las variables que tenga mayor varianza:

$$z_{i1} = \Phi_{11}X_{i1} + \Phi_{21}X_{i2} + \dots + \Phi_{p1}X_{ip} \quad (3.6)$$

Es decir, se busca maximizar el siguiente problema:

$$\begin{aligned} \max_{\Phi_0, \Phi_1, \dots, \Phi_p} & \left[ \frac{1}{n} \sum_{i=1}^n \left( \sum_{j=1}^p \Phi_{j1} x_{ij} \right) \right] \\ \text{S.a: } & \sum_{j=1}^p \Phi_{j1}^2 = 1 \end{aligned} \quad (3.7)$$

Luego, las demás componentes principales se obtienen mediante la misma técnica, pero sin tomar en consideración a los primeros componentes que ya se han obtenido. Así, cada componente que se va obteniendo tiene menor varianza explicada que el anterior.

### **3.5. Ventaja principal de AF sobre ACP**

Según Rencher (2019) una de las ventajas del análisis factorial por sobre el análisis de componentes principales es su flexibilidad para la interpretación, pues es posible hacer rotaciones al análisis factorial y obtener distintos resultados, no así en ACP, que se obtienen siempre los mismos. Es por esto, que en este tipo de estudios es conveniente usar un análisis factorial por sobre un análisis de componentes principales.

## Capítulo 4

# Metodología

### 4.1. Base de datos

Se cuenta con la base de datos “**bfi**” de la librería de R “**Pysch**”. Cuenta con 2800 observaciones y 28 variables, de las cuales 25 corresponden a 5 preguntas de cada factor del modelo de los cinco grandes y 3 variables demográficas de la persona, tales como el sexo (gender), educación (education) y edad (age). Las variables, explicadas una por una, son las siguientes:

1. **A1:** Respuesta de 1 al 6 a: Soy indiferente a los sentimientos de los demás.
2. **A2:** Respuesta de 1 al 6 a: Pregunto sobre el bienestar de los demás.
3. **A3:** Respuesta de 1 al 6 a: Sé consolar a los demás.
4. **A4:** Respuesta de 1 al 6 a: Amo a los niños.
5. **A5:** Respuesta de 1 al 6 a: Hago que la gente se sienta a gusto.
6. **C1:** Respuesta de 1 al 6 a: Soy exigente en mi trabajo.
7. **C2:** Respuesta de 1 al 6 a: Continúo hasta que todo esté perfecto.
8. **C3:** Respuesta de 1 al 6 a: Hago las cosas de acuerdo a un plan.
9. **C4:** Respuesta de 1 al 6 a: Hago las cosas a medias.
10. **C5:** Respuesta de 1 al 6 a: Pierdo mi tiempo.
11. **E1:** Respuesta de 1 al 6 a: No hablo mucho.
12. **E2:** Respuesta de 1 al 6 a: Me resulta difícil acercarme a los demás.
13. **E3:** Respuesta de 1 al 6 a: Sé cautivar a la gente.
14. **E4:** Respuesta de 1 al 6 a: Hago amigos fácilmente.

15. **E5:** Respuesta de 1 al 6 a: Me hago cargo.
16. **N1:** Respuesta de 1 al 6 a: Me enoja fácilmente.
17. **N2:** Respuesta de 1 al 6 a: Me irrito fácilmente.
18. **N3:** Respuesta de 1 al 6 a: Tengo cambios de humor frecuentes.
19. **N4:** Respuesta de 1 al 6 a: A menudo me siento triste.
20. **N5:** Respuesta de 1 al 6 a: Me asusto fácilmente.
21. **O1:** Respuesta de 1 al 6 a: Estoy lleno de ideas.
22. **O2:** Respuesta de 1 al 6 a: Evito el material de lectura difícil.
23. **O3:** Respuesta de 1 al 6 a: Llevo las conversaciones a un nivel superior.
24. **O4:** Respuesta de 1 al 6 a: Paso tiempo reflexionando sobre las cosas.
25. **O5:** Respuesta de 1 al 6 a: No profundizo en temas.
26. **gender:** Género, hombre=1, mujer=2.
27. **education:** Educación, 1=Escuela secundaria, 2= Escuela secundaria finalizada, 3=Educación superior, 4=Universitario graduado, 5=Postgrado graduado.
28. **age:** Edad en años.

En consiguiente, se presenta una mirada resumida de las variables del *Big five*, pues al ser una encuesta es lógico que haya personas que no contesten algunas preguntas:

```
> data(bfi)
> head(bfi)
```

	A1	A2	A3	A4	A5	C1	C2	C3	C4	C5	E1	E2	E3	E4	E5	N1	N2	N3	N4	N5	O1	O2	O3	O4	O5	gender	education	age
61617	2	4	3	4	4	2	3	3	4	4	3	3	3	4	4	3	4	2	2	3	3	6	3	4	3	1	NA	16
61618	2	4	5	2	5	5	4	4	3	4	1	1	6	4	3	3	3	3	5	5	4	2	4	3	3	2	NA	18
61620	5	4	5	4	4	4	5	4	2	5	2	4	4	4	5	4	5	4	2	3	4	2	5	5	2	2	NA	17
61621	4	4	6	5	5	4	4	3	5	5	5	3	4	4	4	2	5	2	4	1	3	3	4	3	5	2	NA	17
61622	2	3	3	4	5	4	4	5	3	2	2	2	5	4	5	2	3	4	4	3	3	3	4	3	3	1	NA	17
61623	6	6	5	6	5	6	6	6	1	3	2	1	6	5	6	3	5	2	2	3	4	3	5	6	1	2	3	21

## 4.2. Análisis de datos

Primero, se verifica si hay valores perdidos:

```
> sum(is.na(bfi))  
[1] 731  
> sum(is.na(bfi[1:25]))  
[1] 508
```

Es notorio que hay muchos valores perdidos, pero no se puede realizar ninguna transformación debido a que no se tiene información sobre estos, por lo que se decide eliminar las observaciones que tienen valores perdidos, quedando:

```
> dim(bfi)  
[1] 2800  28  
> sum(is.na(bfi))  
[1] 731  
> bfi<-bfi[complete.cases(bfi),] #se eliminan los valores nan  
> sum(is.na(bfi)) #ya no hay  
[1] 0  
> dim(bfi)  
[1] 2236  28
```

Se aprecia que ya no hay valores perdidos y quedaron entonces 2236 observaciones.

Ahora, se presenta una tabla de estadísticas descriptivas de las variables:



Variable	Media	desv.est.	mediana	mínimo	máximo
A1	2.41	1.41	2	1	6
A2	4.80	1.17	5	1	6
A3	4.60	1.30	5	1	6
A4	4.70	1.48	5	1	6
A5	4.56	1.26	5	1	6
C1	4.50	1.24	5	1	6
C2	4.37	1.32	5	1	6
C3	4.30	1.29	5	1	6
C4	2.55	1.38	2	1	6
C5	3.30	1.63	3	1	6
E1	2.97	1.63	3	1	6
E2	3.14	1.61	3	1	6
E3	4.00	1.35	4	1	6
E4	4.42	1.46	5	1	6
N1	2.93	1.57	3	1	6
N2	3.51	1.53	4	1	6
N3	3.22	1.60	3	1	6
N4	3.19	1.57	3	1	6
N5	2.97	1.62	3	1	6
O1	4.82	1.13	5	1	6
O2	2.71	1.57	2	1	6
O3	4.44	1.22	5	1	6
O4	4.89	1.22	5	1	6
O5	2.49	1.33	2	1	6
gender	1.67	0.47	2	1	2
education	3.19	1.11	3	1	5
age	28.78	11.13	26	3	86

Cuadro 4.1: Estadísticas descriptivas de las variables

Se observa que según la media de las variables del big five, las preguntas A1=Indiferente a los demás, C1=Exigente en el trabajo, E1=No hablar mucho, N1=enojarse fácilmente, O2=Evitar material de lectura difícil, O5 no profundizar en temas, obtuvieron los promedios menores a 3, por lo cual, en promedio las personas eligen puntajes bajos en variables que indican características que son “negativas”, a excepción de la variable C1 que puede ser considerada buena. Lo anterior, es llamativo, pues hay también más variables que toman cosas negativas como N2 pero que no tuvieron puntajes menores a 3. Se aprecia también que en las desviaciones estándar son bastante parecidas en todas las variables del big five, en cuanto a la mediana, la mayoría tiene un 50 % de sus datos aproximadamente menores que 5, mientras que otras variables (las que tienen media baja) tienen un 50 % de los datos aproximadamente menores que 3 y 2, lo cual es una

diferencia notoria, que posiblemente es debido a la naturaleza de las preguntas. En cuanto al mínimo y máximo, se aprecia que todas tienen resultados extremos.

Siguiendo con lo anterior, se presentan gráficos de barra para las variables que indican sobre la amabilidad (A1, A2, A3, A4, A5):

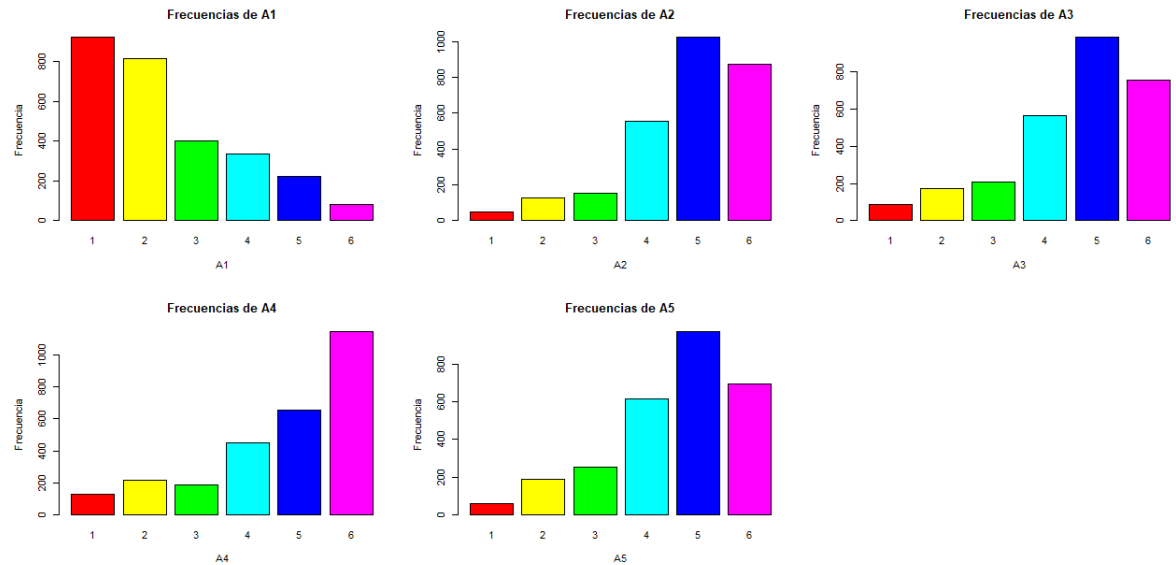


Figura 4.1: Gráfico de barras para las variables indicadoras de amabilidad.

Se aprecia que las variables A1= Soy indiferente a los sentimientos de los demás, A3=Saber consolar a los demás y A4=Amar a los niños, tienen resultados parecidos, lo cual es llamativo, pues si bien las variables A3 y A4 tiene lógica que se comporten similares, A1 describe una característica que no necesariamente tenga que ver con las otras. Por otro lado, A2=preguntar sobre el bienestar de los demás y A5= hacer que la gente se sienta a gusto tienen un comportamiento semejante, lo cual es bastante lógico.

Luego, se presentan los gráficos de barra de las variables que indican la escrupulosidad (C1,C2,C3,C4,C5):

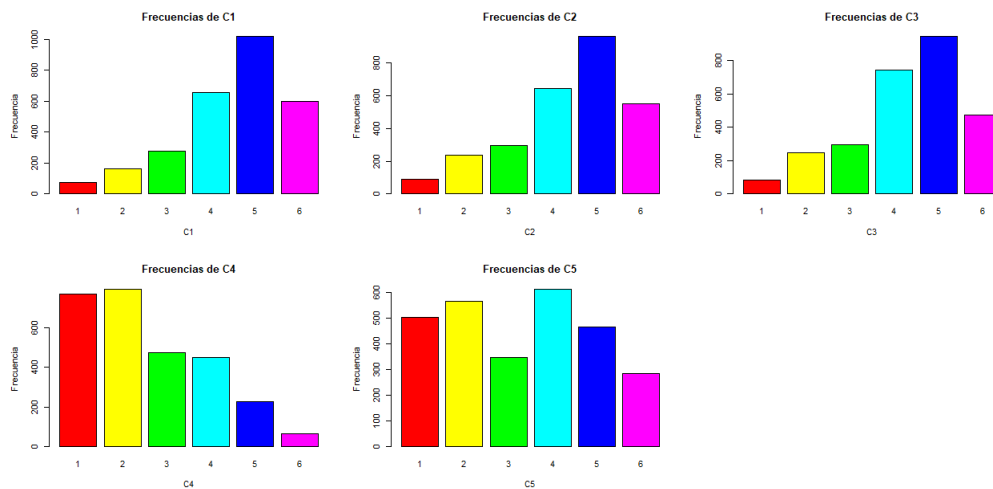


Figura 4.2: Gráficos de barras para las variables indicadoras de escrupulosidad

Se observa que las variables C1= Exigente en mi trabajo, C2= continuó hasta que todo esté perfecto y C3=hacer las cosas de acuerdo a un plan tienen comportamientos parecidos, lo cual es bastante lógico, pues son preguntas que van de la mano. En cuanto a las variables C4= hacer las cosas a media y C5= perder el tiempo, tienen un comportamiento similar y contrario a las variables C1, C2 y C3, la razón es lógica, pues son preguntas opuestas.

En consiguiente, se presentan los gráficos de barra de las variables que indican la extraversión (E1,E2,E3,E4,E5):

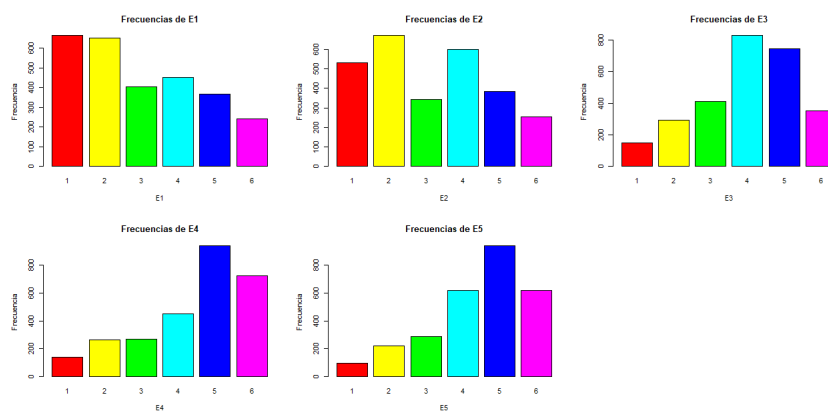


Figura 4.3: Gráficos de barra para variables indicadoras de extraversión.

Es notorio que las variables E1=no hablar mucho y E2=resulta difícil acercarse a los demás, tienen un comportamiento similar, donde se concentra la mayoría de datos hacia valores bajos y medianos (3,4,5,6) mientras que E3=saber cautivar a la gente, E4= hacer amigos fácilmente y E5= hacerse cargo es lo contrario, teniendo la mayoría de frecuencias en 4,5 y 6 lo cual es lógico según la naturalidad de las variables.

Siguiendo con lo anterior, se presentan los gráficos de barra de las variables que indican en neuroticismo (N1,N2,N3,N4,N5):

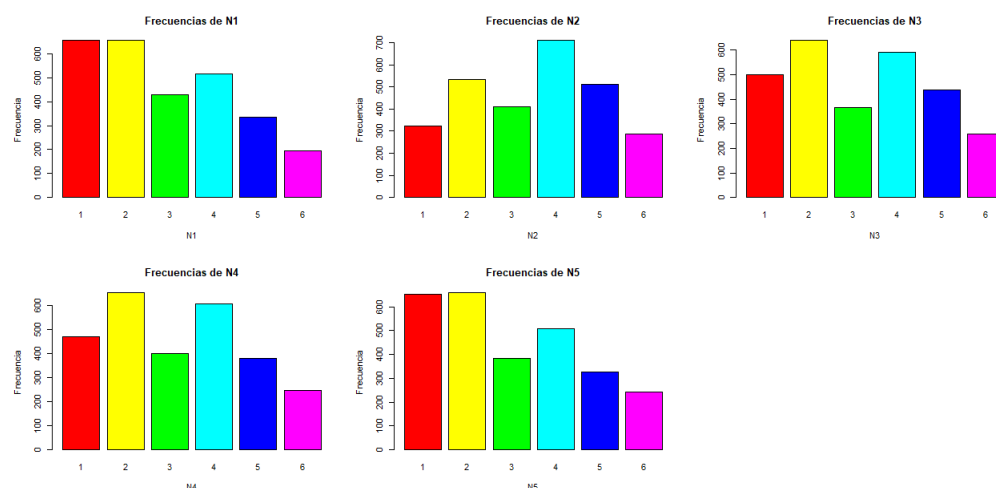


Figura 4.4: Gráficos de barras para variables indicadoras de neuroticismo.

Es claro ver que todas las variables tienen un comportamiento bajo-medio, destacando en varias la frecuencia de 4. Tiene sentido, pues esta variable es normal que la gente prefiera responder un nivel medio, pues todos tenemos algunas veces cambios de humor o enojos.

Finalizando, se presentan gráficos de barra de las variables que indican apertura a la experiencia (O1,O2,O3,O4,O5):

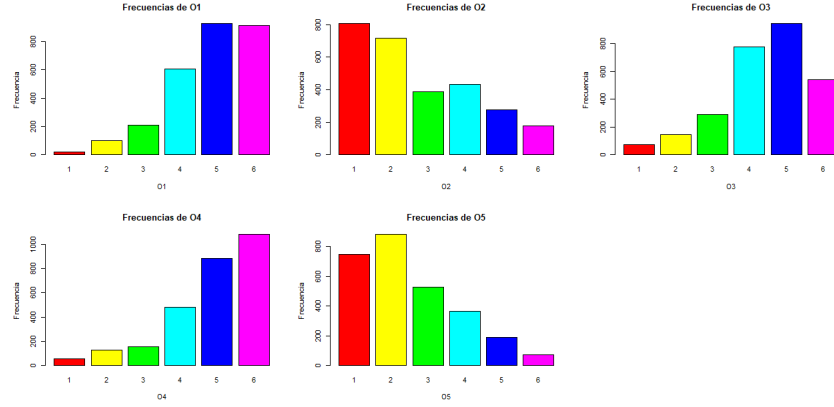


Figura 4.5: Gráficos de barra para variables indicadoras de apertura a la experiencia.

Es claro ver que las variables O1=estar lleno de ideas, O3=llevar las conversaciones a nivel superior y O4=pasar tiempo reflexionando sobre cosas tienen un comportamiento similar en las 4 primeras frecuencias, lo cual es lógico, pues son variables que preguntan cosas muy pensativas, donde las personas buscan destacar en este ámbito, además se destaca que para las frecuencias 5 y 6 tienen valores altos y además varían las proporciones lo cual puede ser producto al azar, mientras que O2=evitar el material de lectura fácil con O5=no profundizar en temas parecieran ser idénticas y contrarias a las demás, lo cual es lógico, pues son preguntas opuestas.

Para poder analizar si las variables de cada grupo del big five e incluso entre variables de distintos grupos en cuestión, tienen un comportamiento correlacional, se realiza un análisis mediante la correlación  $\tau$  de Kendall, el cual se refleja representado en la siguiente matriz:

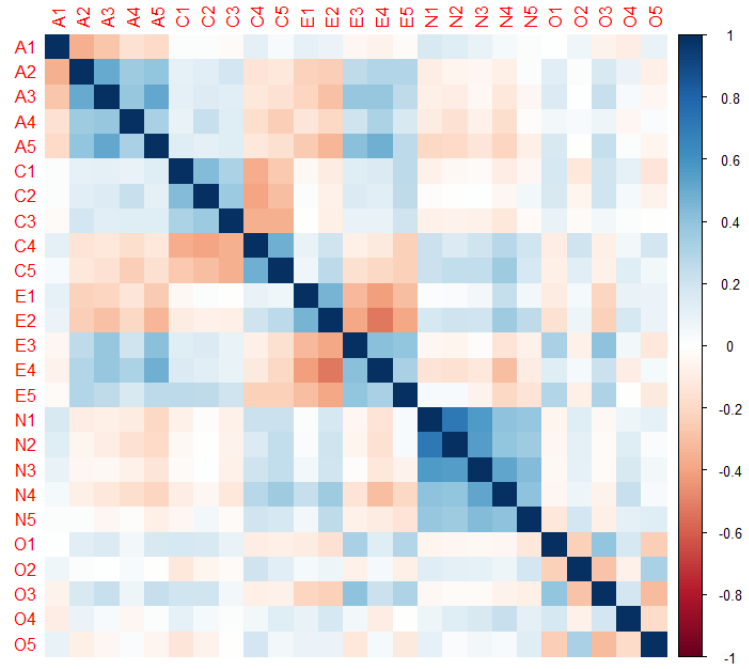


Figura 4.6: Matriz de correlación  $\tau$  de Kendall para todas las variables del Big five.

Es imprescindible indicar que en este tipo de análisis, donde se tienen variables que representan a muchas otras más como lo son las del Big five, una correlación de 0.3 puede ser mucho, por lo cual, se aprecia que las variables A2, A3, A4 y A5 tienen correlación significativa con C4, C5, E1 y E2 y muy poca entre ellas, el resto de correlaciones que no sean entre su propia naturaleza de variable no sobrepasan en valor absoluto el 0.2 lo cual es bastante bueno en este set de datos.

### 4.3. Aplicación análisis factorial

Primero, se realiza un gráfico que indica el número de factores que requiere cada modelo, en donde la decisión heurística se basa en que si los valores propios de los factores sean mayores a 1, también se observa cuando existe un punto de inflexión en la curva del *screeplot*:

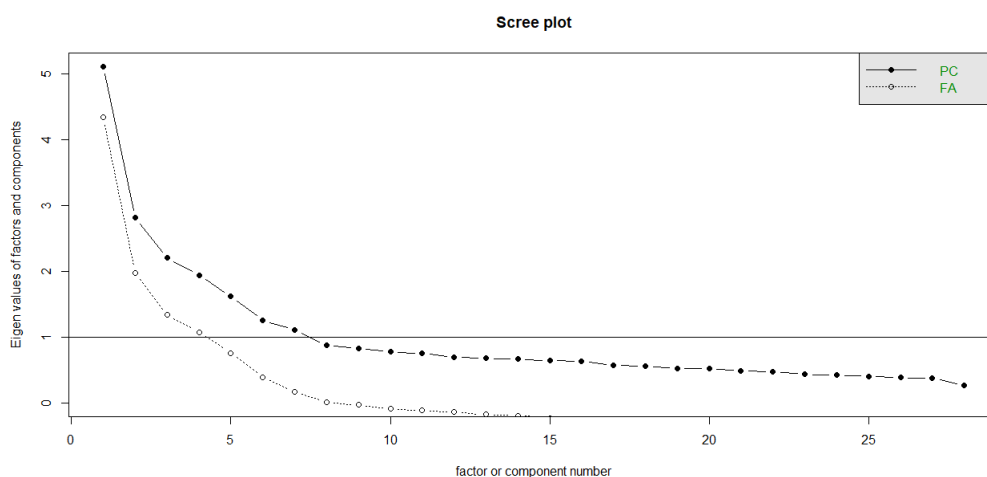


Figura 4.7: Valores propios según el número de factores y de componentes principales tanto en AF y ACP, respectivamente.

Es notorio que para el análisis factorial se obtienen 4 factores significativos, al contrario de lo que se propuso en el trabajo de que deberían ser 5. Sin embargo, el quinto factor no se aleja tanto del 1, por tanto, se tomarán en cuenta los 2.

En consiguiente, se presenta el gráfico del árbol AF con 4 factores:

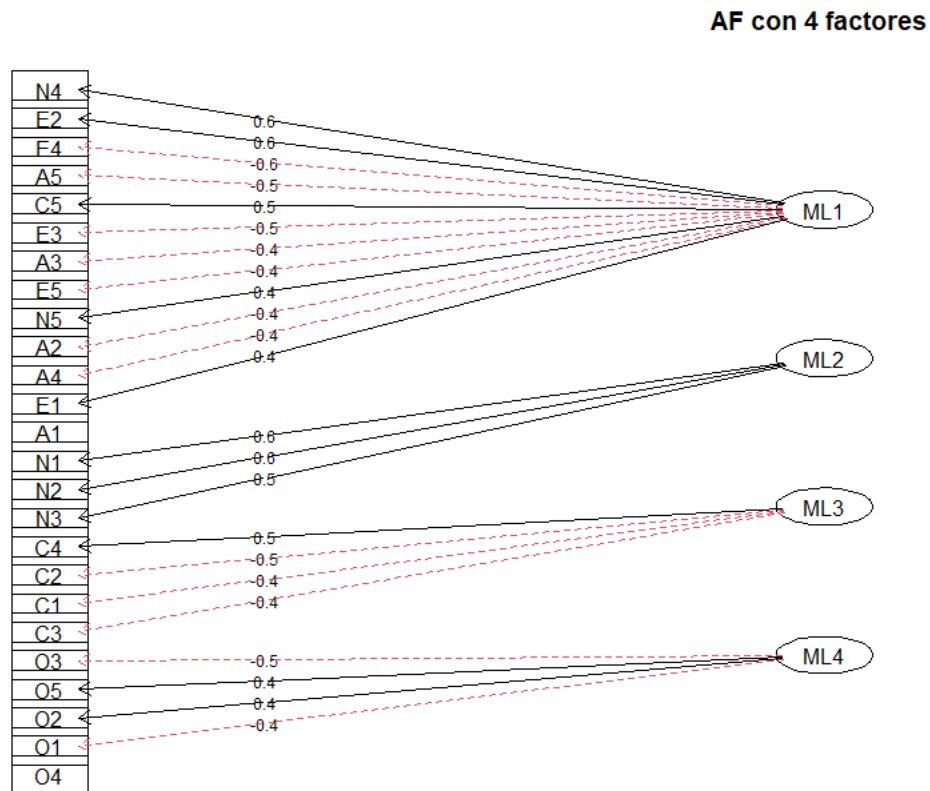


Figura 4.8: Árbol de Análisis factorial con 4 factores

Es notorio que al tener 4 factores se juntan la mayoría de variables  $A_i$  y  $E_i$ , además las variables  $A_1$  y  $O_4$  no forma parte de ninguna variable. Las otras variables  $O_i$  si se juntan en una variable, lo mismo con las  $C_i$ .

Ahora se presenta el árbol del modelo de análisis factorial pero ahora con 5 factores:





No se logra separar las variables como uno pensara, se destaca que parecieran explicar la misma información la mayoría de las variables  $A_i$ ,  $N_i$  y  $E_i$ , además, la E5 por si sola origina un factor.

Debido a que no se ha logrado encontrar un resultado en donde los factores sean interpretables, por lo cual se realiza una rotación VARIMAX para analizar que resultados se obtienen. El gráfico de análisis factorial con rotación VARIMAX y 4 factores se presenta a continuación:

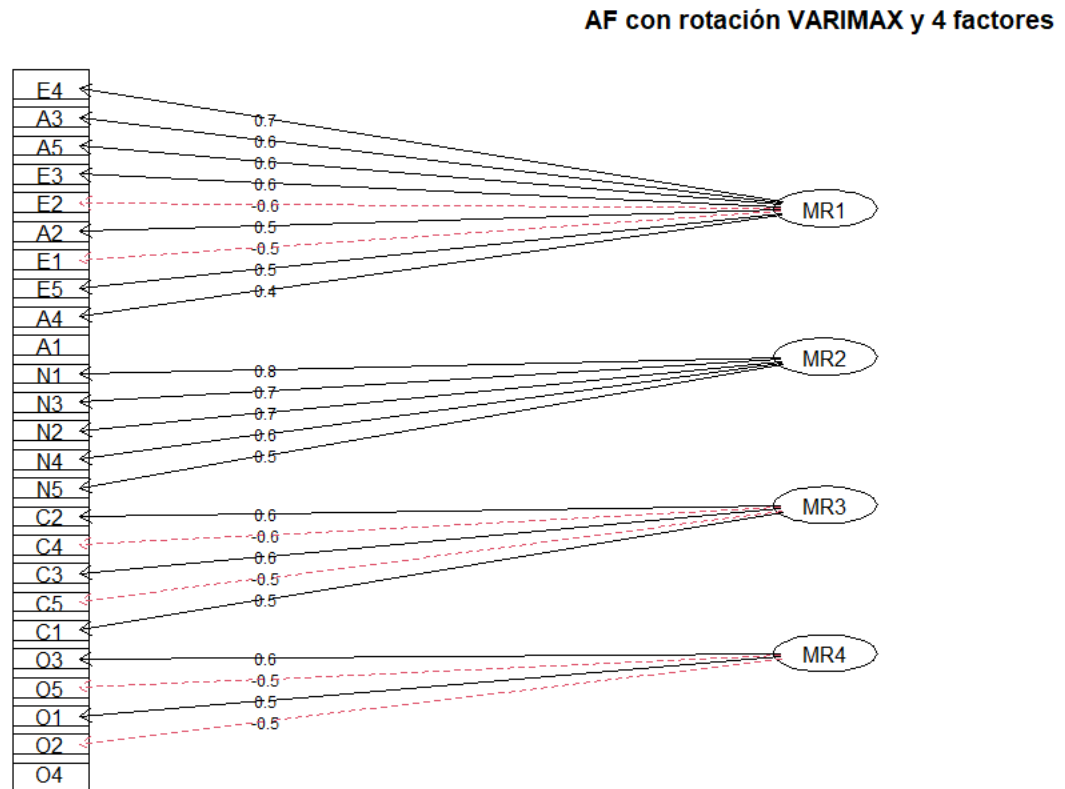


Figura 4.10: Árbol de análisis factorial con rotación VARIMAX y 4 factores

Es notorio que ahora con rotación VARIMAX las variables  $N_i$ ,  $C_i$  y  $N_i$  si ocupan factores por separado, sin embargo, aún las variables  $A_i$  y  $E_i$  juntas hacen un factor, además aún la A1 y O4 no son consideradas. En consiguiente, se realiza el gráfico del AF con rotación VARIMAX pero ahora con 5 factores:

### AF con rotación VARIMAX y 5 factores

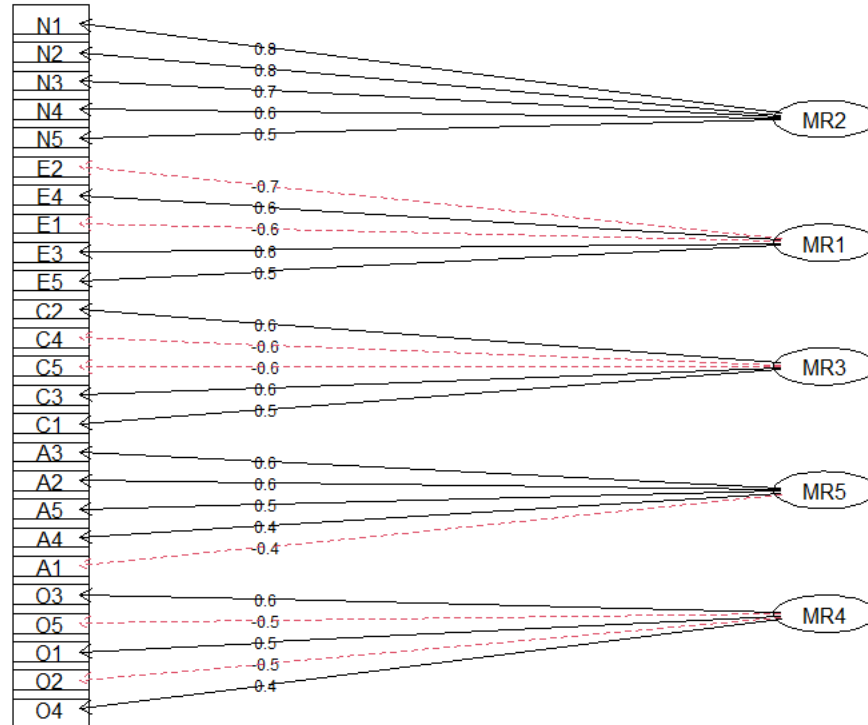


Figura 4.11: Árbol de análisis factorial con rotación VARIMAX y 5 factores

Ahora si se logra un modelo adecuado e interpretable, pues cada variable perteneciente a los rasgos del Big five hacen un factor. Aquí, se puede decir que MR2 por si solo explica el neuroticismo, el MR1 por si solo explica la extraversión, el MR3 la escrupulosidad, MR5 la amabilidad y el MR4 la apertura a la experiencia. Por tanto, el conjunto total de datos puede reducirse a solo 5 mediante análisis factorial y explicar cada una de los rasgos del Big five.

En adición a lo anterior, ahora se agregan las variables demográficas para ver que explicabilidad tienen con las 5 variables obtenidas mediante AF:

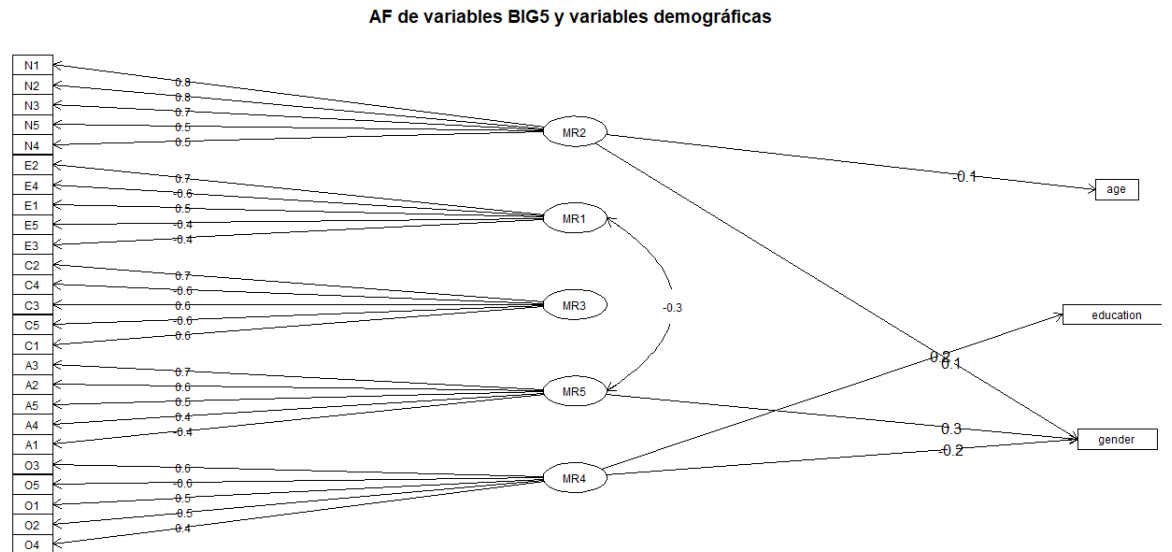


Figura 4.12: Árbol de análisis factorial con rotación VARIMAX y 5 factores incluídas variables demográficas.

Es notorio que la edad es explicada por MR2 que a su vez explica el neuroticismo, lo cual tiene sentido debido a que cuando uno va creciendo aprende a solucionar los problemas de mejor manera, en cuanto la educación es explicada por MR4 que es la apertura a la experiencia lo cual es bastante lógico, pues cuando uno tiene más estudios uno obtiene más experiencia y se tiene más conciencia sobre el futuro y por último el género está asociado a MR4 y MR2 los cuales son la apertura a la experiencia y el neuroticismo, lo cual es concordante con la investigación de Barrio & et al (2006) que llegaron a la conclusión que existe diferencia entre mujeres y hombres sobre la conciencia.

Para realizar un análisis más detallado sobre el último modelo visto, se presenta la matriz de loadings:

Variable	MR2	MR1	MR3	MR5	MR4
A1	0.205	-0.178	0.075	-0.425	-0.072
A2	-0.015	-0.008	0.072	0.638	0.020
A3	-0.023	-0.106	0.0332	0.667	0.027
A4	-0.055	-0.074	0.185	0.449	-0.167
A5	-0.121	-0.216	0.004	0.535	0.044
C1	0.078	0.0397	0.550	-0.004	0.159
C2	0.145	0.104	0.658	0.092	0.040
C3	0.026	0.053	0.575	0.084	-0.065
C4	0.153	0.007	-0.636	0.043	-0.037
C5	0.167	0.150	-0.562	0.026	0.0945
E1	-0.065	0.538	0.103	-0.113	-0.098
E2	0.086	0.671	-0.016	-0.065	-0.069
E3	0.091	-0.423	0.010	0.242	0.297
E4	-0.009	-0.572	0.019	0.318	-0.066
E5	0.163	-0.425	0.268	0.065	0.206
N1	0.832	-0.103	-0.002	-0.103	-0.047
N2	0.781	-0.037	0.013	-0.095	0.019
N3	0.700	0.115	-0.026	0.089	0.015
N4	0.464	0.411	-0.136	0.097	0.087
N5	0.477	0.206	-0.019	0.205	-0.161
O1	0.000	-0.091	0.079	0.004	0.515
O2	0.188	-0.057	-0.084	0.169	-0.480
O3	0.036	-0.159	0.003	0.073	0.623
O4	0.109	0.330	-0.045	0.161	0.365
O5	0.117	-0.099	-0.027	0.0381	-0.551
gender	0.128	-0.018	0.080	0.284	-0.172
education	-0.072	0.096	0.000	0.041	0.151
age	-0.126	0.055	0.067	0.093	0.054

Cuadro 4.2: Loadings del modelo VARIMAX con 5 factores

Es claro ver que los loadings en cada factor son significativamente más altos en las 5 variables del Big five que concuerdan en el gráfico, por ejemplo, en MR2 los loadings de las variables N1,N2,N3,N4 y N5 son mayores a 0.4 incluso unas son 0.7, a diferencia de las demás variables que todas son menores a 0.2.

En cuanto a la varianza explicada puntual y acumulada del modelo se obtiene:

<b>Proporciones</b>	<b>MR2</b>	<b>MR1</b>	<b>MR3</b>	<b>MR5</b>	<b>MR4</b>
Proporción varianza explicada	0.11	0.10	0.08	0.07	0.06
Proporción varianza acumulada	0.11	0.21	0.29	0.36	0.42

Cuadro 4.3: Varianza puntual y explicada de cada factor en AF con rotación varimax y 5 factores

Se observa que es bastante poco lo que estos 5 factores explican a la variabilidad total de la base de datos, lo cual es bastante lógico, pues como se ha mencionado las variables del Big five tienen por detrás muchísimas variables que intentan explicar, por tanto, el reducir la dimensionalidad de estas va a conllevar perder información de no solo la reducción de las variables presentes en la base de datos, sino que también de otras que no se tiene información.

## 4.4. Aplicación análisis de componentes principales

En consiguiente, debido al análisis de la investigación propuesta por Lovik & et al (2017) que se planteó en la formulación teórica, se realiza un análisis de componentes principales(ACP) para comparar los resultados obtenidos con el análisis factorial con rotación VARIMAX. La proporción de variabilidad acumulada explicada al realizar un ACP se presenta en el siguiente gráfico:

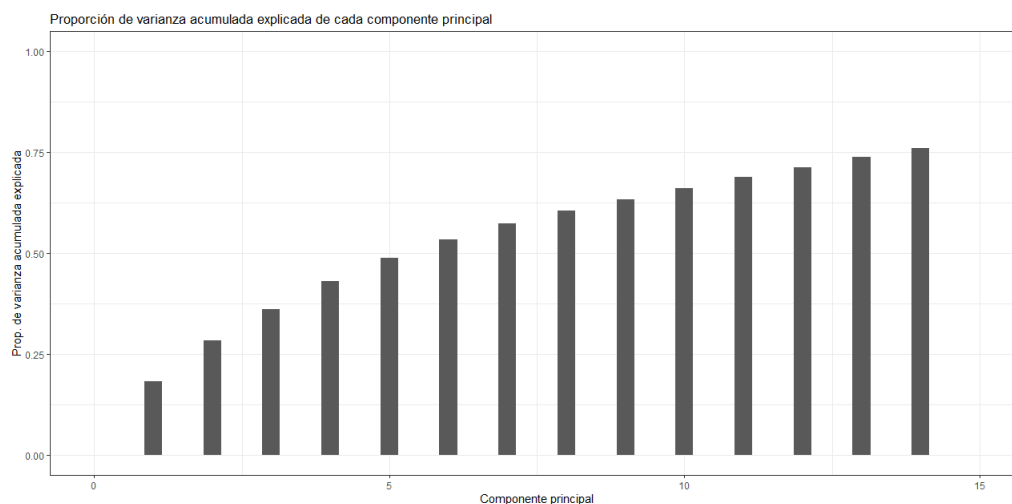


Figura 4.13: Proporción de variabilidad acumulada explicada por cada componente principal de ACP sobre todas las variables de BFI

Es notorio que si se decide tomar 5 componentes principales, se obtiene un poco más de variabilidad explicada acumulada que en análisis factorial (casi un 50 %) lo cual es bastante bueno. Ahora se presenta un gráfico para saber como se explican las variables mediante los componentes principales:

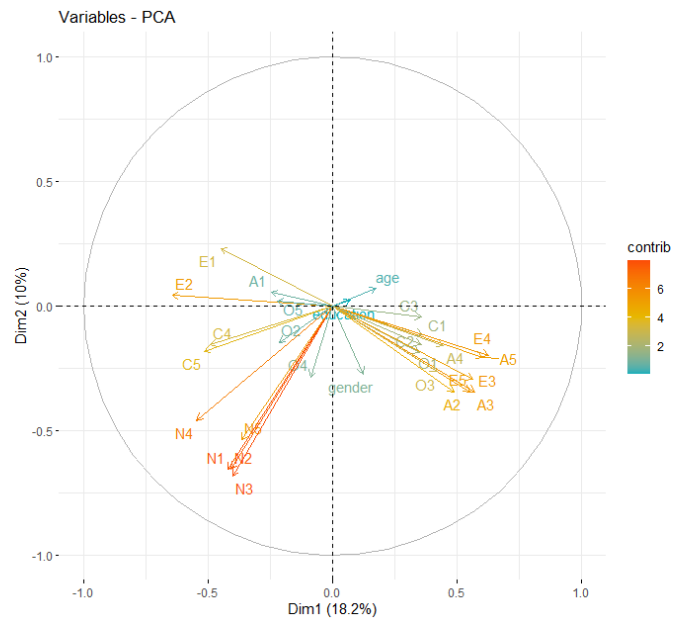


Figura 4.14: Gráfico de contribución de cada variable en ACP.

Es notorio que en cuanto a explicabilidad, se pierde bastante en ACP, pues se observa en la imagen que hay variables de diferentes grupos del big five que parecieran explicar lo mismo (ya que la dirección que tienen es parecida). Este es uno de las desventajas de trabajar con ACP, por ende, en este trabajo es preferible perder un poco de proporción de variabilidad acumulada explicada con los mismos factores y componentes principales, pero a la vez ganar una interpretabilidad mucho mayor.



## Capítulo 5

# Conclusiones

Debido a la poca información acerca de la base de datos y el cómo contestaron las preguntas las personas, es posible que al realizar la eliminación de las variables con valores perdidos se haya desaprovechado información valiosa de estos usuarios.

En consiguiente, los resultados obtenidos indican que de las 28 variables que tenía la base de datos, se puede reducir a solo 5 variables mediante análisis factorial, las cuales en conjunto representan un 42 % de la variabilidad total del set de observaciones, lo cual es bastante bueno, pues según la naturalidad de las variables es muy difícil lograr llegar a resumir en 5 variables y que a la vez tenga una proporción de información parecida a ellas, sin embargo, los factores obtenidos si representan cada rasgo de las variables del Big five, por tanto, se puede decir que la división entre factores de las variables ha sido la esperada.

Además, se logra identificar la ventaja que tiene el análisis factorial sobre el análisis de componentes principales, pues si bien el ACP obtiene una mayor proporción de variabilidad acumulada explicada con el mismo número de factores, en éste se pierde la interpretabilidad, por otro lado, en análisis factorial la interpretabilidad es perfecta, por ende, en este caso se sugiere considerar los resultados del análisis factorial por sobre en análisis de componentes principales.

## Capítulo 6

# Referencias

- Alderete, M. (2022). *Evaluación de la calidad de aguas superficiales en la cuenca del río Rímac mediante análisis multivariado para el período 2011-2018* [Tesis para optar al título de ingeniera ambiental]. Universidad Nacional Agraria La Molina. [[Ver aquí](#)]
- Barrio, M. & et al. (2006). Análisis transversal de los cinco factores de personalidad por sexo y edad en niños españoles. *Revista Latinoamericana de Psicología*, 38(3), 567-577. [[Ver aquí](#)]
- Depaula, P. & Azzollini, S. (2013). Análisis del modelo big five de la personalidad como predictor de la inteligencia cultural. *PSIENCIA. Revista latinoamericana de ciencia psicológica*, 5(1) 35-43 [[Ver aquí](#)]
- Dominguez, S. & Merino, Cesar. (2018). Dos versiones breves del Big Five Inventory en universitarios peruanos: BFI-15p y BFI-10p. *Revista Peruana de Psicología*, 24(1), 80-97. [[Ver aquí](#)]
- Goldberg, L. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychologs*, 59(6), 1216-1229. [[Ver aquí](#)]
- Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200. [[Ver aquí](#)]
- Kendall, M. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81-93. [[Ver aquí](#)]

- Lovik, A. & et al. (2017). Psychometric properties and comparison of different techniques for factor analysis on the Big Five Inventory from a Flemish sample. *Personality and individual differences*, 117, 122-129. [[Ver aquí](#)]
- Pearson, K. (1908). On lines and planes of closest fit to system of points. *Philosophical Magazine*, 2, 559-572. [[Ver aquí](#)]
- RDocumentation. (s.f). bfi: 25 Personality items representing 5 factors. [[Ver aquí](#)]
- Rencher, A. (2019). Methods of Multivariate Analysis. *2ed: John Wiley Sons, Inc.* [[Ver aquí](#)]