

Taller 3: Análisis Multivariado

Análisis de correspondencias

Julián Camilo Riaño Moreno

sábado, junio 20, 2020

Contents

| | |
|--|---|
| Actividad 3 | 1 |
| Especificaciones sobre la base de datos utilizada. | 1 |
| Análisis de tablas de contingencia y proporciones. | 3 |
| Prueba de independencia χ^2 | 4 |
| Análisis de correspondencia y gráficas | 5 |
| Biplots y contribuciones de las variables. | 6 |

Actividad 3

Elija dos variables categóricas de una base de datos relacionada con su área de conocimientos y desarrolle un análisis de correspondencias (AC), mostrando la tabla de contingencia, la tabla de frecuencias relativas, la gráficas de barras de los perfiles fila y columna, y la gráfica de los puntos de los perfiles fila y columna superpuestos en el primer plano factorial (biplot). Haga una interpretación práctica de los resultados. >Están relacionadas las dos variables seleccionadas o son independientes?

Especificaciones sobre la base de datos utilizada.

Para la actividad de análisis de correlaciones, se decidió utilizar parte de una base de datos resultado de un estudio multicéntrico realizado en 61 países donde participe como investigador principal del equipo de Colombia. El estudio tiene como título *COVID19 International Collaboration on Social & Moral Psychology* y tiene como objetivo, entender las características morales y sociales de las personas en el mundo durante la pandemia por COVID-19.

El instrumento aplicado es una encuesta que consta de 41 preguntas definidas por variables categóricas y cuantitativas y fue elaborada y difundida a través del sistema *qualtrics* (<https://www.qualtrics.com>). En Colombia se lograron obtener 735 encuestados entre los días 1-4 de abril del 2020.

A partir de esto se obtuvo una base de datos de 735 filas y 94 columnas. Para desarrollar esta actividad, se decidió seleccionar únicamente dos columnas de interés de manera conveniente para ajustarla para realizar el análisis solicitado. Las columnas seleccionadas fueron las que corresponden a: *inclinación política* y *círculo moral*.

Estas dos corresponden a variables categóricas que fueron asignadas de manera numérica para facilitar el análisis final. Para los fines de esta actividad se decidió hacer una readaptación de la variable *inclinación*

política la que originalmente los encuestado debían responder en una escala de 0-10 siendo 0 ultraizquierda y 10 ultraderecha; para facilitar el análisis se decidió realizar rangos y definir las inclinaciones de manera nominal como se muestra en la tabla 2; con lo que se obtuvo únicamente tres categorías de la variable: *izquierda, derecha y centro*.

La variable *círculo moral* se entiende como el círculo de personas u otras entidades por las cuales la persona se preocupa o procura el bien o el mal que se les pueda hacer. Para evaluar esto se asignaron 16 indicadores ¹ (Id) descritos en la tabla 1. a cada uno se les asignó un nombre corto para este ejercicio y mejor interpretación, que también se puede apreciar en la tabla 1.

Finalmente y luego de realizar una limpieza de la base de datos, se obtuvo 687 una base de datos final de 687 observaciones, la cual fue utilizada para esta actividad. Seguidamente, esta base de datos fue organizada para obtener una tabla de contingencia donde la *inclinación política* se ajustó como la variable columna y el *círculo_moral* se asignó como variable fila.

Table 1: Tabla de descripción de las variables filas

| Id | Variable | Descripción |
|----|---------------------|---|
| 1 | familiar_inmediato | toda su familia inmediata |
| 2 | familiar_extendido | toda su familia extendida |
| 3 | amigos_cercanos | todos sus amigos cercanos |
| 4 | amigos_inclu_lejano | todos sus amigos, incluyendo los distantes |
| 5 | todos_conocidos | todos sus conocidos |
| 6 | personas_cruzado | todas las personas con las que se ha cruzado |
| 7 | personas_pais | todas las personas de su país |
| 8 | personas_continente | todas las personas en su continente |
| 9 | personas_mundo | todas las personas en todos los continentes |
| 10 | mamiferos | todos los mamíferos |
| 11 | animal_cordado | todos los anfibios, reptiles, mamíferos, peces y aves |
| 12 | animal_bact | todos los animales en la tierra, incluyendo bacterias y amibas |
| 13 | animal_extrat | todos los animales en el universo, incluyendo formas de vida extraterrestres |
| 14 | todo_biotico | todas las cosas vivientes en el universo, incluyendo plantas y árboles |
| 15 | biotico_abiotico | todas las cosas naturales en el universo, incluyendo entidades inertes como las rocas |
| 16 | todo_existente | todas las cosas que existen |

Table 2: Tabla de descripción de columnas

| Variable | Descripción |
|-----------|--|
| derecha | Todos aquellos que marcaron entre 7 y 10 |
| izquierda | Todos aquellos que marcaron entre 0 y 3 |
| centro | Todos aquellos que marcaron entre 4 y 6 |

¹En esta escala el número que seleccione incluye todos los números bajo el también. Por ejemplo, si usted selecciona 10 (todos los mamíferos), está incluyendo, también los números 1 al 9 (“todas las personas en todos los continentes) en su círculo moral.

Analisis de tablas de contingencia y proporciones.

Inicialmente se realizó una inserción de los valores marginales a la tabla de contingencia que se trabajó (tabla 3). Esto muestra que la mayor cantidad de las observaciones se obtuvieron de población que se percibe asimismo en una inclinación política de **centro** ($n = 373$). El grupo de inclinación política menos representado fue **derecha** ($n = 115$).

Table 3: Tabla de contingencia y marginales

| | centro | derecha | izquierda | Sum |
|----------------------------|--------|---------|-----------|-----|
| amigos_cercanos | 15 | 5 | 12 | 32 |
| amigos_inclu_lejano | 17 | 3 | 10 | 30 |
| animal_bact | 11 | 2 | 6 | 19 |
| animal_cordado | 28 | 2 | 8 | 38 |
| animal_extrat | 3 | 3 | 2 | 8 |
| biotico_abiotico | 11 | 5 | 5 | 21 |
| familiar_extendido | 5 | 4 | 12 | 21 |
| familiar_inmediato | 47 | 21 | 18 | 86 |
| mamiferos | 16 | 3 | 7 | 26 |
| personas_continente | 4 | 1 | 2 | 7 |
| personas_cruzado | 9 | 1 | 4 | 14 |
| personas_mundo | 11 | 9 | 6 | 26 |
| personas_pais | 5 | 2 | 9 | 16 |
| todo_biotico | 48 | 12 | 28 | 88 |
| todo_existente | 112 | 33 | 62 | 207 |
| todos_conocidos | 31 | 9 | 8 | 48 |
| Sum | 373 | 115 | 199 | 687 |

Por otra parte, a través de un análisis de proporciones mostrado en la tabla 4. se identificó que el círculo moral más representado en los datos de estudio fue el que corresponde a la categoría **todo_existe** ($n = 207$, 31.16%), es decir, se puede asignar un estatus moral a cada cosa existente en el universo; el grupo que más contribuyó a esta categoría es **centro** ($n = 112$, 30.03%). La categoría de menor representación fue **animal_extrat** y **personas_continente** ($n = 8$ y 7 ; 1.01% y 1.01% respectivamente).

Table 4: Tabla de proporciones

| | centro | derecha | izquierda |
|----------------------------|--------|---------|-----------|
| amigos_cercanos | 4.02 | 4.35 | 6.03 |
| amigos_inclu_lejano | 4.56 | 2.61 | 5.03 |
| animal_bact | 2.95 | 1.74 | 3.02 |
| animal_cordado | 7.51 | 1.74 | 4.02 |
| animal_extrat | 0.8 | 2.61 | 1.01 |
| biotico_abiotico | 2.95 | 4.35 | 2.51 |
| familiar_extendido | 1.34 | 3.48 | 6.03 |
| familiar_inmediato | 12.6 | 18.26 | 9.05 |
| mamiferos | 4.29 | 2.61 | 3.52 |
| personas_continente | 1.07 | 0.87 | 1.01 |
| personas_cruzado | 2.41 | 0.87 | 2.01 |
| personas_mundo | 2.95 | 7.83 | 3.02 |
| personas_pais | 1.34 | 1.74 | 4.52 |
| todo_biotico | 12.87 | 10.43 | 14.07 |
| todo_existente | 30.03 | 28.7 | 31.16 |

| | centro | derecha | izquierda |
|------------------------|--------|---------|-----------|
| todos_conocidos | 8.31 | 7.83 | 4.02 |

Para evaluar mejor los datos de las tablas 3 y 4, se decidió realizar una gráfica de barras y esferas de los perfiles (figura 1.), en esta podemos ver que el radio de la esfera en azul representa la proporción o número de personas representadas en la intersección de dicha fila y columna. En esta imagen se ve claramente lo mencionado anteriormente respecto a la relación entre la categoría **centro** y **todo_existente**. Aca también se puede evidenciar en las barras grises que aparecen en cada uno de los nombres la proporción de los marginales, confirmando gráficamente lo antes descrito. Además, se puede apreciar mejor como **izquierda** también aporta grandemente a la categoría **todo_existe**.

Circulo moral por disposición política

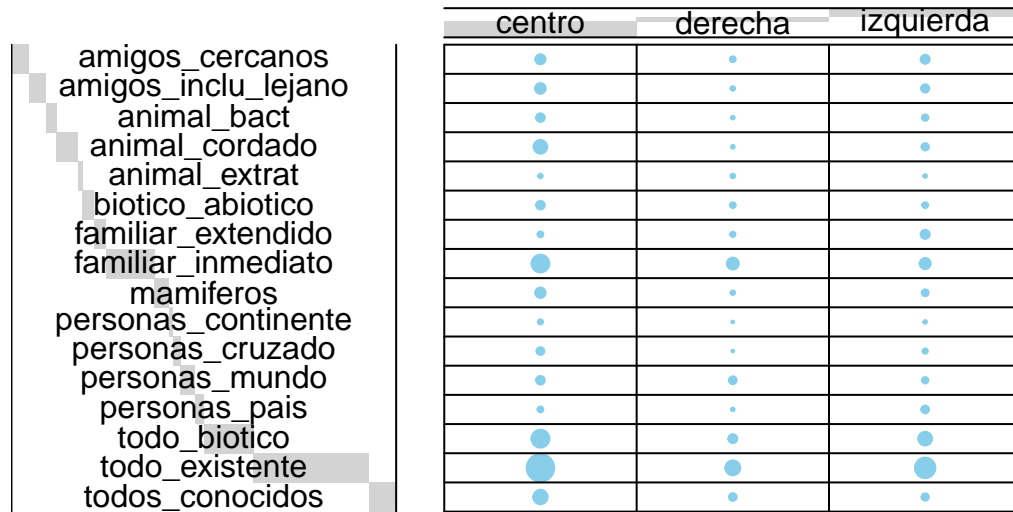


Figure 1: Gráfica de perfiles

Prueba de independencia χ^2

Para evaluar la independencia entre las variables estudiadas se decide realizar una prueba de χ^2 a través de método de Pearson, con la función `chisq.test`. Esta prueba corresponde a una prueba de hipótesis entendida de la siguiente manera:

H_0 : las variables son independientes

H_1 : las variables son dependientes

Table 5: Prueba de independencia por Ji^2

| ji^2 | $p - value$ |
|--------|-------------|
| 44.93 | 0.03917 |

El resultado de la ji^2 puede verse en la tabla 5. El valor del estadístico (44.93) fue obtenido con 30 grados de libertad con un $p - value = 0.03917$ lo que asumiendo un erro estandar de 5% para la prueba se puede concluir que no es posible rechazar la hipotesis nula (H_0), de manera que las variables son dependientes.

Análisis de corresponencia y gráficas

Se elabora un análisis de correspondencia a través de la función **CAdel** paquete **factoextra**. Se obtiene los valores propios (o *eigenvalues*) de las dimensiones mínimas obtenidas del análisis de correspondencia (2). En otras palabras, con tan solo dos dimensiones es posible explicar el 100 % de la varianza de los datos analizados, como se observa en la tabla 6. Del total de la varianza, la dimensión 1 explicar el 54.3 % de la varianza y la dimensión 2 el 45.7% de esta.

La figura 2 corresponde al Scree plot de los % de contribución de las variables en general para cada una de las obtenidas, mostrando que los perfiles entre variables explican en general casi de manera similar las dimensiones 1 y 2.

Table 6: Tabla de *eigenvalues* y dimensiones

| | eigenvalue | variance.percent | cumulative.variance.percent |
|--------------|------------|------------------|-----------------------------|
| Dim.1 | 0.03551 | 54.3 | 54.3 |
| Dim.2 | 0.02989 | 45.7 | 100 |

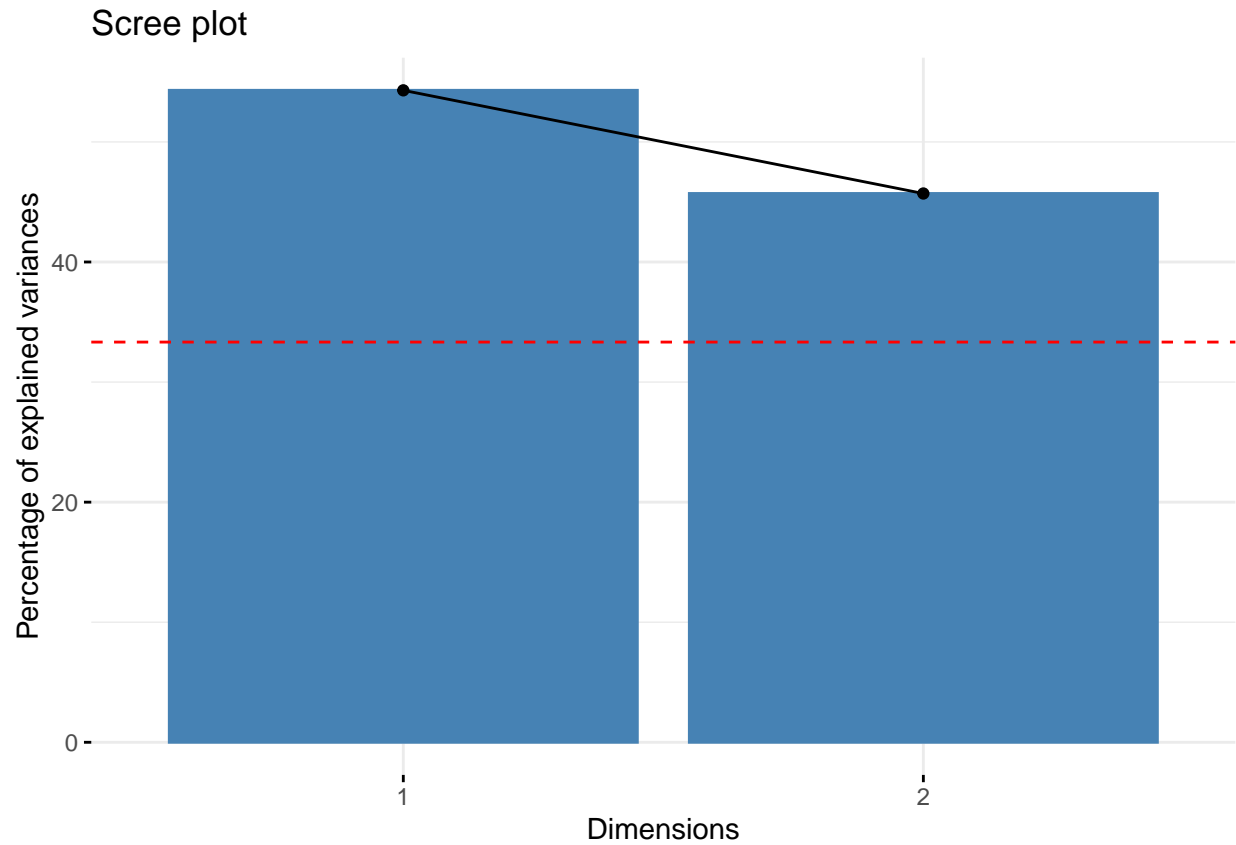


Figure 2: gráfica de sedimentación

Biplots y contribuciones de las variables.

Finalmente se realizaron dos tipos de Biplots, uno de estos simétrico (figura 3) y un biplot asimétrico (figura 4.). Para darle mayor sentido las gráficas se les asignó color en cuanto al peso de contribución para la vector fila como a la variable columna.

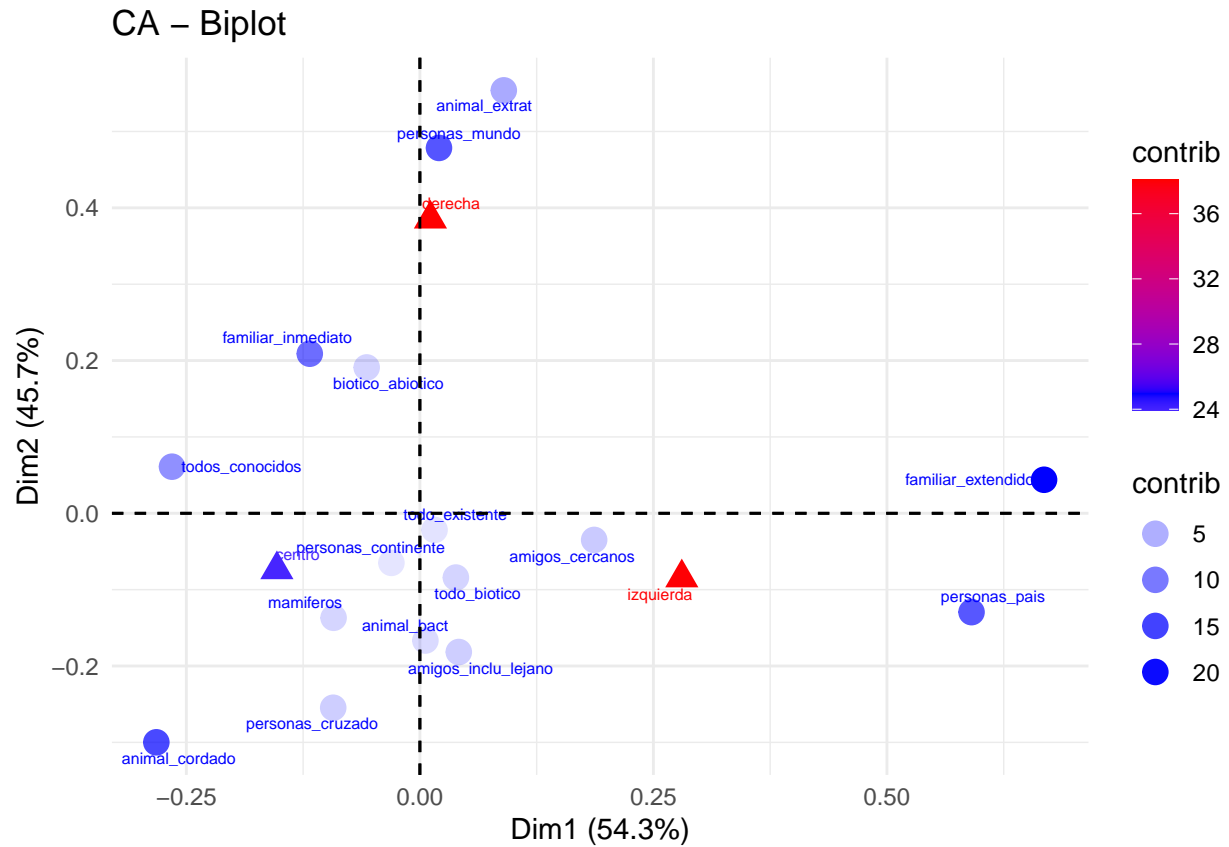


Figure 3: Biplot de las variables categóricas

En el caso del Biplot simétrico (figura 3.) se puede observar que las vector columna fue representada con colores que van de rojo para la máxima contribución de la variable a la dimensión y azul la menor contribución. En el caso de la vector fila, se asigno azul oscuro para la mayor contribución y azul más claro (o gris) para la menor contribución. En el caso del biplot asimétrico se le asignó tanto a la vector fila como al vector columna, color rojo a la máxima contribución y azul a la mínima contribución.

destacar que `derecha` y `persona_mundo` tienen un ángulo muy agudo lo que apunta a lo dicho anteriormente, la definición de un perfil de relación entre las dos variables a través de estas categorías; lo que es también perceptible en el caso de `izquierda` y `persona_pais`.