



Análisis de ENVIGMU: 2019

Antonio Alvarez

14.6.2020

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(readr)
library(readxl)
library(tibble)
library(stringr)
library(FactoMineR)
library(factoextra)
library(patchwork)
```

1. Introducción

La violencia de genero es un problema perverso que afecta a todos los miembros de la sociedad. Este se refiere a actos cometidos exclusivamente contra mujeres o niñas, los cuales son causados por diversos culturales y socioeconomicos. Para recoger información de estas posturas, el Instituto Nacional de Estadísticas y Censos (INEC) recogieron opiniones a nivel nacional a través de la Encuesta Nacional sobre Relaciones Familiares y Violencia de Genero contra las Mujeres (ENVIGMU) en el 2019.

La metodología, sus datos, y demás metainformacion la podemos encontrar en su pagina oficial. Debido a que las mujeres pueden experimentar actos violentos de diversas formas y en diferentes escenarios, se pueden identificar grupos de variables que siguen estas membresias. En nuestro análisis, nos enfocaremos en los escenarios en donde se da esta violencia: escolar, familiar, social, laboral y en pareja.

Para investigar la estructura de las información a través de estos grupos, usaremos una metodología ya discutida en clases: análisis factorial múltiple (MFA). Además de descomponer toda mi observación en factores principales, la información de cada grupo se la resumirá y sus diferencias en este plano se las dividirá. Es crucial notar la presencia de pesos en el análisis de encuesta y su inclusión en los métodos para el análisis tienen que ser de principal consideración para el investigador

2. Configuración del modelo

```
vio_df <- read_delim("data/violencia.csv", ";")[-1]

# Transformacion de pesos en numericos
pesos <- vio_df[,72] %>%
  mutate(pesos=as.numeric(str_replace(pesos, ",", "."))) %>%
  pull(pesos)

vio_mfa <- MFA(vio_df[-72], group = c(15, 18, 14, 22, 1, 1),
  type = c(rep("n", 4), "s", "n"),
  name.group = c("educacion", "laboral",
    "familiar", "social", "edad", "civil"),
  num.group.sup = c(5, 6), graph = FALSE, row.w = pesos)
```

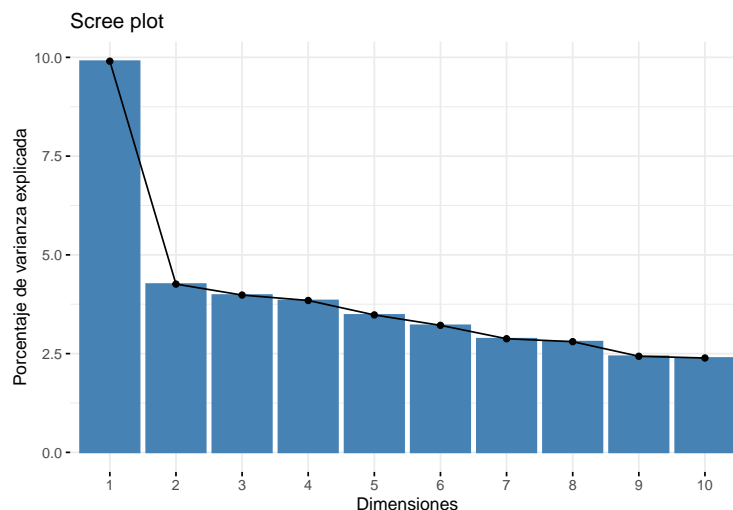
3. Resultados

3.1 Autovalores

```
get_eig(vio_mfa)[1:10,]
```

##	eigenvalue	variance.percent	cumulative.variance.percent
## Dim.1	1.9041869	9.903087	9.903087
## Dim.2	0.8192443	4.260636	14.163723
## Dim.3	0.7658169	3.982777	18.146499
## Dim.4	0.7394681	3.845744	21.992244
## Dim.5	0.6690743	3.479648	25.471892
## Dim.6	0.6184195	3.216209	28.688100
## Dim.7	0.5531636	2.876832	31.564933
## Dim.8	0.5389837	2.803088	34.368020
## Dim.9	0.4678780	2.433288	36.801309
## Dim.10	0.4594150	2.389275	39.190584

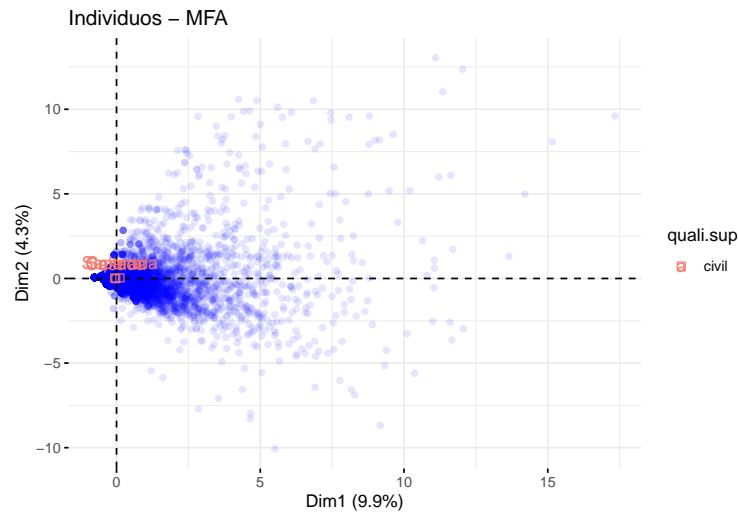
```
fviz_screepLOT(vio_mfa) + labs(x="Dimensiones", y="Porcentaje de varianza explicada")
```



Los 10 primeros autovalores no corresponden ni al 50% de la varianza acumulada, lo cual indica una alta dimensionalidad en la estructura de los datos. Esto no es inesperado, es común en análisis de variables categóricas donde las variables están muy dispersas. En el curso, tuvimos un ejemplo durante la explicación MCA donde los componentes igual no capturaban mucha varianza.

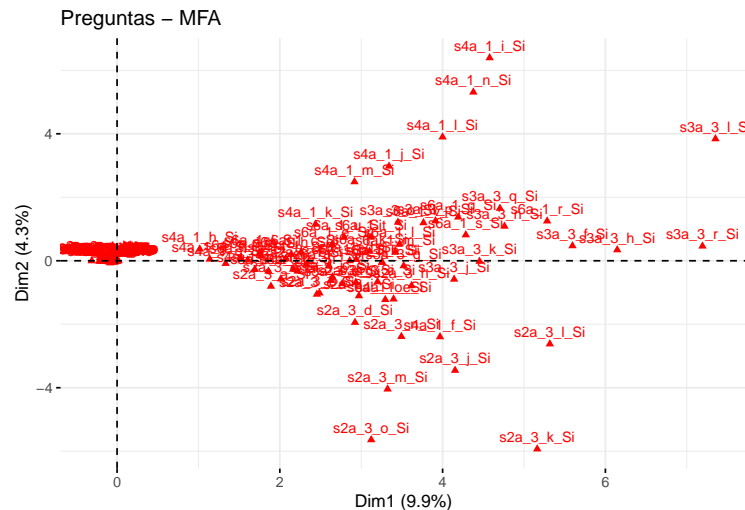
3.2 Nube de individuos y preguntas

```
fviz_mfa_ind(vio_mfa, geom = "point", alpha.ind = 0.1) +
  labs(title = "Individuos - MFA")
```



En la dirección de la primera dimensión, se ve una importante desviación de individuos. Mientras los puntos mas se alejan del origen, se dirigen a diferentes partes del espacio dimensional.

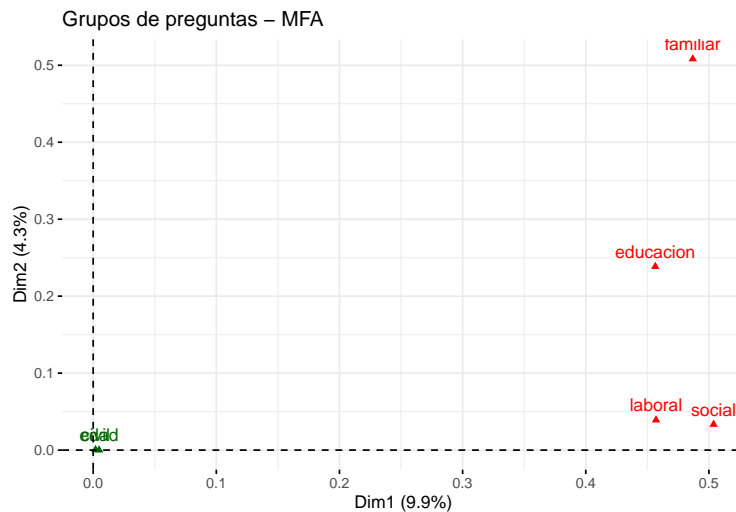
```
fviz_mfa_var(vio_mfa, geom = c("point", "text"),
  choice = "quali.var", labels=3) +
  labs(title = "Preguntas - MFA")
```



Como los individuos, las variables también siguen una dispersión importante en la dirección del primer componente. Podemos distinguir 3 direcciones importantes desde esta representación. Una diagonal superior con categorías como `s4a_1_i_Si` y `s4a_1_n_Si`, una dirección paralela al primer componente, y a otra diagonal inferior con categorías como `s2a_3_m_Si` y `s2a_3_k_Si`.

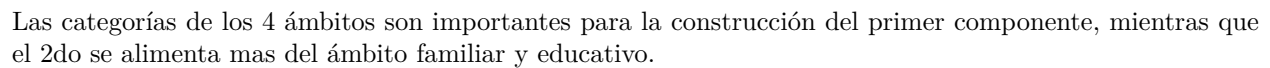
En este gráfico no se aprecia adecuadamente, pero las categorías con el valor No se encuentran muy cercanas unas o otra en el origen, mientras las que se responde Si están alejadas en el primer componente. Recomiendo que recreen este gráfico agregando el parámetro `repel = TRUE` al código y verificar esta forma.

```
fviz_mfa_var(vio_mfa, "group") +  
  labs(title="Grupos de preguntas - MFA")
```



Cuando vemos las grupos, dentro de las componentes de mayor varianza, los grupos suplementarios (edad y estado civil) no están correlacionados con ellos y no pueden ser interpretados utilizando estos componentes. En contraste, los otros grupos se encuentran correlacionados por lo menos con un componente. En particular, podemos ver tambien que el grupo de variables correspondientes al grupo laboral están muy cerca a las del ámbito social, lo cual significa que ambos grupos son bien parecidos y esperaríamos que sus categorías sean similares.

```
p1 <- fviz_contrib(vio_mfa, "group", axes = 1) +  
  labs(title="Contrib. de los grupos a la 1era Dim", y="Contribución (%)")  
  
p2 <- fviz_contrib(vio_mfa, "group", axes = 2) +  
  labs(title="Contrib. de los grupos a la 2da Dim", y="Contribución (%)")  
  
p1 + p2
```



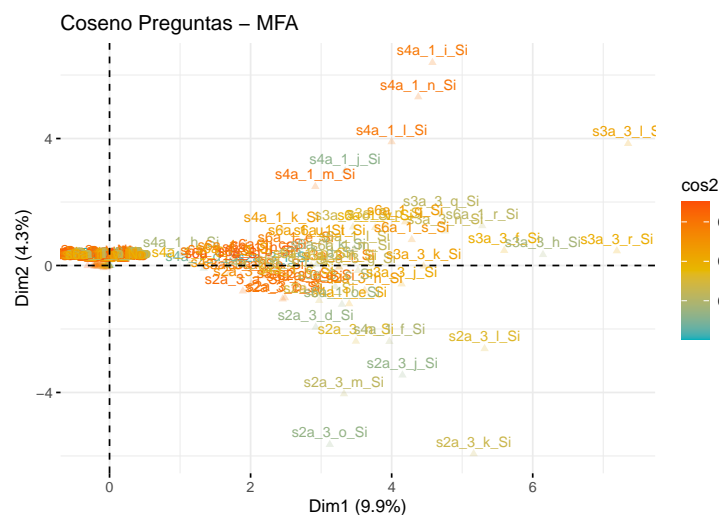
```
fviz_mfa_var(vio_mfa, "quali.var", col.var = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             col.var.sup = "violet",
             geom = c("point", "text"), labels=3) +
labs(title="Contribución Preguntas - MFA")
```



Vemos una alta contribución para la categoría Si de las preguntas s4a_1_i, s4a_1_n, s4a_1_l y s4a_1_m. Usando la información de la base de etiquetas (`etiquetas.xlsx`), encontramos que estas preguntas se refieren al forzoso uso de la fuerza para tener relaciones sexuales. Con eso, podríamos decir que esta dirección se refiere mujeres que han sido violadas en un ambiente social.

Debido también a que su contribución es alta para los 2 primeros componentes, podríamos decir que estas 2 dimensiones dividen a las mujeres que han sufrido un nivel mínimo de violencia con otras que han sido víctimas de niveles más altos de violencia.

```
fviz_mfa_var(vio_mfa, "quali.var", col.var = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             col.var.sup = "violet", alpha=0.2,
             geom = c("point", "text"), labels=3) +
labs(title="Coseno Preguntas - MFA")
```



La calidad de la representación también nos ayuda a informarnos de la lectura que hicimos de los componentes anteriormente. Para encontrar la dirección de otras categorías sería adecuado estudiar la estructura de estas nubes en otros componentes.

4 Limitaciones

Una limitación importante es la baja cantidad de varianza que estamos explicando en los primeros componentes. A pesar que estos 2 son los de máxima varianza, esperaríamos que otros componentes también nos informen de más información sobre la estructura de las respuestas.

5 Sigüientes pasos

El análisis presentado no es para nada un análisis completo de los resultados de esta encuesta. Más preguntas podríamos hacer sobre las similitudes de los grupos. ¿Qué preguntas en particular son las que los ámbitos suelen tener una distribución en común? En cuanto a los individuos, podríamos seguir investigando su nube y analizar candidatos en diferentes direcciones para informarnos de los componentes y la estructura general de los datos. Sería adecuado, también, agregar la información del quinto ámbito sobre la violencia en Pareja y observar cuál es el grupo que más se acerca o si los primeros componentes que hemos creado también están correlacionados con este nuevo ámbito.