**ESTADÍSTICA PARA EL ANÁLISIS POLÍTICO 2**

**HORARIO 690-B | Docente: Christian Chiroque Ruiz**

**ENTREGA 5 - ANÁLISIS FACTORIAL EXPLORATORIO - GRUPO 2**

**Información general:**

**Tema**: Desigualdad de género

**Objetivo**: Determinar los factores que explican e influyen en la desigualdad de género a nivel mundial.

**Link del repositorio:** <https://github.com/DanaeNorah/Estadistica-para-el-Analisis-Politico-2-/tree/master/TRABAJO/Hito%204>

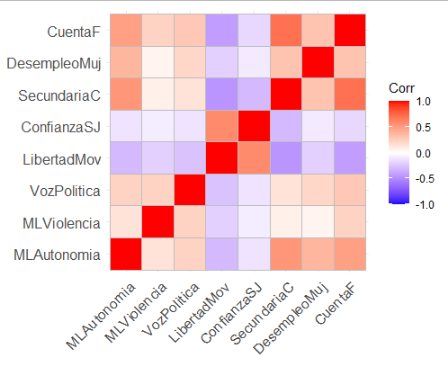
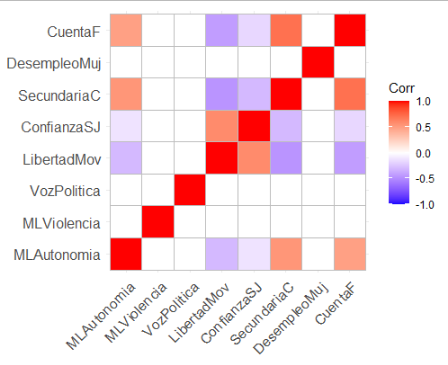
**Análisis con técnica multivariada descriptiva**

Con la finalidad de poder procesar con mayor facilidad las variables que explican la brecha de la desigualdad de género y contrastar con lo que nos dice la literatura, se realiza el análisis factorial exploratorio para crear un resumen que permita explicar las ideas principales del trabajo, es decir, encontrar el número mínimo de dimensiones o factores para explicar el máximo de la desigualdad de género con las variable observables elegidas. Asimismo, con estas variables latentes se puede redimensionar la data para hacer un mejor análisis por dimensiones empleando la totalidad del marco literario seleccionado.

Primero, se tiene que preparar la data para poder realizar el análisis factorial exploratorio. Por ello, se ha tenido que invertir la dirección de las variables que miden el marco legal de autonomía reproductiva, el marco legal de violencia, la libertad de movimiento, el nivel de desconfianza en el sistema judicial y la proporción de mujeres a hombres que están desempleadas, ya que lo ideal es que tengan un mismo sentido. Asimismo, fue necesario normalizar esta última y mantener los valores de 0 a 100, como con las demás variables.

Segundo, se realizó la matriz de correlación reportada en la Figura 1, en la cual se puede observar que hay un bloque de correlación entre las variables libertad de movimiento y confianza en el sistema judicial, lo cual hace referencia a la variable latente libertades civiles. Asimismo, se puede observar una correlación entre mujeres con secundaria completa, desempleadas y con cuenta financiera, que hacen referencia a la variable latente de educación y economía, las cuales están muy correlacionadas porque a un mejor nivel educativo mayor probabilidad de tener recursos económicos.

**Figura 1: Matriz de correlación sin significancia Figura 2: Matriz de correlación con significancia**

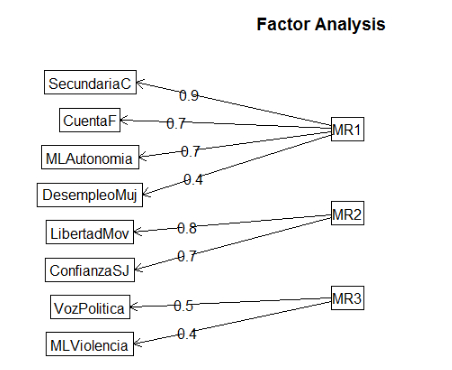
 

A continuación, se debe hacer un diagnóstico de la matriz de correlaciones. En primer lugar, es necesario verificar si los datos permiten hacer la factorización. Para ello, se empleará el test de KMO. Se tiene que revisar el valor de Overall MSA. Cuanto más cerca se encuentre este número a 1 implica que hay una relación alta entre variables. En este caso se observa que es un valor de 0.73, lo cual indica que el análisis factorial exploratorio es apropiado.

Luego, corresponde verificar si la matriz de correlaciones es adecuada, es decir, si hay relación entre variables. Por un lado, se aplica el test de Bartlett, que permite descartar si las variables solo se relacionan consigo mismas. Al obtener FALSO como resultado es un indicador de que el p-value es menor a 0.05, por lo tanto, se rechaza la hipótesis nula y la relación es significativa. Por otro lado, se emplea el test Singular Square Matrix, que permite descartar si la matriz de correlación es una matriz singular. El resultado es FALSO. Entonces se rechaza la hipótesis nula.

Después, el siguiente paso, consiste en determinar en cuántos factores o variables latentes se puede redimensionar la data. De acuerdo con lo sugerido, se han formado tres factores. Al visualizar el resultado se observa que la variable que mide el marco legal con respecto a la autonomía reproductiva (MLAutonomía) contribuye a los factores 1 y 3, con 0.663 y 0286 respectivamente. La variable que mide el marco legal con respecto a la violencia (MLViolencia) contribuye a dos factores, teniendo más peso en el factor 3 (0.427). La variable que se refiere a la voz política (VozPolitica) contribuye a los factores 1 y 3, con 0.167 y 0.531 respectivamente. La variable Libertad de movimiento (LibertadMov) contribuye a los factores 1, 2 y 3, con -0.268, 0.767 y -0.294 respectivamente.

La variable que mide la desconfianza de las mujeres en el sistema de justicia (ConfianzaSJ) contribuye sólo al factor 2 con 0.713. La variable que mide el porcentaje de mujeres de 25 años a más que ha llegado al nivel de educación secundaria (SecundariaC) contribuye a los factores 1 y 2 con 0.912 y -0.326 respectivamente. La variable que mide la proporción de mujeres a hombres que están desempleadas (DesempleoMuj) contribuye a los factores 1 y 3, con 0.386 y 0.237 respectivamente. La variable que mide el acceso a servicios financieros (CuentaF) contribuye a los factores 1, 2 y 3, con 0.708, -0.171 y 0.301 respectivamente. Finalmente, con respecto a la varianza acumulada, se obtiene 0.510, lo cual no es tan alto como se esperaría. Incluso antes de mejorar el resultado, se puede observar en general que gran parte de las variables no son muy significativas, pocas de las que se observan son mayores a |0.7|.

Al momento de mejorar el resultado con un corte en 0.5, se puede observar dos aspectos importantes. Primero, la proporción de mujeres a hombres que están desempleadas y el marco legal sobre la violencia doméstica no contribuyen significativamente a ningún factor creado. Segundo, solo el factor de educación contribuye con un valor de más de 0.8 a un factor. Entonces, si se observa los resultados de manera gráfica, las variables que contribuyen al factor 1 son SecundariaC, CuentaF, MLAutonomía (de manera inversa) y DesempleoMuj.

Esta combinación de variables no es muy extraña, primero porque la variable secundaria completa y la variable desempleo están intrínsecamente relacionadas a nivel teórico, las mujeres con mejores grados académicos tienden a ser menos desempleadas. Asimismo, contar con una cuenta financiera y con un marco legal para la autonomía reproductiva están relacionadas con las 2 variables previas, porque hablan de una independencia de los hombres para realizar acciones básicas de la vida cotidiana, lo cual se logra a través de la educación y del empleo. Estas variables en conjunto se pueden considerar dentro de un factor denominado Independencia.

Al segundo factor contribuye la variable libertad de movimiento y la variable desconfianza en el sistema de justicia. Con respecto a esta combinación de variables, según la literatura revisada se sustenta que pertenezcan a una misma latente denominada Libertades Civiles, así, ambas se relacionan en el sentido de que si hay desconfianza en el sistema judicial, la libertad de movimiento no va ser tan alta, pues, si a una mujer la acosan en la calle y la justicia no responde ante ese tipo de situaciones, tanto la libertad de movimiento como la confianza en el sistema de justicia van a ser bajos. Este factor puede continuar llamándose libertades civiles. Al tercer factor contribuyen la variable voz política y el marco legal sobre violencia. Esta combinación de variables pueden estar relacionadas en el sentido de que la cantidad de mujeres en el parlamento pueden contribuir a que exista un mejor marco legal que proteja a las mujeres violentadas. Este tercer factor puede considerarse como Visibilización.

Entonces, es necesario evaluar el Análisis Factorial exploratorio que se ha solicitado para comprobar la validez del del EFA calculado. Para ello, esta debe de cumplir con determinadas condiciones. En primer lugar, nuestro análisis debe de cumplir con una raíz de error cuadrático corregido cercano a cero; no obstante, en el caso que la cifra obtenida sea mayor al valor que se espera, es necesario observar. En este caso, no es necesario reportar nuestra raíz pues el resultado es 0.057. El segundo requisito a cumplir es que la raíz del error cuadrático medio de aproximación (RMSEA) sea menor a 0.5 para que sea considerado como óptimo. En este caso, el índice RMSEA obtenido es 0.0501; el cual es mayor a 0.5, por ende, es necesario reportarlo ante cualquier eventualidad. El siguiente requisito es que el índice de Tucker- Lewis debe de ser mayor a 0.9. En nuestro caso, es 0.9651; es decir, cumple con la condición.

Asimismo, como parte de la evaluación del análisis factorial realizado, las variables que aportaron más a los factores son "Secundaria Completa", "Libertad de Movimiento" y "Cuenta Financiera". En cambio, las variables que aportan menos a los 3 factores son "Violencia", "Desempleo femenino" y "Voz política". Por último, a grandes rasgos, todas nuestras variables cuentan con presencia de decimales, es decir cada una de ellas aportan a más de un factor; sin embargo, las que presentan un valor de "complexity" más elevado son "Desempleo de Mujeres", "Libertad de Movimiento" y "Cuenta Financiera".

Empero, ante todo ello, es necesario reportar una particularidad que ha llamado la atención durante esta evaluación. En el momento en que realizamos el paso anterior, la visualización del EFA solicitado, en la tabla requerida para analizar el resultado inicial, se puede observar que la variable “desempleo en las mujeres” aporta a más de un factor, en este caso al 1 y 3; sin embargo, este, en ninguno de los casos, es significativo. Asimismo, respecto al "resultado mejorado" —con la cual logramos que cada variable se vaya a un factor— es visible que la variable no aporta a ninguno de nuestros tres factores. Además, esta misma variable es la segunda que menos aporta a los factores. No obstante, a pesar de los resultados de esas tres situaciones, en la prueba "complexity", obtuvimos que esta variable Proporción desempleo de mujeres sobre hombres es la que contribuye a más de un factor (1.670).

Finalmente, cabe resaltar que la variable proporción de mujeres a hombres que están desempleadas altera significativamente los resultados. Es decir, al realizar el análisis factorial exploratorio sin la variable desempleo, se observa que si permite factorizar y también sugiere 3 factores o variables latentes. Asimismo, en el diagnóstico del EFA, se observa que la raíz del error cuadrático medio de aproximación es menor a 0.05 al ser 0. El índice Tucker-Lewis indica que hay suficiencia de validez de este trabajo al ser 1.076. Finalmente, al realizar la regresión lineal multivariada de los factores y del índice de desigualdad, este modelo explicaría la variabilidad del índice de desigualdad de género en un 82.14%, que es un porcentaje mayor al primer modelo en el que se usa la variable Desempleo (80.48%).