**Materiales y Métodos.**

**Reducción de dimensiones.**

Con el inicio de la pandemia surgen distintas formas de analizar el impacto en el que esta afecta a un determinado sector. Estas variables son numerosas para distintos ámbitos en los que se quiera enfocar, es decir, si nuestro estudio posee p-variables explicativas entonces estas serán de dimensión p. Reducir las dimensiones será fundamental para nuestro análisis, llevar a una expresión condensada de estas que explique su interrelación nos ayudará a concluir los objetivos de este estudio. Tomar que variables son las que más aportan dentro de un estudio ha sido tema de investigación durante mucho tiempo dentro del campo estadístico (Ker-ChauL I,Y ve ARAGON, Kerby SHEDDEN, and C. Thomas AGNAN, 2003). Existen dos distintos enfoques para esto, uno que queda en manos del investigador el cual de todas las covariables disponibles toma un grupo de estas según su criterio, estas siendo las que el cree son las menos redundantes y las más significativas al momento de explicar el fenómeno que estudia. El segundo enfoque, el cual es el que desarrollaremos más de fondo en este estudio, será el de reducción de dimensiones suficiente. Este enfoque producirá una combinación lineal del grupo de covariables tratadas. De lo anterior, se pueden producir varias combinaciones lineares de un grupo de covariables, la meta es decir que combinación o combinaciones son las que mejor explican el efecto que se estudia (Yanyuan Ma and Liping Zhu, 2014). Para el análisis de correspondencias, este paso es crucial y permite hacerse con un costo de perdida de información mínima, nuestro objetivo será el restringir esta mínima perdida de información para que este explique el máximo de información.

**Ordenamiento en espacio reducido.**

Al reducir la dimensión de los datos es necesario que estos estén ordenados alrededor de una serie de ejes ortogonales reducida, estos de manera descendiente siendo el primero el que más información provee de esta estructura. Esto se hace por medio de la extracción de valores propios asociado a la matriz de datos de nuestras variables. Si esta matriz de datos es de nxp, n observaciones en p variables estudiadas, estas n observaciones pueden ser representadas por grupos en un espacio p dimensional. Este grupo de datos usualmente se acumula más en ciertas direcciones y es plano en otras dentro de nuestro espacio p dimensional y no necesariamente este tomara el curso de alguna de las variables. Donde más esta alargado esta nube de puntos es donde reside la mayor varianza de este grupo, será el mayor gradiente que representan nuestros datos. Este será nuestro primer eje que se extraerá y es el que más información dota nuestro análisis, el siguiente eje será el segundo más impórtate solo si este es ortogonal al primero. El numeró de ejes que interpretan la estructura de los datos se determina por el método a tomar que en nuestro caso ser el de análisis de correspondencia.

**Análisis de Correspondencia.**

El análisis de correspondencia (AC) es una técnica estadística la cual es útil cuando dentro de nuestra matriz de datos se tienen filas y columnas de naturaleza categórica. En particular, se usa cuando se tiene tablas de contingencia la cual contiene frecuencias numéricas de los perfiles, produciendo un análisis gráfico más simple y formal, haciendo que su interpretación sea eficiente (Weller, 2005). Esta técnica buscara la mejor representación de las variables de la matriz en un plano de baja dimensión, los cuales llamaremos ejes que de manera ordinal se organizaran siendo el primero el que mejor explique la asociación entre los perfiles filas y columnas. Los otros factores trataran explicar la mayor parte del residuo que no explico su inmediato anterior, organizándose de manera descendiente (Cuadras, 2020). Este método buscara encontrar estas dimensiones tal que la variabilidad geométrica de la nube de puntos (también llamada inercia) sea máxima.

Esencialmente, el análisis de correspondencia tiene 3 fases: La transformación inicial de la matriz de datos, la descomposición en valores singulares de la transformación y el cambio de escala de los vectores proprios resultantes. Del primer paso, la transformación de la matriz se lleva acabo de la siguiente forma:

La matriz H tiene los valores transformados de la matriz X. Las matrices son matrices diagonales que contienen la raíz cuadrada de las marginales totales reciprocas tanto de las dimensiones de la fila y columna. Esta transformación remueve los efectos de las magnitudes que se formas de las diferencias entre marginales totales. En el caso de tablas de contingencia lo anterior seria equivalente a remover los valores esperados del estadístico Chi-cuadrado del en el modelo de independencias (Weller, 2005).

La segunda fase constaría en obtener los vectores característicos de nuestra matriz, mediante descomposición de valores singulares.

Donde:

es la matriz de la dimensión fila

es la matriz de la dimensión columna

es la matriz diagonal de valores propios.

Las matrices U, V y d se pueden encontrar directamente con una descomposición de valores singulares en la matriz de datos X o indirectamente mediante la extracción de los valores propios de las matrices de productos cruzados de X. Como última fase, las columnas tanto de las matrices U y V deben ser reescaladas para obtener los valores óptimos, también llamados valores canónicos. El resultado de esto da un ordenamiento donde se preservará la distancia entre los perfiles. Gráficamente el resultado de un análisis de correspondencias puede ser descrito por medio de un “Biplot” el cual grafica puntos del perfil fila xi y puntos del perfil columna yj tal que los productos escalares entre los vectores de fila y columna se aproximan a los elementos correspondientes de los datos de la matriz lo más cerca posible en un espacio reducido.

**Análisis de Correspondencia Múltiple.**

El análisis de Correspondencia (ACM) visto en el punto anterior es el caso mas sencillo cuando se tiene dos variables categóricas. El análisis de correspondencia múltiple involucra 3 o más variables dentro del conjunto de variables a estudiar. El análisis de correspondencia múltiple equivale a realizar un análisis de correspondencia simple, pero a la matriz indicadora de nuestros datos. En estas matrices indicadoras cada fila representa un caso y las columnas representan todas las categorías de las variables. Para el caso de variables de naturaleza continua estas pueden ser redefinidas por categorías, con una variable por categoría. Empleando un análisis por medio del producto cruz o matriz de Burt de las variables indicadoras se puede llegar a la solución óptima.

**Análisis de Correspondencia Canónica.**

Como hemos visto, el análisis de correspondencia nos ayuda a visualizar los datos de una matriz dentro de un espacio de dimensión reducida, donde la inercia de esta nube de puntos será la más optima. Un análisis de correspondencia canónico (ACC) se puede emplear cuando encontramos estas dimensiones a partir del AC, pero con la condición de que estas son una combinación lineal de variables explicativas adicionales. Podemos ver esto como un problema de regresión lineal múltiple, solo que en vez de hacer regresión sobre las variables explicativas lo haremos a las dimensiones de estas variables explicativas. El ACC restringe el espacio donde este buscara los ejes principales óptimos de un espacio total, el complemento de este será un espacio no restringido el cual podría también tener la solución optima. El porcentaje total de inercia estará repartido entre estos dos espacios, siendo el restringido el lugar donde se encontraría la inercia máxima. En consecuencia, a lo anterior el ACC explicara un poco menos del total de inercia existente. Para interpretar el modelo de manera gráfica se sigue la de un AC clásico y se sigue la misma explicación que la del “Biplot”.

# Bibliografía

Cuadras, C. M. (2020). *Nuevos Metodos de Analisis Multivariante.* Barcelona, Spain: CMC Editions.

Daniel Borcard.François Gillet Pierre Legendre. (2017). *Numerical Ecology with R.* Montréal: Département de sciences biologiques.

Greenacre, M. (2009). Canonical correspondence analysis ins ocial science research. *Department of Economics and Business*, 25-27.

Ker-ChauL I,Y ve ARAGON, Kerby SHEDDEN, and C. Thomas AGNAN. (2003). Dimension Reduction for Multivariate Response Data. *Journal of the American Statistical Association*, 99-109.

Weller, S. C. (2005). Correspondence Analysis. *Encyclopedia of Biostatistics*, 5.

Yanyuan Ma and Liping Zhu. (2014). A Review on Dimension Reduction. *NIH Public Access*, 18.