**ANÁLISIS DE VARIANZA MULTIVARIANTE. (MANOVA)**

El análisis de varianza multivariante es una generalización a p > 1 variables del análisis de varianza (ANOVA). (Cuadras, 2014)

Es decir, ANOVA prueba la diferencia de medias entre dos o más grupos, mientras que MANOVA prueba la diferencia en dos o más vectores de medios. (French, Macedo, Poulsen, & Yu, 2008)

Nótese que a las medias se les permite ser distintas, pero las varianzas se suponen todas iguales. Por suponerse las varianzas iguales diremos que el modelo es homocedástico (Sellero., 2008)

El objetivo principal al usar MANOVA es determinar si las variables de respuesta se modifican por la manipulación del observador de las variables independientes. Por lo tanto, hay varios tipos de preguntas de investigación que pueden responderse utilizando MANOVA: (French, Macedo, Poulsen, & Yu, 2008)

1) ¿Cuáles son los principales efectos de las variables independientes?

2) ¿Cuáles son las interacciones entre las variables independientes?

3) ¿Cuál es la importancia de las variables dependientes?

**Modelo lineal general Multivariante.**

Se tiene n observaciones independientes de p variables observables obtenida de diversas condiciones experimentales. El modelo lineal multivariante es:

Siendo:

* X la matriz de diseño
* βLa matriz de parámetros de regresión desconocidos
* E la matriz de desviaciones aleatorias (errores)

**Estimación de parámetros.**

En el modelo MANOVA se debe estimar los parámetros contenidos en y en la matriz de covarianzas. Donde:

Dado que las hipótesis en el ámbito multivariante puede expresarse como hipótesis del caso univariante, se presenta modelos de una vía y dos vías de clasificación

**Modelos de una vía de clasificación**

Dado

Así la media en cada muestra es

Y la media general se obtiene de

Ahora se plantea el contraste de la igualdad de medias:

*H*0 : *μ*1 = · · · = *μq*

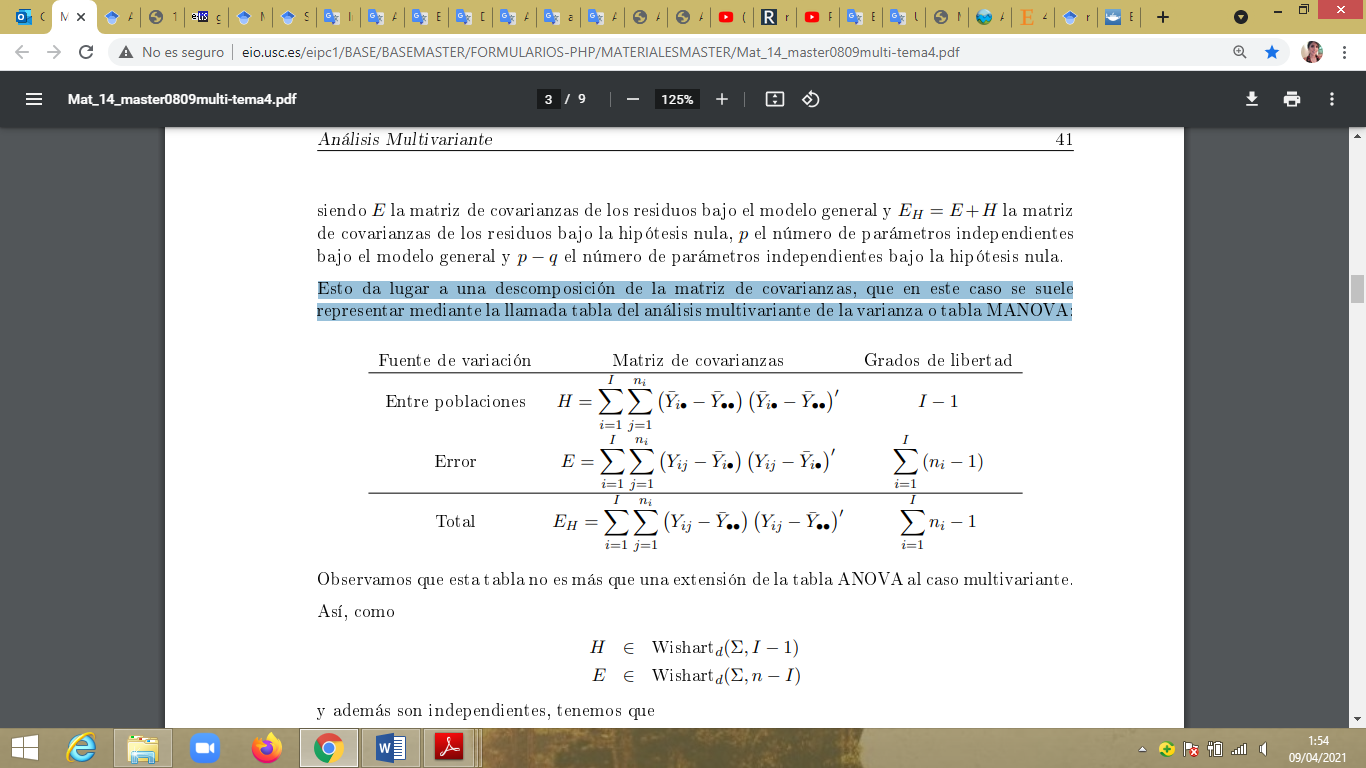
Donde esta hipótesis de puede formular de modo equivalente así:

Ho= µ1 − µq = µ2 − µq = · · · = µq−1 − µq = 0

Este contraste de hipótesis de igualdad se puede llevar a cabo por el método de de Wilks, la cual compara la matriz de covarianzas de los residuos bajo el modelo general y bajo la hipótesis nula. (Sellero., 2008)

Siendo E la matriz de covarianzas de los residuos bajo el modelo general y la matriz de covarianzas de los residuos bajo la hipótesis nula. (Sellero., 2008)

Esto da lugar a una descomposición de la matriz de covarianzas, que en este caso se suele representar mediante la llamada tabla del análisis multivariante de la varianza o tabla MANOVA:



Cuando se obtienen resultados, no es el mismo en análisis de varianza univariante, esto ocurre porque no se considera la relación que puede haber entre las variables. Al igual que en el modelo lineal general univariante, aquí también podemos plantear el procedimiento de unión-intersección para el contraste de la hipótesis nula de igualdad de todas las medias. (Sellero., 2008)

En ese caso, el estadístico de contraste seria el máximo valor propio de y se rechaza la hipótesis nula donde n\*tr () que es mayor que el estadístico

Supuestos

**MANOVA-BIPLOT**

El Manova-Biplot es una técnica estadística multivariante utilizada en situaciones experimentales donde se dispone de varias variables respuesta y se quiere buscar las diferencias entre varios grupos. (Amaro, Vicente-Villardón, & Galindo-Villardón, 2008). En el Manova-Biplot, además de interpretar las diferencias-semejanzas entre los grupos; también podemos interpretar las relaciones entre las variables; y las relaciones entre grupos y variables. Adicionalmente, este nos proporciona medidas de la calidad de representación tanto de medias de grupo como de las variables, que nos permiten una mejor interpretación de los resultados. (Amaro, Vicente-Villardón, & Galindo-Villardón, 2008)

Siguiendo el modelo lineal General Multivariante

Donde se quiere probar .

El Manova-Biplot se construye a partir de la descomposición en valores singulares generalizada:

Siendo

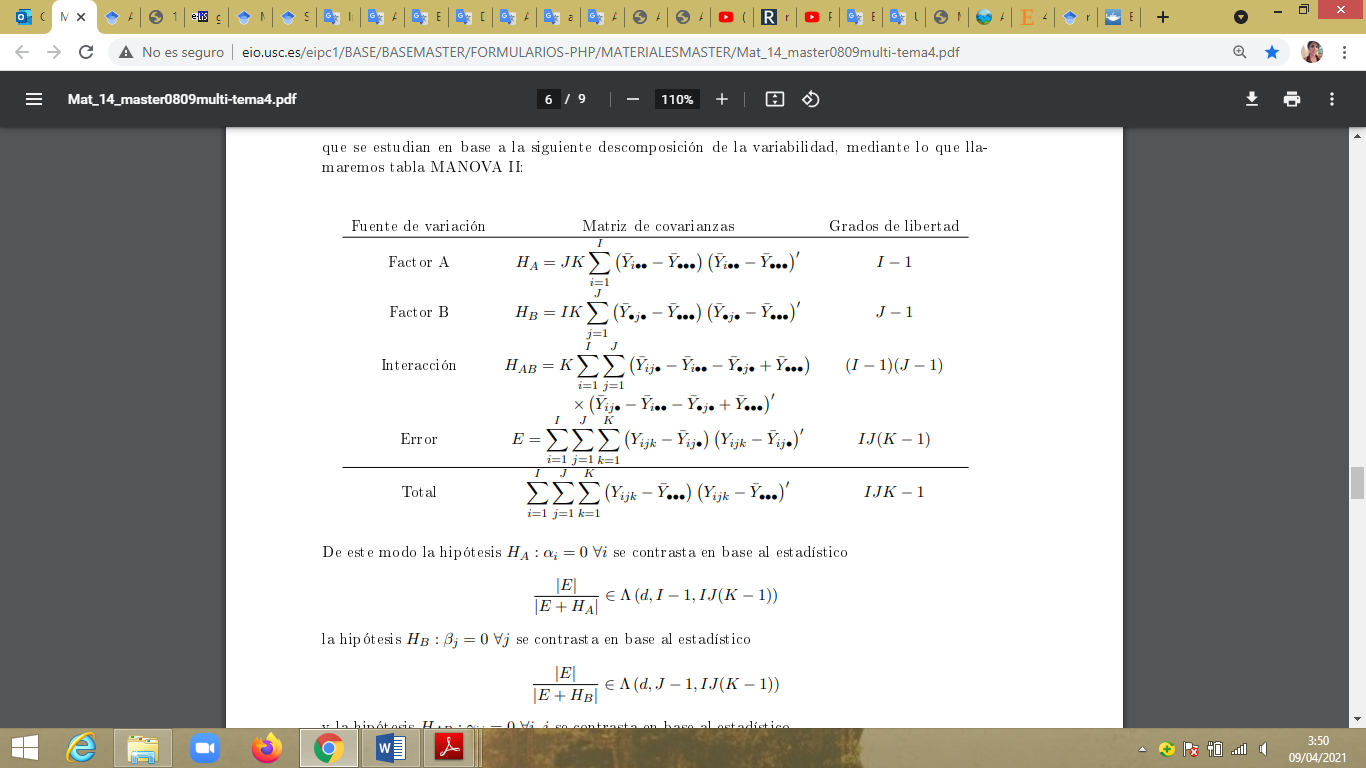
**Manova de dos vías.**

La diferencia en un MANOVA de dos vías es que pueden contrastarse varias hipótesis, para los efectos principales y para la interacción; asociados con cada una de las hipótesis pueden construirse distintas representaciones canónicas que permiten el estudio separado de cada una de ellas. (Amaro, Vicente Villardón, & Galindo Villardón, 2004)

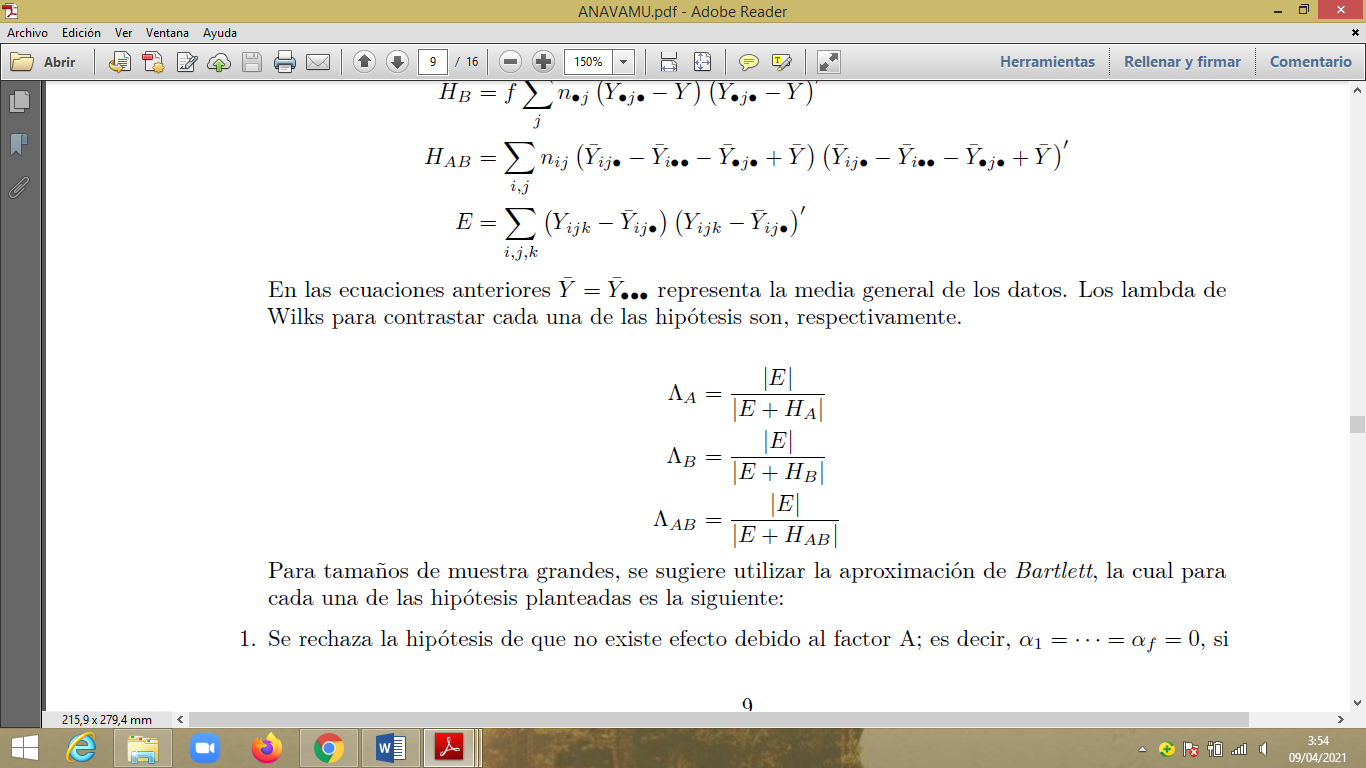
Suponemos un conjunto de datos donde las filas representan los niveles del primer factor y las columnas los niveles del segundo factor, denotados por A y B respectivamente. El modelo que los relaciona seria:

El parámetro µ representa la media global, los parámetros representan el efecto principal del factor A, los parámetros representan el efecto principal del factor B y los parámetros representan la interacción de los factores A y B. (Sellero., 2008)

Para esto se plantean las siguientes hipótesis:



Las lambdas de Wilks para contrastar cada hipótesis son:



**Manova – Biplot de dos vías**

Se proponen 4 tipos de Biplot para representar los resultados de un MANOVA para un diseño de dos vías, todos construidos a partir del Biplot de la matriz de parámetros del modelo MANOVA: el primero que denominamos Biplot Total; y, los otros tres: Biplot de interacción, de filas y de columnas. (Amaro, Vicente Villardón, & Galindo Villardón, 2004)

El Manova- Biplot total tiene las mismas propiedades que el Manova- Biplot de una vía. En particular nos permite estudiar semejanzas y diferencias entre grupos, que en este caso son las combinaciones de niveles de los factores; variables que más importan en las comparaciones entre los grupos, la dirección de máxima separación entre grupos, y las calidades de representación. Sin embargo, en general no es la forma más adecuada para estudiar la interacción, ni los efectos de filas y/o columnas, ya que muestra combinados los efectos principales y la interacción. (Amaro, Vicente Villardón, & Galindo Villardón, 2004)

Las propiedades del Biplot de Interacción son:

1. El Biplot obtenido de esta forma es equivalente a la obtención de las variables canónicas para la hipótesis de interacción.
2. La bondad de ajuste para las medias no se calcula porque la intención de la presentación es la interacción.
3. Para cada uno de los contrastes, la longitud al cuadrado del marcador aproxima la distancia de Mahalanobis al vector nulo, es decir, aproxima la desviación de la hipótesis nula de igualdad a cero para ese contraste particular.

Las propiedades del Biplot de Filas son:

1. El Biplot de filas los ejes son aquellas combinaciones lineales de las variables que maximizan la F de Snedecor para el efecto del factor fila.
2. Este Biplot permite estudiar semejanzas y diferencias entre los grupos construidos según el factor de clasificación por filas, cuando el efecto es significativo.
3. Se pueden construir círculos de confianza alrededor de los centroides de los grupos, de tal forma que si los círculos no se interceptan el efecto del factor fila es significativo. Si se interceptan es necesario hacer otras pruebas para estudiar la significación.
4. Se puede interpretar la importancia de las variables en la separación de los grupos, de manera análoga a como se hace en Manova-Biplot de una vía.

El Biplot de Columnas tiene propiedades semejantes a las del Biplot de filas, solo que en este caso se hacen comparaciones entre los grupos que resultan del factor columna.

Teniendo en cuenta que el Análisis Multivariante de la Varianza de dos vías puede ser expresado como caso particular del Modelo Lineal General Multivariante, se ha generalizado el Manova-Biplot de una vía, propuesto por Gabriel (1972), al caso de dos vías, en el que es posible llevar a cabo cuatro formas de Manova- Biplot: Manova Biplot Total, Manova- Biplot para filas, Manova-Biplot para columnas y Manova-Biplot de interacción. (Amaro, Vicente Villardón, & Galindo Villardón, 2004)

Las teorías de MANOVA y mG tradicionales no proporcionan errores para la estimación de la varianza y componentes de la covarianza, potencialmente enmascarando la precisión de las estimaciones. Sin embargo, el uso de un enfoque bayesiano puede minimizar los problemas mencionados. En este artículo demostramos un enfoque Bayesiano de la teoría mG utilizando la cadena de Markov Monte Carlo (MCMC) técnicas para estimar los parámetros de interés. Las técnicas de MCMC son capaces de trabajar con modelos complejos y matrices de datos dispersos que muchos métodos tradicionales no pueden (ver Muthén & Asparouhov, [2012](https://translate.googleusercontent.com/translate_f#21) , por ejemplo). Además, los usuarios se les permite incorporar información previa para realizar estimaciones más robusto (Lynch, [2007](https://translate.googleusercontent.com/translate_f#21) , p. 2). Por ejemplo, incorporando Las distribuciones previas de los componentes de la varianza permiten a los usuarios restringir las estimaciones al espacio numérico permisible y, por lo tanto, evitar cuestiones como los casos de Heywood (estimaciones de varianza negativa). Si La información previa no está disponible, los antecedentes no informativos pueden ser especificado; la estimación posterior es entonces esencialmente equivalente a estimación tradicional (Jeffreys, [1946](https://translate.googleusercontent.com/translate_f#21) ). Tenga en cuenta que  Las distribuciones anteriores poco informativo se utiliza en todo el artículo. Otro desventaja de un marco bayesiano es que la imputación de datos perdidos se pueden acomodar simultáneamente. Las muestras de MCMC pueden aplicarse a los valores perdidos como parte del algoritmo (Little &Rubin, [2014](https://translate.googleusercontent.com/translate_f#21)). La literatura ha demostrado que los métodos bayesianos son una forma eficaz de imputar los valores perdidos, y el proceso es consistentemente estable (Enders, [2010](https://translate.googleusercontent.com/translate_f#21) ). Por último, similar a los intervalos de confianza de enfoques frecuentistas, los métodos bayesianos proporcionan intervalos creíbles para cada estimación posterior para que los usuarios puedan Investigar el rango probable de los parámetros con un nivel de probabilidad (por ejemplo, 95%). Esta característica elude las suposiciones asintótico, que pueden no ser siempre razonables (Jackman, [2009](https://translate.googleusercontent.com/translate_f#21)) (Jiang & Skorupski, 2018)

Potentes métodos estadísticos multivariados, como el análisis de varianza multivariado tradicional ( MANOVA ), han existido durante décadas (Hotelling 1931; Wilks 1932; Fisher 1936; Bartlett 1939; Lawley 1939; Pillai 1955), pero las pruebas que utilizan estas estadísticas se basan en suposiciones de que generalmente no se cumplen con los datos ecológicos. El supuesto que los datos se ajustan a una distribución normal multivariante La contribución es particularmente poco realista para la mayoría de los conjuntos de datos. Esto se debe a que las distribuciones de abundancia Las danzas de especies individuales suelen ser muy agregadas. cerrado o sesgado (p. ej., Gaston y McArdle 1994). También, las abundancias toman valores discretos en lugar de ser continua, las especies con pequeños medios a menudo tienen distribuciones asimétricas porque son necesariamente truncado en cero, y las especies raras contribuyen mucho ceros al conjunto de datos. Las estadísticas de prueba MANOVA no son particularmente robusto a las desviaciones de las normas multivariadas maldad (Mardia 1971; Olson 1974; Johnson & Field 1993). Finalmente, muchas de estas estadísticas de prueba son simplemente imposible de calcular cuando hay más variables que las unidades de muestreo, lo que a menudo ocurre en aplicaciones. (ANDERSON, 2001)

El análisis de varianza multivariante permutacional (PERMANOVA) es una prueba estadística multivariante no paramétrica. Se utiliza para comparar grupos de objetos y probar la hipótesis nula de que los centroides y la dispersión de los grupos definidos por el espacio de medida son equivalentes para todos los grupos. Un rechazo de la hipótesis nula significa que el centroide y / o la dispersión de los objetos es diferente entre los grupos. Por lo tanto, la prueba se basa en el cálculo previo de la distancia entre dos objetos incluidos en su experimento. (Ebner, 2018)

Tres suposiciones subyacen a las pruebas de significancia en MANOVA: (1) independiente de las observaciones, (2) normalidad multivariante del grupo de población dependiente variables, y (3) homogeneidad de la varianza-covarianza de la población de grupo matrices. Cada uno de estos supuestos debe evaluarse como parte del análisis. (Grice, 2007)

En estadística , el problema de Behrens-Fisher , que lleva el nombre de Walter Behrens y Ronald Fisher , es el problema de la estimación de intervalos y la prueba de hipótesis sobre la diferencia entre las medias de dos poblaciones distribuidas normalmente cuando las varianzas de las dos poblaciones no se suponen iguales , basado en dos muestras independientes . Problema de Behrens-Fisher - <https://es.qaz.wiki/wiki/Behrens%E2%80%93Fisher_problem>

Vale la pena mencionar que los problemas heterocedásticos de ANOVA y MANOVA han recibido mucha atención en las últimas décadas debido a su naturaleza interesante y desafiante. Hay dos tipos principales de métodos: enfoques basados en simulación y enfoques basados en aproximaciones. Las referencias mencionadas anteriormente, entre otras, son generalmente enfoques basados en aproximaciones. En la literatura, los enfoques basados en simulación también son populares debido a su simplicidad y al avance reciente de las capacidades de cálculo. Por ejemplo, para tratar la heterocedasticidad, [21] propuso una prueba aproximada basada en simulación. [22] propuso otra prueba basada en simulación utilizando el concepto de valores P generalizados, lo que resultó en una prueba F generalizada. [23] y [24] estudian con más detalle el método de los valores P generalizados. [25] comparó el rendimiento de una prueba de arranque paramétrico con la prueba F generalizada de [22]. [26] propuso un método bootstrap y [27] estudió una prueba de permutación para problemas generales de ANOVA. Un inconveniente de los enfoques basados en simulación es que a menudo consumen mucho tiempo, especialmente cuando los tamaños de muestra son grandes. (Zhang, Zhou, Guo, & Liu, 2016)

Este contraste de hipótesis de igualdad se puede llevar a cabo por el método de de Wilks, la cual compara la matriz de covarianzas de los residuos bajo el modelo general y bajo la hipótesis nula.

Siendo E la matriz de covarianzas de los residuos bajo el modelo general y la matriz de covarianzas de los residuos bajo la hipótesis nula, nivel de significancia, son los grados de libertad para la hipótesis, son los grados de libertad del error.