Facultad
de
Ciencias

EVALUACIÓN DEL ANÁLISIS DE LA VARIANZA MEDIANTE SIMULACIÓN
(Evaluation of analysis of variance through simulation)

Trabajo de Fin de Grado
para acceder al

GRADO EN MATEMÁTICAS

Autor: Miguel Pérez Antuñano

Directora: Alicia Nieto Reyes

Octubre, 2024

Resumen

El análisis de la varianza en dimensión 1 es un método estadístico ampliamente utilizado para determinar si existen diferencias significativas entre las medias de varios grupos. En su forma multivariante, este análisis se vuelve aún más interesante, ya que permite evaluar múltiples variables dependientes simultáneamente, proporcionando una visión más completa de las diferencias entre los grupos. En este trabajo se repasará la teoría fundamental que sustenta el análisis de la varianza multivariante (MANOVA), así como el análisis funcional de la varianza (FANOVA), y se presentarán los tests de hipótesis más comúnmente utilizados en este contexto. Además, se realizarán simulaciones para demostrar la aplicación práctica del MANOVA y FANOVA. Finalmente, se propondrá un método bootstrap para evaluar las diferencias de media entre dos grupos, ofreciendo una alternativa robusta cuando los supuestos tradicionales no se cumplen.

Palabras clave: Estadística, MANOVA, FANOVA, bootstrap.

Abstract

The analysis of variance in one dimension is a widely used statistical method to determine if there are significant differences between the means of several groups. In its multivariate form, this analysis becomes even more interesting as it allows the evaluation of multiple dependent variables simultaneously, providing a more comprehensive view of the differences between groups. In this work, we will review the fundamental theory underlying multivariate analysis of variance (MANOVA) and functional analysis of variance (FANOVA) and present the most commonly used hypothesis tests in this context. Additionally, simulations will be conducted to demonstrate the practical application of MANOVA and FANOVA. Finally, a bootstrap method will be proposed to evaluate the mean differences between two groups, offering a robust alternative when traditional assumptions are not met.

Keywords: Statistics, MANOVA, FANOVA, bootstrap.

Índice general

1. Introducción del problema inicial **4**
1.1. Definición del Problema . . . 4
1.2. Planteamiento de las Hipótesis . . . 5
1.3. Enfoques de Análisis . . . 5
1.3.1. MANOVA (Multivariate Analysis of Variance) . . . 5
1.3.2. FANOVA (Functional Analysis of Variance) . . . 5
1.3.3. Análisis de Varianza Basado en Bootstrap . . . 5
1.4. Motivación del Estudio . . . 5
1.5. Estructura del Trabajo . . . 6

2. Conocimientos previos **7**
2.0.1. ANOVA . . . 7
2.1. Distribución normal multivariada . . . 9
2.2. Distribución de Wishart . . . 9
2.3. Distribución Lambda de Wilks . . . 10
2.4. Distribución T^2 de Hotellings . . . 11
2.5. Estimación de máxima verosimilitud . . . 12

3. Análisis de la varianza con MANOVA **15**
3.1. Algoritmo . . . 15
3.2. Métodos de Contraste en MANOVA . . . 16

4. Análisis de la varianza con FANOVA **18**
4.1. Test FP . . . 19
4.2. Test CH . . . 20
4.3. Test CS . . . 20
4.4. Test de las normas L_2 . . . 20
4.4.1. L_{2N} . . . 20
4.4.2. L_{2B} . . . 20
4.4.3. L_{2b} . . . 21
4.5. FN, FB, Fb . . . 21
4.6. GPF . . . 21
4.7. F_{max_b} . . . 21
4.8. TRP . . . 22
4.8.1. TRP_{anova} . . . 22
4.8.2. TRP_{ATS} . . . 22
4.8.3. TRP_{WTPS} . . . 22

5. Otro análisis de la varianza **23**
5.1. Algoritmos genéticos . . . 26
5.2. Algoritmo de simulación de recocido . . . 28

6. Simulaciones **30**
6.1. Prueba 1 . . . 31
6.2. Prueba 2 . . . 32
6.2.1. MANOVA . . . 32
6.2.2. FANOVA . . . 32
6.2.3. Bootstrap . . . 33
6.3. Prueba 3 . . . 33
6.3.1. MANOVA . . . 33
6.3.2. FANOVA . . . 34
6.3.3. Bootstrap . . . 35
6.4. Prueba 4 . . . 35
6.4.1. MANOVA . . . 35
6.4.2. FANOVA . . . 36
6.4.3. Bootstrap . . . 36
6.5. Prueba 5 . . . 37
6.5.1. MANOVA . . . 37
6.5.2. FANOVA . . . 38
6.5.3. Bootstrap . . . 38
6.6. Prueba 6 . . . 38
6.6.1. MANOVA . . . 39
6.6.2. FANOVA . . . 40
6.6.3. Bootstrap . . . 41

7. Conclusiones **43**

8. Apéndice **46**

1 Introducción del problema inicial

Históricamente, un problema común ha sido determinar si existen diferencias significativas entre dos o más grupos dentro de una misma población, en donde cada grupo recibe un 'tratamiento' o se diferencian en un cierto factor. Este problema fue introducido para el caso de una variable por el matemático Fisher en 1919, ampliado en años posteriores hasta que en 1925 publicó su libro "Statistical Methods for Research Workers" [21] en donde expone el ANOVA, un análisis de la varianza con el que se pueden analizar varios grupos en busca de diferencias.
Posteriormente, esta idea fue ampliada para el caso multivariante, creando así el análisis MANOVA, que pretendía generalizar las ideas del ANOVA para el caso con varias variables dependientes. Su principal finalidad es poder analizar varias variables de forma simultánea. En este desarrollo, Harold Hotelling fue uno de los matemáticos que expandió el conocimiento sobre la estadística en el caso multivariante, introduciendo la distribución T^2 [28], una generalización de la t de Student [29]. En cuanto al MANOVA, fue introducido en torno al año 1930, pero no fue tan popular y utilizado hasta décadas después. Uno de estos libros donde se muestra de forma formal el MANOVA es en "An Introduction to Multivariate Statistical Analysis" [22], escrito por T. W. Anderson. En las últimas décadas, a partir del año 1990, se han visto otros enfoques como el análisis funcional de la varianza, FANOVA. La diferencia principal entre este análisis y los dos anteriores es que supone que se tienen datos funcionales y no observaciones discretas. En este campo, algunos matemáticos que han aportado conocimientos nuevos son M. Antoniadis [25], contribuyendo a la estadística funcional, Pedro Galeano [26] o Antonio Cuevas [27], cuyos aportes en la teoría del FANOVA son significativos. Entre las numerosas obras que se encuentran sobre este tema, están "Functional Data Analysis", de J. O. Ramsay y B. W. Silverman [23] y "Nonparametric Functional Data Analysis: Theory and Practice" de Ferraty, F. y Vieu, P [24]. Además de estos matemáticos, muchos otros han aportado su conocimiento en esta última rama del FANOVA, que, gracias a que en la actualidad es un tema muy interesante por las innumerables aplicaciones, está siendo muy estudiado. En este trabajo se buscará seguir ampliando el caso multivariante, aportando un nuevo algoritmo para el análisis multivariante de la varianza.

1.1. Definición del Problema

Primero, se definirá el problema multivariante de forma general, para posteriormente reducirlo al problema que se analizará en el trabajo.
Se fija un espacio probabilístico $(\Omega, \sigma, \mathbb{P})$. Se definen las variables aleatorias independientes Y_{ij}, con $i = 1, \ldots, m$, $j = 1 \ldots, n_i$. Dado un i fijo, se tiene que las variables aleatorias Y_{ij} son idénticamente distribuidas por una distribución Q_i. Cada variable aleatoria Y_{ij} se puede reescribir como $Y_{ij} = (Y_{ij1}, \ldots, Y_{ijd})$.

En este problema se estudia la existencia de diferencias entre las distribuciones Q_i. El objetivo del trabajo se limitará el estudio a dos grupos, es decir, i tomará los valores 1 o 2, por lo que se reduce el problema a estudiar la existencia de diferencias entre dos distribuciones. Entonces, suponiendo que exista la esperanza en las dos distribuciones, se tendrá que si sus esperanzas son distintas, podremos confirmar que son distribuciones distintas, pero, si sus esperanzas son iguales, no se podrá asegurar que en verdad son distribuciones idénticas. (Un ejemplo sencillo que muestra esto podrían ser dos distribuciones normales de media 0 y desviaciones típicas distintas, o una distribución normal de media 0 y una distribución uniforme U[-a,a], tienen misma media pero no son las mismas distribuciones).
Por tanto, una vez explicado todo lo anterior, se introduce la siguiente sección:

1.2. Planteamiento de las Hipótesis

Para abordar esta cuestión, planteamos las siguientes hipótesis estadísticas:

- $H_0 : \mu_1 = \mu_2$
- $H_a : \mu_1 \neq \mu_2$.

Para poder hacer un estudio, se toma $\omega \in \Omega$ y se evalúan las variables aleatorias Y_{ij}. Se tienen por tanto $Y_{ij}(\omega) = y_{ij} = (y_{ij1}, \ldots, y_{ijd})$.
Para estimar las esperanzas de cada grupo, se puede usar la media maestral de cada grupo. Se define la media maestral del grupo i como $\overline{y_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}(\omega) = \overline{Y_i}(\omega)$.

1.3. Enfoques de Análisis

Para evaluar estas hipótesis, se seguirán tres caminos diferentes. Estas formas del análisis se revisarán más a fondo en cada capítulo correspondiente. Las dos primeras propuestas son técnicas ya conocidas, por lo que se explicarán y se probará su funcionamiento, mientras que la última propuesta, es un algoritmo novedoso que se explicará y se comparará con el resto de técnicas ya conocidas.

1.3.1. MANOVA (Multivariate Analysis of Variance)

El análisis multivariante de la varianza (MANOVA) es una extensión del ANOVA que permite analizar múltiples variables dependientes simultáneamente. MANOVA evalúa si los vectores de medias diferenciales entre los grupos son significativamente diferentes, considerando la correlación entre las variables dependientes.

1.3.2. FANOVA (Functional Analysis of Variance)

El análisis funcional de la varianza (FANOVA) es una técnica que extiende el análisis de varianza a datos funcionales, es decir, datos que se consideran como funciones en lugar de puntos discretos. Esta técnica es útil cuando los datos tienen una estructura temporal o espacial que puede ser modelada como funciones continuas.

1.3.3. Análisis de Varianza Basado en Bootstrap

Este método novedoso basado en el bootstrap, garantiza robusted en los resultados y unas hipótesis iniciales más relajadas al no necesitar de las suposiciones tradicionales del MANOVA. Primeramente se hablará del algoritmo, sus posibles variantes y se estudiará el método para ver su comportamiento en simulaciones.

1.4. Motivación del Estudio

El análisis de la varianza en contextos multivariantes tiene aplicaciones en diversas áreas como la biología, la economía, la psicología y la ingeniería, entre otras. Por ejemplo, en estudios de genética, se puede utilizar MANOVA para evaluar si hay diferencias significativas en la expresión de múltiples genes entre diferentes grupos de tratamiento. En economía, el análisis de la varianza puede ayudar a determinar si diferentes políticas económicas tienen efectos distintos en varias variables macroeconómicas.
La motivación principal de este trabajo es proporcionar una comprensión profunda de estos métodos y su aplicación en la detección de diferencias significativas entre conjuntos de datos multivariantes. Al

comparar y contrastar MANOVA, FANOVA y el nuevo análisis de varianza basado en bootstrap, se espera ofrecer una guía comprensiva para los investigadores que enfrentan problemas similares en sus propias disciplinas.

1.5. Estructura del Trabajo

Este trabajo se estructura de la siguiente manera:

- **Capítulo 1: Introducción del Problema Inicial**:
  Se presenta la motivación y el planteamiento de las hipótesis.
- **Capítulo 2: Conocimientos previos**:
  Se detallarán todos los conocimientos previos que son necesarios para el análisis MANOVA.
- **Capítulo 3: Fundamentos de MANOVA**:
  Se detallan los conceptos y la metodología detrás del MANOVA.
- **Capítulo 4: Análisis Funcional de la Varianza (FANOVA)**:
  Se exploran los métodos del FANOVA. Debido a que existen muchas variantes y en algunos casos la teoría detrás de estos modelos es avanzada y podría extender demasiado la memoria, se opta por explicar de una forma más superficial que el MANOVA en esta sección.
- **Capítulo 5: Análisis de Varianza Basado en Bootstrap**:
  Se describen las técnicas detrás de este análisis, las variantes a la hora de optimizar una parte del código que se verá en el capítulo y un análisis de la complejidad temporal del algoritmo.
- **Capítulo 6: Simulaciones**:
  Se comparan los resultados obtenidos con cada método por las simulaciones y se discuten sus implicaciones.
- **Capítulo 7: Conclusiones**:
  Se presentan las conclusiones del estudio.

2 Conocimientos previos

En esta sección se repasarán conocimientos previos que serán necesarios para el desarrollo del trabajo, como la distribución normal multivariada, que se suele ver en clases de probabilidad y estadística avanzada, o como otras distribuciones no vistas en clase como la distribución de Wishart o la Λ de Wilks, o conocimientos sobre estimación de máxima verosimilitud, que si bien se introdujo en clase de inferencia estadística, en este trabajo se hablará con el fin de introducir uno de los test multivariado. La información de esta sección es un compendio de información extraída principalmente de dos libros; Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate Analysis.[10] y Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis [11]. Del primer libro se ha extraído casi toda la parte que tiene que ver con las distribuciones normal, Wishart, Λ, (definiciones y teoremas). Mientras que para la parte donde se habla sobre la estimación de máxima verosimilitud, se ha usado principalmente la segunda referencia, aunque también se han seguido usando las ideas del primer libro como referencia.
El motivo por el que se han introducido estas distribuciones es para acabar introduciendo la distribución de Λ de Wilks, que será importante para el desarrollo del MANOVA.
También en esta sección se incluirá un breve resumen del ANOVA clásico [30]. Este modelo que se presentó en clase de inferencia estadística se va a resumir porque aunque en el resto del trabajo no se tenga en cuenta, ya que no es un modelo propio para un caso multivariante, es una de las primeras ideas que se tuvieron para el análisis de la varianza, lo que dio lugar mas adelante al MANOVA y al resto de modelos que se usan en el análisis multivariante de la varianza.

2.0.1. ANOVA

El ANOVA es un modelo para el problema planteado en una dimensión. (d=1). Dentro de este modelo, se pueden tener en cuenta, o bien un factor o bien múltiples factores. Un factor es aquella variable categórica que divide la población en grupos. Por ejemplo, en una población de personas que toman pastillas para una cierta enfermedad, un factor podría ser el tipo de pastilla. En esta sección solo se introducirá el caso de un factor.
Para el análisis de un factor, una vez tomadas las muestras de una población, el modelo usado en el ANOVA puede escribirse de la siguiente forma:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Donde $i = 1, \ldots, k$, $j = 1, \ldots, n_i$ y $N = \sum_{i=1}^{k} n_i$. Y_{ij} es la muestra j-ésima del grupo i-ésimo. μ es la media correspondiente al 'tratamiento', α_i corresponde con el valor de la corrección por el tratamiento i-ésimo al que está siendo sometido y por último, ϵ_{ij} es el valor del error del modelo para la muestra j del grupo del tratamiento i. Además, se deberá tener en cuenta que $\sum_{i=1}^{k} n_i\alpha_i = 0$ para que todo esté bien definido.

Las hipótesis iniciales eran que todas las medias de cada grupo son iguales. $\mu_1 = \ldots = \mu_k$. Donde $\mu_i = \mu + \alpha_i$. Por tanto si se cumple la hipótesis nula, se tendrá que $\alpha_1 = \ldots = \alpha_k = 0$. (Se puede comprobar fácilmente sustituyendo todos los valores μ_i y viendo que la única solución posible para la cadena de igualdades es 0). Tendremos además como hipótesis, que todos los ϵ_{ij} deberán ser independientes, normales de media 0 y misma desviación típica σ.

Tomando $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)^2$. Se puede descomponer como: $\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}((Y_{ij}-Y_{i\cdot}) + (Y_{i\cdot} - \overline{Y} - \alpha_i) + (\overline{Y} - \mu))^2 =$
$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Y_{i\cdot})^2 + (Y_{i\cdot} - \overline{Y} - \alpha_i)^2 + (\overline{Y} - \mu)^2 + 2((y_{ij} - Y_{i\cdot})(Y_{i\cdot} - \overline{Y} - \alpha_i) + (Y_{ij} - Y_{i\cdot})(\overline{Y} - \mu) + (Y_{i\cdot}\overline{Y} - \alpha_i)(\overline{Y} - \mu)) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Y_{i\cdot})^2 + \sum_{i=1}^{k}(Y_{i\cdot} - \overline{Y} - \alpha_i)^2 + N(\overline{Y} - \mu)^2$
La última expresión se puede obtener gracias a que se tiene que se pueden seguir simplificando los productos cruzados sabiendo que $\sum_{i=1}^{k} n_i(Y_{i\cdot} - \overline{Y})^2 = 0$ y $\sum_{j=1}^{n_i}(Y_{ij} - Y_{i\cdot})^2 = 0$, además de que $\sum_{i=1}^{k} n_i\alpha_i = 0$ cuando se está en las condiciones de la hipótesis nula.

Por último, se necesitará introducir el teorema de Cochran:

Teorema 2.0.1 *(Teorema de Cochram) Sean las variables aleatorias $X_1, \ldots, X_n$ independientes e idénticamente distribuidas por $N(0,1)$, se verifica la descomposición:*

$$\sum_{i=1}^{n} X_i^2 = Q_1(X_1, \ldots, X_n) + \ldots + Q_k(X_1, \ldots, X_n)$$

Donde Q_i es una forma cuadrática de rango r_i. Si $n = r_1 + \ldots + r_k$, entonces las variables aleatorias $Q_1(X_1, \ldots, X_n), \ldots, Q_k(X_1, \ldots, X_n)$ son independientes y la distribución de cada una de ellas es $\chi^2_{r_i}$ [1]

Habría que introducir también la distribución F.

Definición 2.0.1 *Dadas dos variables aleatorias independientes, A con distribución χ^2_n y B con distribución χ^2_m. Entonces se llama distribución F con (n,m) grados de libertad a la de la variable*

$$F_{(n,m)} = \frac{A/n}{B/m}$$

Sabiendo esto, se puede llegar a la conclusión de que si se cumple la hipótesis nula, entonces, $\alpha_i = 0$ y todos los μ_i son iguales, entonces

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Y_{i\cdot})^2 + \sum_{i=1}^{k} n_i(Y_{i\cdot} - \overline{Y})^2$$

Si se cumple la hipótesis nula, con el teorema de Cochram, se cumple que la parte derecha son formas cuadráticas, luego siguen una distribución χ^2 de $N - k$ y k-1 grados de libertad respectivamente siempre y cuando estas formas cuadráticas se dividad por σ^2. Luego se puede aplicar la definición de la distribución F y se tiene que en el caso de cumplirse la hipótesis nula, se tiene:

$$F = \frac{\frac{1}{k-1}\sum_{i=1}^{k} n_i(Y_{i\cdot} - \overline{Y})^2}{\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - Y_{i\cdot})^2}$$

Cuando se cumple la hipótesis nula seguirá una distribución $F_{k-1,N-k}$. También se tendrá que todos los valores de $Y_{i\cdot}$ serán parecido y cercanos a $\overline{Y}$. Entonces el numerador será pequeño. Si la hipótesis nula es falsa, existirán ciertos $Y_{i,\cdot}$ que tendrán diferencias significativas, luego se tendría que el numerador será mayor. Por otro lado, se cumpla o no la hipótesis nula, el denominador es solamente un estimador de σ^2. En conclusión, el valor de F solo depende del numerador a la hora de tomar los valores esperados de una distribución $F_{k-1,N-k}$ o mayores de lo esperado.

[1] Como algo puramente notacional, se tiene que una variable aleatoria cualquiera, no necesariamente las variables aleatorias Y_{ij} de las que se habla en esta memoria, se expresará con una notación distinta, como en este caso, que se han llamado $X_1, \ldots, X_n$. El motivo principal es intentar mantener la mayor generalización posible y la mejor comprensión.

Para finalizar esta introducción, hay que saber que en esta sección en muchos momentos se habla de 'matriz de datos'. Cuando se toma una muestra de unas variables aleatorias $X_1, \ldots, X_n$, donde cada X_i tiene su imagen en $\mathbb{R}^d$:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix}$$

Entonces, cuando se nombre una matriz de datos, se hará referencia a una muestra aleatoria simple de las variables aleatorias multidimensionales que correspondan en ese contexto.

2.1. Distribución normal multivariada

Definición 2.1.1 *Sea $X = (X_1, \ldots, X_d)$ un vector aleatorio, se dice que X sigue una distribución normal multivariante, ($X \sim N_d(\mu, \Sigma)$) si cumple una de las siguientes condiciones equivalentes:*

- *Toda combinación lineal $Y = a_1X_1 + \ldots + a_nX_d$ esta normalmente distribuida.*
- *Hay un vector $Z = (Z_1, \ldots, Z_d)$ cuyas componentes son variables aleatorias distribuidas idénticas e independientes de una distribución normal estándar y hay un vector μ y una matriz A tal que $X = AZ + \mu$*
- *Hay un vector μ y una matriz semidefinida positiva Σ tal que la función característica es: $\Phi_X(u; \mu, \Sigma) = exp(i\mu^T u - \frac{1}{2}u`T\Sigma u)$*

Teorema 2.1.1 *Sea X es una muestra aleatoria simple de tamaño $n \times d$ proveniente de una distribución $N_d(\mu, \Sigma)$ y sea $Y \in \mathbb{R}^{m \times q}$ tal que $Y = AXB$, entonces Y es una matriz de datos de una normal $N_q(\alpha B^T\mu, \beta B^T\Sigma B)$ si y solo si se cumplen las dos condiciones siguientes:*

- *$A1 = \alpha 1$ para algún escalar o bien $B^T u = 0$*
- *$AA^T = \beta I$ para un escalar β o bien $B^T\Sigma B = 0$*

2.2. Distribución de Wishart

Definición 2.2.1 *Sea M una matriz de tamaño $p \times p$. Si $M = X^TX$ donde X es una matriz de tamaño $m \times p$ proveniente de una distribución $N_p(0, \Sigma)$, entonces M sigue una distribución Wishart con matriz escala Σ y m grados de libertad. Se denota como $M \sim W_p(\Sigma, m)$.*

Se dice que la distribución de Wishart es estándar si σ es la identidad. Se podría decir que es una generalización matricial de la distribución $\sigma^2\chi^2$, ya que si $p = 1$, entonces $M = x^Tx$, con x un vector de dimensión m, cuyas componentes son variables aleatorias identicamente distribuidas e independientes provenientes de $N(0, \sigma^2)$. entonces en este caso la distribución $W(\Sigma, m) = \sigma^2\chi^2_m$, por lo que vemos que Σ juega el mismo papel que σ^2 en la distribución $\sigma^2\chi^2_m$. También se puede ver que $E[M = \sum_{i=1}^m E[x_ix - i^T] = m\Sigma$.

Teorema 2.2.1 *Si $M \sim W_p(\Sigma, m)$ y B es una matriz de dimensión $p \times q$, entonces $B^TMB \sim W_p(B^T\Sigma B, m)$*

Demostración: $B^TMB = B^TX^TXB = (XB)^T(XB) = Y^TY$ siendo $Y = IXB$. Con I la identidad. Como se tiene que $X \sim N_p(0, \Sigma)$, el teorema 2.1.1 se puede aplicar, ya que:

- $I1 = \alpha 1$ con $\alpha = 1$

- $II^T = 1I$.

Y por tanto $Y \sim N_q(0, B^T \Sigma B)$. De esta manera, usando la definición de distribución de Wishart, se tiene demostrado. Otro teorema interesante es el siguiente:

Teorema 2.2.2 *Sean $M_1 \sim W(\Sigma, m_1)$ y $M_2 \sim W(\Sigma, m_2)$ independientes, entonces $M_1 + M_2 \sim W(\Sigma, m_1 + m_2)$*

Demostración: Se tiene por la definición, que $M_1 + M_2 = X_1^T X_1 + X_2^T X_2$. Por otro lado, se tiene que como M_1 y M_2 son independientes, entonces X_1 y X_2 deberán ser independientes también. Por lo que si

$$X = \begin{pmatrix} X_1^T \\ X_2^T \end{pmatrix}$$

entonces X tiene $m_1 + m_2$ filas independientes y, además, se tiene que

$$X^T X = \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = X_1^T X_1 + X_2^T X_2 = M_1 + M_2$$

Por lo que $M_1 + M_2 \sim W(\Sigma, m_1 + m_2)$.

2.3. Distribución Lambda de Wilks

Definición 2.3.1 *Sean $A, \sim W_p(I, m)$ y $B \sim W_p(I, n)$ matrices independientes, $p \geq m$. definimos a la distribución conjunta del* $\text{mín}(n, p)$ *de los valores propios no cero de $A^{-1}B$ como $\psi(p, m, n)$.*

Teorema 2.3.1 *Para $m \geq p$ y $p, n \geq 1$, las distribuciones $\psi(p, m, n)$ y $\psi(n, n + m - p, p)$ Son equivalentes.*

Demostración: Lo primero, ambas distribuciones tienen el mismo número de valores propios distintos de cero. Suponiendo que $n \leq p$, entonces por la definición, $\psi(p, m, n)$ es la distribución de los valores propios no ceros de $A^{-1}B$, donde $A \sim W_p(I, m)$ y $B \sim W_p(I, n)$ son independientes. Si $B = X^T X$ y si definimos $G := (XX^T)^{-1/2} X$, veremos que $GG^T = I$.

$$GG^T = (XX^T)^{-1/2} X((XX^T)^{-1/2} X)^T = (XX^T)^{-1/2} XX^T ((XX^T)^{-1/2})^T = BB^{-1} = I$$

Entonces,

$$A^{-1}B = A^{-1}(X^T, X) = A^{-1}(XG^T G)^T (XG^T G) = A^{-1} G^T G X^T X G^T G$$

Tiene los mismos valores propios que $(GA^{-1}G^T)(GX^T XG^T) = C^{-1}D$.
Hay que ver que $C \sim W_n(I, m+n-p)$ y $D \sim W_n(I, p)$. Vemos que $D = GX^T XG^T = X^T X \sim W_n(I, p)$ y también como A es independiente de X, también lo será de D y por tanto tenemos independencia entre C y D. De una forma análoga, se hace para el caso de $n > p$, se empieza por $\psi(n, n + m - p, p)$.

Definición 2.3.2 *Dados $A, \sim W_p(I, m)$ y $B \sim W_p(I, n)$ matrices independientes, $p \geq m$. Decimos que $\Lambda = \frac{|A|}{|A+B|} = |I + A^{-1}B|^{-1} \sim \Lambda(p, m, n)$.*

Esta distribución suele verse en el contexto del test del principio de la razón de verosimilitud, donde m representa los grados de libertad del "error" y n los grados de libertad de la hipótesis nula, por lo que los grados de libertad totales son m + n.

Teorema 2.3.2 *Las distribuciones $\Lambda(p, m, n)$ y $\Lambda(n, m + n - p, p)$ son equivalentes.*

Demostración: Sean $\lambda_1, \ldots, \lambda_n$ los valores propios de $A^{-1}B$, siendo A y B las matrices de la definición 2.3.2. Sea $k = \text{mín}(n, p)$ el número de valores propios distintos de 0.

Como paso previo, se tiene que dada una matriz M, con valores propios $\alpha_1, \ldots, \alpha_n$, el determinante de M, $|M| = \prod_{i=1}^{n} \alpha_i$.

Los valores propios de $M = I + A^{-1}B$ son $1 + \lambda_i$ con λ_i un valor propio de $A^{-1}B$, porque dado v_i el vector propio asociado a λ_i, se tiene que $(I + A^{-1}B)v_i = Iv_i + A^{-1}Bv_i = v_i + \lambda_i v_i = (1 + \lambda_i)v_i$.
Entonces, $|I + A^{-1}B|^{-1} = \prod_{i=1}^{p}(1 + \lambda_i)^{-1} = \prod_{i=1}^{k}(1 + \lambda_i)^{-1}$. Como Λ es una distribución de los valores propios distintos de 0 de $A^{-1}B$, entonces se demuestra con el teorema 2.3.1.
Con este resultado, puede demostrarse que:

$$\frac{1 - \Lambda(p, m, 1)}{\Lambda(p, m, 1)} \sim \frac{p}{m - p + 1} F_{p, m-p+1}$$

$$\frac{1 - \Lambda(1, m, n)}{\Lambda(1, m, n)} \sim \frac{n}{m} F_{n,m}$$

$$\frac{1 - \sqrt{\Lambda(p, m, 2)}}{\Lambda(p, m, 2)} \sim \frac{p}{m - p + 1} F_{2p, 2(m-p+1)}$$

$$\frac{1 - \sqrt{\Lambda(2, m, n)}}{\Lambda(2, m, n)} \sim \frac{n}{m - 1} F_{2n, 2(m-1)}$$

En la literatura hay muchas otras aproximaciones de la lambda de Wilks con otros parámetros distintos. Estas aproximaciones son las propuestas en el libro de Multivariate Analysis [10] Esta distribución surge de los test de razón de verosimilitud. Estas aproximaciones pueden ser de utilidad para los test de hipótesis del MANOVA.

2.4. Distribución T^2 de Hotellings

Definición 2.4.1 *Se dice que α está distribuido por una distribución T^2 de Hotellings si existen m, d, M de modo que $\alpha = md^T M^{-1} d$ con $d \sim N_p(0, I)$ y $M \sim W_p(I, m)$ y además d y M son independientes.*

Esta distribución se usa para el test T de Hotelling. El test puede encontrarse desarrollado en el libro de "The generalization of Student's ratio", escrito por Hotelling [28]. El test es la generalización del test t de Student del caso univariante.
Como una pequeña introducción, el test T de Hotelling tiene usos similares al test t de Student pero en el caso multivariante. También tiene unos supuestos adaptados para el caso multivariante: Igualdad en las matrices de covarianza, independencia entre las observaciones, normalidad de la muestra y un tamaño muestral adecuado. Como un supuesto extra que tiene este test multivariante, es ventajoso para el test T que las variables tienen una relación aproximadamente lineal y las variables poco correlacionadas. Esta suposición no es necesaria, pero si ayuda al test a funcionar mejor ya que simplifica cálculos necesarios.

2.5. Estimación de máxima verosimilitud

La estimación de máxima verosimilitud es aquella estimación de un parámetro que maximiza la probabilidad de las observaciones dadas en la muestra. Esta teoría se vio en la asignatura de inferencia estadística, pero aquí se verá en una mayor profundidad el caso multivariante. La utilidad de comentar este tema es, para iniciar la teoría para el análisis de la varianza multivariante.

Empezando por unas muestras obtenidas de una normal $N_p(\mu, \Sigma)$, $X_1, \ldots, X_n$, donde cada X_i es un vector de la muestra de tamaño p. Con esto, la función de densidad conjunta de $X_1, \ldots, X_n$ es:

$$\prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp((-x_j - \mu)^T \Sigma^{-1}(-x_j - \mu)/2)$$

$$= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp(-1/2 \sum_{j=1}^{n} (x_j - \mu)^T \Sigma^{-1}(-x_j - \mu))$$

Ahora veamos unos resultados interesantes que servirán para simplificar la expresión anterior.

Proposición 2.5.1 *Dado M una matriz simétrica de dimensión $(k \times k)$ y x un vector de dimensión k, entonces se tiene que:*

- $x^T M x = traza(x^T M x) = traza(M x x^T)$
- *traza(M)* $= \sum_{i=1}^{n} \lambda_i$ *donde cada* λ_i *es un valor propio de la matriz* M

Demostración: Tenemos para empezar que $x^T M x$ es un escalar, luego se tiene que $x^T M x = traza(x^T M x)$. Por otro lado, se puede demostrar que dadas dos matrices A, B con dimensión $(n \times k)$ y$(k \times n)$, se tiene la siguiente igualdad: $traza(AB) = traza(BA)$, entonces, si tomamos x^T como primera matriz y Ax como segunda matriz, aplicando esta propiedad se tiene que $traza(x^T M x) = traza(M x x^T)$. Para finalizar, si se aplica el teorema de descomposición espectral, se puede tener $M = P^T \Lambda P$ con $PP^T = I$ con Λ una matriz diagonal con los valores propios de M. Aplicando nuevamente la propiedad de $traza(AB) = traza(BA)$, tenemos que $traza(M) = traza(P^T \Lambda P) = traza(\Lambda P p^T) = traza(\Lambda) = \sum_{i=1}^{n} \lambda_i$.

Ahora, si aplicamos la proposición 2.5.1 a $(x_j - \mu)^T \Sigma^{-1}(-x_j - \mu)$, tenemos la igualdad siguiente:

$$(x_j - \mu)^T \Sigma^{-1}(-x_j - \mu) = traza(\Sigma^{-1}(x_j - \mu)(x_j - \mu)^T)$$

Además,

$$\sum_{j=1}^{n} (x_j - \mu)^T \Sigma^{-1}(-x_j - \mu) = \sum_{j=1}^{n} traza(\Sigma^{-1}(x_j - \mu)(x_j - \mu)^T) = traza(\Sigma^{-1} \sum_{j=1}^{n} (x_j - \mu)(x_j - \mu)^T)$$

El siguiente resultado es útil para definir la función de máxima verosimilitud de μ y Σ.

Proposición 2.5.2 *Sea $B \in \mathbb{R}^{p \times p}$ una matriz simétrica definida positiva y un escalar $b > 0$, entonces se tiene la desigualdad:*

$$\frac{1}{|\Sigma|^b} e^{-traza(\Sigma^{-1}B)/2} \leq \frac{1}{|B|^b} (2b)^{pb} e^{-bp}$$

Demostración: Tomamos $B^{1/2}$. tenemos que entonces $traza(\Sigma^{-1}B) = traza(\Sigma^{-1}B^{1/2}B^{1/2}) = traza(B^{1/2}\Sigma^{-1}B^{1/2})$. Podemos ver que la matriz $B^{1/2}\Sigma^{-1}B^{1/2}$ es simétrica definida positiva. Tomando un vector $t \in \mathbb{R}^p$ distinto del 0, entonces $t^T(\Sigma^{-1}B^{1/2}B^{1/2})t = (B^{1/2}t)^T\Sigma^{-1}B^{1/2}t > 0$ porque $B^{1/2}t$ es un vector distinto del 0 ya que t no es 0 y Σ^{-1} era definida positiva. Como es definida positiva, si η_i $i = 1, \ldots, p$ son los valores propios de $B^{1/2}\Sigma^{-1}B^{1/2}$ y tenemos:

$$traza(\Sigma^{-1}B) = traza(B^{1/2}\Sigma^{-1}B^{1/2}) = \sum_{j=1}^{p} \eta_j$$

por la proposición 2.5.1.
Además, por propiedades de los determinantes se tiene que $|B^{1/2}\Sigma^{-1}B^{1/2}| = \prod_{j=1}^{n} \eta_j$. Por otra parte también tenemos que $|B^{1/2}\Sigma^{-1}B^{1/2}| = |B^{1/2}||\Sigma^{-1}||B| = \frac{|B|}{|\Sigma|}$.

Para finalizar, se tiene que $\frac{1}{|\Sigma|} = \frac{|B^{1/2}\Sigma^{-1}B^{1/2}|}{|B|} = \frac{\prod_{j=1}^{n}\eta_j}{|B|}$ y con esto, se tiene que:

$$\frac{1}{|\Sigma|^b} e^{-\frac{\text{traza}(\Sigma^{-1}B)}{2}} = \left(\frac{\prod_{j=1}^{p}\eta_j}{|B|}\right)^b e^{-\frac{\sum_{j=1}^{p}\eta_j}{2}} = \frac{1}{|B|^b}\prod_{j=1}^{p}\eta_j^b e^{-\eta_j/2}$$

Si tenemos en cuenta la función de una variable real $f(\eta) = \eta^b e^{-1/2\eta}$, podemos calcular los máximos y mínimos. Derivando f, se tiene que: $f'(\eta) = b\eta^{b-1}e^{-\eta/2} + \eta^b(-1/2)e^{-1/2\eta}$. Si se iguala a 0 para obtener los puntos: $0 = e^{-\eta/2}\eta^{b-1}(b - \eta/2)$ De aquí obtenemos dos soluciones;

$$\eta = 0$$

$$b - \eta/2 = 0 \rightarrow \eta = 2b$$

De estos dos puntos, se observa que $\eta = 2b$ es el máximo. Si se sustituye, se obtiene que:

$$\frac{1}{|\Sigma|^b} e^{-\frac{\text{traza}(\Sigma^{-1}B)}{2}} = \frac{1}{|B|^b}\prod_{j=1}^{p}\eta_j^b e^{-\eta_j/2} \leq \frac{1}{|B|^b}(2b)^{pb}e^{-bp}$$

Para finalizar, se puede observar que hay una igualdad para el caso de que $\Sigma = \frac{1}{2b}B$. $B^{1/2}\Sigma B^{1/2} = B^{1/2}(2bB^{-1})B^{1/2} = 2bI$. Por tanto la $traza(B^{1/2}\Sigma B^{1/2}) = 2bp$. Además, $|B^{1/2}\Sigma^{-1}B^{1/2}| = |2bI| = (2b)^p$, luego $\frac{1}{|\Sigma|} = \frac{|B^{1/2}\Sigma^{-1}B^{1/2}|}{|B|} = \frac{(2b)^p}{|B|}$. Sustituyendo en $\frac{1}{|\Sigma|^b}e^{-\frac{\text{traza}(\Sigma^{-1}B)}{2}}$ las últimas igualdades de la traza y el determinante, se obtiene $\frac{1}{|B|^b}\prod_{j=1}^{p}\eta_j^b e^{-\eta_j/2} \leq \frac{1}{|B|^b}(2b)^{pb}e^{-bp}$.

Volviendo atrás, se tenía la expresión: $\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}e^{-1/2\sum_{j=1}^{n}(x_j-\mu)^T\Sigma^{-1}(-x_j-\mu)}$. Si nos fijamos en $\sum_{j=1}^{n}(x_j - \mu)^T\Sigma^{-1}(-x_j - \mu)$, se vio que aplicando la proposición 2.5.2 se obtenía:

$$\sum_{j=1}^{n}(x_j - \mu)^T\Sigma^{-1}(-x_j - \mu) = traza(\Sigma^{-1}\sum_{j=1}^{n}(x_j - \mu)(x_j - \mu)^T)$$

y fijándose únicamente en $\sum_{j=1}^{n}(x_j - \mu)(x_j - \mu)^T)$ sumamos y restamos $\overline{X} = \frac{1}{n}\sum_{j=1}^{n}x_j$, que es la media de la muestra inicial, se tiene que,

$$\sum_{j=1}^{n}(x_j - \mu)(x_j - \mu)^T = \sum_{j=1}^{n}(x_j - \overline{X} + \overline{X} - \mu)^T(x_j - \overline{X} + \overline{X} - \mu)$$

.

$$\sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T + \sum_{j=1}^{n}(\overline{X} - \mu)(\overline{X} - \mu)^T$$

$$= \sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T + n(\overline{X} - \mu)(\overline{X} - \mu)^T$$

Ahora, si se sustituye, en la función de densidad conjunta, se puede obtener la función:

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-traza(\Sigma^{-1}(\sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T + n(\overline{X} - \mu)(\overline{X} - \mu)^T))/2}$$

En esta función, la parte de la traza en el exponente se puede expresar de la siguiente manera, que será útil en un futuro:

$$traza(\Sigma^{-1}(\sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T + n(\overline{X} - \mu)(\overline{X} - \mu)^T))$$

$$traza(\Sigma^{-1}(\sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T)) + n\ traza(\Sigma^{-1}(\overline{X} - \mu)(\overline{X} - \mu)^T)$$

$$= traza(\Sigma^{-1}(\sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T)) + n(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu)$$

Ahora se verá cual es el máximo de $L(\mu, \Sigma)$.

Proposición 2.5.3 *Sean $X_1, \ldots X_n$ una muestra aleatoria de una normal de media μ y matriz de covarianza Σ. Se tiene que*

$$\hat{\mu} = \overline{X}\ \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})(X_j - \overline{X})^T = \frac{n-1}{n}S$$

Son los estimadores máximos verosímiles de μ y Σ respectivamente.

Demostración: Teniendo en cuenta una parte del exponente en la función $L(\mu, \Sigma)$,

$$L(\mu, \Sigma) = traza(\Sigma^{-1}(\sum_{j=1}^{n}(x_j - \overline{X})(x_j - \overline{X})^T)) + n(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu)$$

Tenemos que como Σ es definida positiva, entonces Σ^{-1} también lo es, por tanto para que $(\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu) = 0$, se debe tener $\overline{X} - \mu = 0$. A si que $\hat{\mu} = \overline{X}$. Solo se debe maximizar

$$L(\hat{\mu}, \Sigma) = \frac{1}{(2\pi)^n p/2|\Sigma|} e^{-traza(\Sigma^{-1}(\sum_{j=1}^{n}(X_j - \overline{X})(X_j - \overline{X})^T))/2}$$

Con la proposición 2.5.2, si se aplica con $b = n/2$ y $B = \sum_{j=1}^{n}(X_j - \overline{X})(X_j - \overline{X})^T$, el máximo es en el caso de $\hat{\Sigma} = 1/n\sum_{j=1}^{n}(X_j - \overline{X})(X_j - \overline{X})^T$

Por último, se tiene que

$$L(\hat{\mu}, \hat{\Sigma}) = \frac{1}{(2\pi)^{np/2}} e^{-np/2} \frac{1}{|\hat{\Sigma}|^{n/2}}$$

Si se sustituye usando las proposiciones 2.5.2 y 2.5.3.

Definición 2.5.1 *Dada una muestra aleatoria simple $X = (x_1, \ldots, x_n)^T$, un vector de parámetros, θ, con hipótesis $H_0 : \theta \in \Sigma_0$ y $H_1 : \theta \in \Sigma_1$ Se llama el test del ratio de verosimilitud del test de H_0 frente a H_1 a:*

$$\lambda = \frac{\text{máx}\, L_0(\theta)}{\text{máx}\, L_a(\theta)}$$

Donde $L_i(\theta)$ está definido en la región θ_i, $i = 0, 1$

3 Análisis de la varianza con MANOVA

Empezaremos por un análisis clásico de la varianza, el MANOVA. Para esta sección, se ha empleado la teoría anteriormente vista. También se ha extraído el algoritmo con el cual se hace el MANOVA de unos apuntes de la universidad estatal de Pensilvania (EEUU),[6], [7]
En el análisis del ANOVA, se puede crear una tabla donde aparezcan las sumas de las diferencias de cuadrados de los tratamientos y errores. El análisis MANOVA busca extender el proceso de creación de la tabla que se hacía en el ANOVA y, también, extender el test F en el caso multivariante. Se vio en la sección anterior que la distribución de Wishart es una generalización de la distribución χ^2. También se vio el test de máxima verosimilitud y la distribución Λ de Wilks, que se podría decir que tiene un estilo "parecido" al test F.
Por tanto, el algoritmo que se muestra a continuación buscan extender esa suma de las diferencias de los cuadrados del ANOVA y extender el test usado.

3.1. Algoritmo

Empezamos con algo de notación. Sea $Y_{i,j,k}$ la observación de la variable k del grupo i del sujeto j. Entonces $Y_{i,j} = (Y_{i,j,1}, \ldots, Y_{i,j,d})$ es el vector de variables para el sujeto j en el grupo i. n_i es el número de observaciones del grupo i. $N = n_1 + \ldots + n_m$. Suponemos que los datos del grupo i tienen una media común $\mu_i = (\mu_{i,1}, \ldots, \mu_{i,d})$. Los datos de todos los grupos deben tener la misma matriz de covarianza. La muestra es una muestra aleatoria y los datos tienen una distribución normal multivariada, (aunque si la muestra es suficientemente grande, apelando a la ley fuerte de los grandes números, esta suposición no tiene por qué cumplirse completamente).
Las hipótesis iniciales para aplicar este modelo son las mismas que se mostraron en la introducción, pero se vuelven a escribir pero de una forma genérica para tenerlas en mente:

- $H_0 : \mu_1 = \ldots = \mu_m$
- $H_a : \mu_{ik} \neq \mu_{jk}$ para algún i=j y al menos un k.

Definimos como el vector de la media de la muestra del grupo i como: $\overline{y_{i.}} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$

Definimos también el vector de media grande como:
$\overline{y_{i..}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} Y_{ij}$. Ahora, habiendo definido algunas cosas de una forma parecida al ANOVA, en el MANOVA también se hacen las sumas de cuadrados y productos cruzados, con la diferencia que en el análisis univariado estas sumas son escalares mientras que en el MANOVA son matrices.
La suma de los cuadrados total es:

$$T = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \overline{y_{..}})(Y_{ij} - \overline{y_{..}})'$$

Esta suma puede dividirse de la siguiente manera:

$$T = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (Y_{ij} - \overline{y_{..}})(Y_{ij} - \overline{y_{..}})' = \sum_{i=1}^{m} \sum_{j=1}^{n_i} ((Y_{ij} - \overline{y_{i.}}) + (\overline{y_{i.}} - \overline{y_{..}}))((Y_{ij} - \overline{y_{i.}}) + (\overline{y_{i.}} - \overline{y_{..}}))' =$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(Y_{ij}-\overline{y_{i\cdot}})(Y_{ij}-\overline{y_{i\cdot}})'+\sum_{i=1}^{m}n_i(\overline{y_{i.}}-\overline{y_{..}})(\overline{y_{i.}}-\overline{y_{..}})'=E+H$$

Donde E es el error de la suma de los cuadrados y productos cruzados y H es la hipotética suma de los cuadrados y productos cruzados. Por tanto, con todos estos datos podemos hacer esta tabla:

	Grados de libertad	Suma de cuadrados y productos cruzados
Tratamientos	$m-1$	H
Error	$N-m$	E
Total	$N-1$	T

Cuadro 3.1: Tabla del MANOVA.

Volviendo a las hipótesis, desearíamos tener un test del estadístico para el MANOVA. Tenemos que $T = H + E$. En el caso de que H sea grande en comparación con E, se tiene que las diferencias de las sumas de los cuadrados y productos cruzados de los tratamientos son grandes, por lo que las diferencias entre grupos son grandes y, por tanto, se debería rechazar la hipótesis inicial.

3.2. Métodos de Contraste en MANOVA

En este apartado, se discutirán cuatro pruebas estadísticas comunes utilizadas en el análisis multivariado de la varianza (MANOVA) para contrastar la hipótesis nula. Los métodos considerados son: la Lambda de Wilks, la Traza de Hotelling-Lawley, la Traza de Pillai y la Raíz Máxima de Roy. A continuación, se presenta una descripción detallada de cada uno de estos test y sus implicaciones en la toma de decisiones.

1. **Lambda de Wilks**

   Recordando que los datos $Y_{i,j,k}$ provienen de una distribución normal, se tiene que, si se cumple la hipótesis nula, $E \sim W_d(\Sigma, N-m)$ y $H \sim W_d(\Sigma, m-1)$, donde $W(\Sigma, \cdot)$ denota la distribución de Wishart con parámetro de covarianza Σ. Además, según el teorema 2.2.2, se cumple que $E + H \sim W_d(\Sigma, N-1)$.

   El test de razón de verosimilitud se basa en comparar los determinantes de las matrices H y $T = E+H$. Específicamente, se puede comprobar que el estadístico $\frac{|E|}{|E+H|}$, sigue una distribución $\Lambda(d, N-m, N-1)$, pues H y E provienen de la distribución de Wishart, se tiene que $\frac{E}{E+H} \sim \Lambda(d, m-1, N-1)$.
   Por último, si este cociente es pequeño, se debe rechazar la hipótesis nula.

   En la definición 2.3.2, se muestra que $\frac{|E|}{|E+H|}$ es equivalente a $|I + HE^{-1}|$. Este determinante puede calcularse como el producto de los valores propios, $\prod_{j=1}^{N}\phi_j$, donde ϕ_j son los valores propios de $I + HE^{-1}$. Estos valores propios se pueden expresar como $\phi_j = 1 + \lambda_j$, donde λ_j es un valor propio de HE^{-1}. Por lo tanto, el producto $\prod_{j=1}^{N}\phi_j$ depende directamente de los valores propios de HE^{-1}, determinando si el estadístico es grande o pequeño.

2. **Traza de Hotelling-Lawley**

   La estadística de la Traza de Hotelling-Lawley se define como $T_0^2 = \text{traza}(HE^{-1})$. La traza de una matriz es la suma de sus valores propios, por lo que en este caso, $\text{traza}(HE^{-1}) = \sum_{j=1}^{N}\lambda_j$. En el test de Lambda de Wilks, se ha visto que los valores propios de $I + HE^{-1}$ dependen de los de HE^{-1}, por lo que resulta razonable estudiar la suma de estos valores.

   Si H es grande en comparación con E, la traza tomará un valor grande, lo que llevará al rechazo de la hipótesis nula. Este test es útil para detectar diferencias significativas cuando la suma de las contribuciones individuales es considerable.

3. **Traza de Pillai**

   La estadística de la Traza de Pillai se define como $V = \text{traza}(H(H+E)^{-1})$. Similar al método anterior, esta prueba se basa en la suma de los valores propios de la matriz $H(H+E)^{-1}$. Si H es grande en relación con E, entonces $H+E$ también lo será, y la traza será correspondientemente grande.

   Este test es conocido por su robustez, especialmente en situaciones donde las suposiciones de normalidad del MANOVA no se cumplen completamente. Esta robustez lo convierte en una opción preferible en análisis donde la normalidad de los datos es cuestionable.

4. **Raíz Máxima de Roy**

   La estadística de la Raíz Máxima de Roy se basa en el valor propio más grande de HE^{-1}. En el test de Lambda de Wilks, se observó que el estadístico se relaciona con el producto de los valores propios de HE^{-1}. Por lo tanto, es razonable considerar el valor propio más grande para estimar la magnitud de dicho producto.

   En este caso, el estadístico del test es el mayor de los valores propios de HE^{-1}. Si H es grande en comparación con E, entonces el valor propio más grande de HE^{-1} también será grande, lo que indicará la necesidad de rechazar la hipótesis nula.

   Cabe destacar que los estadísticos mencionados anteriormente no siempre cuentan con tablas para el cálculo directo del p-valor, por lo que es común recurrir a aproximaciones mediante la distribución F. Esta aproximación necesita de un número considerable de observaciones de la muestra original para que sean precisas. En la página web de SAS Institute Inc., se pueden encontrar diferentes aproximaciones que facilitan este proceso (SAS Institute Inc. [8]).

4 Análisis de la varianza con FANOVA

En esta sección se verá desde otra perspectiva el análisis del problema inicial. El análisis funcional es un campo que se ha estado investigando mucho. Vidakovic, Brani. "Wavelets for Kids: A Tutorial Introduction", [1], donde explica las wavelets, unas funciones con las que se puede tener una base funcional y que es usado para guardar videos, eliminacion de ruido en audio y análisis fnancieros entre muchos otros campos. También documentos orientados al análisis estadístico funcional como [2], de Aneiros Pérez, Gemma y otros investigadores, libros como "Inference for Functional Data with Applications" [13] de Horváth, L., & Kokoszka, P. y libros como Ferraty, F., & Vieu, P. (2006). "Nonparametric Functional Data Analysis: Theory and Practice" [24], o las investigaciones y revisiones de Cuevas junto con otros investigadores, [12], [26], [27].
El análisis de la varianza funcional, es una técnica que se ha ido consolidando este último siglo. En este método, se plantea un análisis parecido al ANOVA clásico, pero los datos son funciones reales. Es decir, el análisis se hace en un conjunto infinito de puntos t de unas funciones $Y_{ij}(t)$. Pero cuando se toma una muestra, se deben tomar un conjunto finito de valores t.
Como una forma preliminar, en este análisis con datos funcionales, se puede aplicar el ANOVA para cada t, pero esta idea es poco escalable por su problema de multiplicidad. Por otro lado, para acabar de hacer el test de hipótesis, se debe usar un método de combinación de p-valores, aumentando el error de tipo I, es decir, se rechaza en situaciones donde no debería rechazarse. Otra técnica sería aplicar el MANOVA que se comentó anteriormente, pero nuevamente, no es buena idea con grandes dimensiones de los datos, que es lo común en el caso funcional. Por tanto, se han buscado numerosos métodos alternativos para este problema.
En este análisis de la varianza, normalmente para simplificar el problema, es común usar un sistema de bases. Una base funcional se puede definir como:

Definición 4.0.1 *Un sistema de funciones ϕ_i se llaman base si para cualquier función f, f se puede escribir como combinación lineal de funciones ϕ_i.*

$$f = \sum_{i=1}^{\infty} c_i \phi_i$$

Y se tiene que todos los ϕ_i son linealmente independientes y los c_i son los coeficientes.

Se tiene por tanto que cualquier función f puede aproximarse por una combinación lineal con un conjunto finito de estas funciones ϕ_i, todas ellas provenientes de una base.
Algunas bases conocidas son las compuestas por monomios, polinomios, la base de Fourier, b-splines, bases con funciones exponenciales o funciones escalón.
El uso principal de usar un sistema de bases es hacer un suavizado inicial de los datos, porque las funciones de algunas bases son suaves y la combinación lineal de algunas de ellas conservan esta propiedad y también se busca simplificar el problema, ya que si el conjunto de datos es denso, se puede reducir la dimensionalidad gracias a la base y también, se puede reducir al estudio de las funciones o coeficientes de la base.

Para adaptar la notación del problema inicial al caso funcional y que se adecúe más a la notación esperada de las funciones, tendremos $Y_{ij}(t)$, que es la muestra j-ésima del grupo i, en vez de Y_{ijk}, que es la notación antigua que se usaba para el caso no funcional. Tendremos también que $j = 1, \ldots, n_i$ y

$i = 1, \ldots, K$. Además, ahora tendremos, $t \in T$ en vez de $k = 1, \ldots, d$.
También se podría modificar la hipótesis nula para que se adecúe al contexto de funcional: La media del grupo i será $\mu_i(t) = \frac{1}{n_1}\sum_{j=1}^{n_i} Y_{ij}(t)$, entonces la hipótesis nula es que $\mu_i(t) = \mu_k(t)\ \forall i, k \in K, i \neq k$. La hipótesis alternativa a su vez será la negación de la nula.
Como se mencionó anteriormente en la introducción de la memoria y en la de este capítulo, existe una gran cantidad de información sobre estadística funcional y concretamente sobre el análisis de la varianza funcional. Entonces, en este capítulo se ha decidido estudiar únicamente algunos de los test que existen. Estos test que se van a analizar serán los test que aparecen en la librería de R "fdANOVA"[3], creada por Lukasz Smaga y Tomasz Górecki, de la universidad de Adam Mickiewicz (Polonia). Esta librería es la que más adelante se usará para hacer las simulaciones y por eso se ha decidido por limitar la explicación de estos test a los de esta librería. La documentación de esta librería, junto con las referencias que contiene, se han usado para la exposición de estos test en esta memoria. Por último, los test disponibles en esta librería son los siguientes:

- Test FP
- Test CH
- Test CS (Cuadros Integrados)
- L2N, L2B, L2b (Normas L2)
- Test funcional de la norma (FN)
- Bootstrap funcional (FB, Fb)
- GPF
- Fmaxb
- TRP_{anova}, TRP_{ATS}, TRP_{WTPS}

4.1. Test FP

En este test se busca una base ortogonal para las funciones aleatorias $Y_{ij}(t)$ de modo que:

$$Y_{i,j}(t) = \sum_{m=1}^{K} \beta_{ijm}\phi_m(t) + \epsilon_{ij}(t).$$

Donde β_{ij} es el coeficiente j-ésimo de la función i, $\phi_m(t)$ son las funciones de la base y $\epsilon_{ij}(t)$ es un error cometido. Para encontrar los coeficientes de la base y, además reducir la dimensionalidad sin perder mucha información sobre los datos originales, se hace un análisis de componentes principales como el visto en clase de Advanced Statistics. Por tanto, si α_{ijk} y ζ_k son los coeficientes y las componentes principales, los coeficientes β_{ijm} serán: $\beta_{ijm} = \sum_{k=1}^{M} \alpha_{ijk}\zeta_k$. Tras obtener los coeficientes de la base, se procede a usarlos para un análisis ANOVA clásico: se calculan las sumas de cuadrados y se obtiene el estadístico F, con el que se rechazará o no la hipótesis inicial.
La motivación de este ANOVA es que, los coeficientes que se nombran, si se tiene que existen diferencias significativas entre los distintos grupos de coeficientes, esas mismas diferencias se verán reflejadas en las funciones a analizar.

4.2. Test CH

En este test, se busca nuevamente una base para los datos funcionales; $Y_{i,j}(t) = \sum_{m=1}^{K} \beta_{ijm}\phi_m(t) + \epsilon_{ij}(t)$. Se hace un análisis de componentes principales para obtener los coeficientes α_{ijm}, que provienen de los coeficientes β_{ijm} que al ser proyectados en las componentes principales dan lugar a estos nuevos coeficientes α_{ijm}. Finalmente, se calcula el estadístico como:

$$T = traza\left(\frac{B}{W}\right)$$

Donde $B = \sum_{i=1}^{K} n_i(\alpha_i - \alpha)(\alpha_i - \alpha)^T$ y $W = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(\alpha_{ij} - \alpha_i)(\alpha_{ij} - \alpha_i)^T$ Donde $\alpha_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\alpha_{ij}$ y $\alpha = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\alpha_{ij}$.
Para decidir si se rechaza o no, es necesario una distribución nula, que se obtiene con métodos de remuestreo y cálculos del estadístico mencionado con las nuevas muestras.

4.3. Test CS

En este método, nuevamente se busca obtener los valores $Y_{ij}(t)$ como combinación lineal de elementos de una base ortogonal. Se calculan las medias de cada grupo i, $\mu_i(t) = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}(t)$. Se calcula la matriz de distancias donde la entrada $d_{ij} = \left(\int_T (\mu_i(t) - \mu_j(t))^2 dt\right)^{1/2}$. Por último, se considera el estadístico de pruebas como: $CS = max_{ij} d_{ij}$. Para determinar la significancia de este estadístico calculado, se procede a hacer una distribución nula mediante simulaciones o técnicas de remuestreo.

4.4. Test de las normas L_2

Estos test tienen cosas en común, como el uso de la norma L_2 y que en todos se busca un enfoque basado en datos que son combinación lineal de una base ortogonal. Para repasar la notación, la norma L_2 de la diferencia de dos funciones f(t) y g(t) es:

$$||f(t) - g(t)||_2 = \left(\int_T (f(t) - g(t))^2\right)^{1/2}$$

Los tres test que se tienen implementados en la librería son:

4.4.1. L_{2N}

En este test el estadístico es:

$$T = \sum_{i<j} ||\mu_i(t) - \mu_j(t)||^2_{L_2}$$

Donde para cada grupo i, $\mu_i(t) = \sum_{j=1}^{n_i} Y_{ij}(t)$.
Nuevamente, para obtener la distribución nula y poder estudiar el nivel de significancia del estadístico y poder rechazar, se procede a un remuestreo aleatorio bajo la hipótesis nula, con el que se obtienen un número significativo de estadísticos con lo que poder rechazar o no.

4.4.2. L_{2B}

En este test, se definen un conjunto de M bandas, B_m $(m = 1, \ldots, M)$ en el dominio de t. Una banda es un intervalo dentro de este dominio.

Se calcula la media funcional de cada grupo i en cada banda m. $\mu_i^{(m)}(t) = \sum_{j=1}^{n_i} Y_{ij}(t)\ t \in B_m$. El estadístico usado en este caso es:

$$T = \sum_m \sum_{i<j} ||\mu_i(t) - \mu_j(t)||^2_{L_2(B_m)}.$$

Donde $||\mu_i(t) - \mu_j(t)||_{L_2(B_m)} = \left(\int_{B_m} (f(t) - g(t))^2\right)^{1/2}$ Posteriormente se crea la distribución nula como en anteriores test para luego decidir si se puede rechazar.

4.4.3. L_{2b}

En este test se calcula el estadístico

$$S = \int_T \sum_{i=1}^{k} n_i (Y_{ij}(t) - \mu(t))^2.$$

Posteriormente, por medio del bootstrap, se crea una distribución que siga la hipótesis nula.

4.5. FN, FB, Fb

Estos tres test son similares a los anteriores. Son casos donde la base para representar los datos es una base de Fourier. El test FN tendría el mismo estadístico que el test L2N, el FB, igual estadístico que el L2B y el Fb, tendría el mismo que el L2b. La base de Fourier puede ser de utilidad en ciertos casos donde se sabe que los datos de la muestra provienen de funciones suaves. Con una base de Fourier se tiene una mayor suavidad.

4.6. GPF

En este test el estadístico se calcula como:

$$SSB(t) = \sum_{i=1}^{K} n_i (\mu_i(t) - \mu(t))^2$$

$$SSW(t) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} (Y_{ij}(t) - \mu_i(t))^2$$

$$F = \int_T F(t)dt = \int_T \frac{SSB(t)(N-k)}{SSW(t)(k-1)}.$$

Para saber si se rechaza o no, se crea una distribución nula.

4.7. F_{max_b}

Para este test, primero hay que calcular

$$SSB(t) = \sum_{i=1}^{K} n_i (\mu_i(t) - \mu(t))^2$$

$$SSW(t) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} (Y_{ij}(t) - \mu_i(t))^2$$

$$F = \int_T F(t)dt = \int_T \frac{SSB(t)(N-k)}{SSW(t)(k-1)}.$$

En este test el estadístico es

$$S = \max_T F(t)$$

y para decidir si se rechaza o no, se requiere de una distribución nula que se genera de manera similar a otros test.

4.8. TRP

Estos tres test mantienen una misma filosofía; proyectar los datos originales; se eligen M proyecciones aleatorias para proyectar los datos y con esas proyecciones, calcular el estadístico y posteriormente combinar los p-valores. La fórmula para calcular las proyecciones es la siguiente:

$$P_{ij}^r(t) = \frac{\int_I Y_{ij}(t)h_r(t)dt}{\left(\int_T h_r^2(t)dt\right)^{1/2}}$$

Donde $h_r(t)$ es una función aleatoria que se elige para proyectar los datos y $r = 1, \ldots, M$. Estas funciones pueden obtenerse de diversas maneras, como por ejemplo por un puente Browniano o distribuciones Gaussianas u otros procesos estocásticos.

4.8.1. TRP_{anova}

Para este test, una vez se tienen las proyecciones, para cada proyección r se hace un análisis ANOVA, se obtiene un p-valor, p_r y posteriormente se combinan para tomar una decisión.

4.8.2. TRP_{ATS}

Para cada proyección se calcula el estadístico ATS, un test no paramétrico, basado en rangos alineados. Finalmente, se combinan los p-valores para decidir si se rechaza o no.

4.8.3. TRP_{WTPS}

En este test, se calcula un test para cada proyección r para evaluar si existen diferencias significativas (ej, ANOVA), posteriormente se realiza el cálculo del estadístico del test T_r.
Finalmente, se calcula el estadístico final de este test,

$$S = \sum_{r=1}^{M} T_r w_r$$

Donde w_r representa el peso que tiene el estadístico de la proyección r-ésima. Estos pesos pueden ser todos iguales o asignarse con diferentes reglas. Por último, se crea una distribución nula por métodos de remuestreo para decidir si se rechaza o no.

5 Otro análisis de la varianza

En esta sección se mostrará otro método que usará un doble bootstrap para hacer el test de hipótesis. Este algoritmo se ha diseñado para el caso de dos grupos, por lo que, se tendrán solamente las funciones aleatorias $Y_{1,j}$ e $Y_{2,j}$. El algoritmo es el siguiente:

Descripción del Algoritmo

Expresamos estos conjuntos como matrices:

$$Y_1 = \begin{pmatrix} Y_{1,1,1} & Y_{1,1,2} & \cdots & Y_{1,1,d} \\ Y_{1,2,1} & Y_{1,2,2} & \cdots & Y_{1,2,d} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{1,n_1,1} & Y_{1,n_1,2} & \cdots & Y_{1,n_1,d} \end{pmatrix} \in \mathbb{R}^{N\times d}$$

$$Y_2 = \begin{pmatrix} Y_{2,1,1} & Y_{2,1,2} & \cdots & Y_{2,1,d} \\ Y_{2,2,1} & Y_{2,2,2} & \cdots & Y_{2,2,d} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{2,n_2,1} & Y_{2,n_2,2} & \cdots & Y_{2,n_2,d} \end{pmatrix} \in \mathbb{R}^{M\times d}$$

Cada fila es un elemento de los conjuntos respectivos.

Definiciones y Cálculos Iniciales

Primero, se puede reescribir $\overline{Y_i}$ como:
$\overline{Y_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij1},\ldots,Y_{ijd}) = (\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{i,j,1},\ldots,\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ijd}) = (\overline{Y_{i1}},\ldots,\overline{Y_{id}})$

- **Definición del Conjunto I_c:**

$$I_c := \left\{k : |\overline{Y_{1,k}} - \overline{Y_{2,k}}| > c\right\}$$

- **Cálculo de ϵ_c:**

$$\epsilon_c := \int_{I_c} |\overline{Y_{1,k}} - \overline{Y_{2,k}}|\, dk\, \frac{1}{\lambda(I_c)}$$

  donde $\lambda(I_c)$ es la medida del conjunto I_c.

Procedimiento de Bootstrap

Primer Nivel de Bootstrap

1. Generar 999 muestras bootstrap de las matrices Y_1 e Y_2 con reemplazamiento de $Y_1 - \overline{Y_1}$ e $Y_2 - \overline{Y_2}$, obteniendo:

$$Y_1^1, Y_1^2, \ldots, Y_1^{999} \quad \text{e} \quad Y_2^1, Y_2^2, \ldots, Y_2^{999}$$

2. Para cada muestra $j = 1,\ldots,999$:

   - **Definir I_c^j:**

$$I_c^j := \left\{k : |\overline{Y_{1k}^j} - \overline{Y_{2k}^j}| > c\right\}$$

- **Calcular** ϵ_c^j:

$$\epsilon_c^j := \int_{I_c^j} |\overline{Y_{1k}^j} - \overline{Y_{2k}^j}|\, dt \, \frac{1}{\lambda(I_c^j)}$$

Segundo Nivel de Bootstrap

1. Para cada $j = 1, \ldots, 999$:

   - Generar 999 muestras bootstrap con reemplazamiento de $Y_1^j - \overline{Y_1^j}$ e $Y_2^j - \overline{Y_2^j}$, obteniendo:

$$Y_1^{j,1}, Y_1^{j,2}, \ldots, Y_1^{j,999} \quad \text{e} \quad Y_2^{j,1}, Y_2^{j,2}, \ldots, Y_2^{j,999}$$

   - Para cada muestra $i = 1, \ldots, 999$:
     - **Definir** I_c^{ji}:

$$I_c^{ji} := \left\{ k : |\overline{Y_{1k}^{ji}} - \overline{Y_{2k}^{ji}}| > c \right\}$$

     - **Calcular** ϵ_c^{ji}:

$$\epsilon_c^{ji} := \int_{I_c^{ji}} |\overline{Y_{1k}^{ji}} - \overline{Y_{2k}^{ji}}|\, dt \, \frac{1}{\lambda(I_c^{ji})}$$

2. Ordenar los ϵ_c^{ji} para cada j y seleccionar el valor en la posición 950 para obtener $\epsilon_c^{j,(950)}$.

Función de Maximización

1. Definir la función $R(c)$:

$$R(c) = \sum_{j=1}^{999} I\left(\epsilon_c^j > \epsilon_c^{j,(950)}\right)$$

   donde I es la función indicadora.

2. Maximizar $R(c)$ para obtener un vector $C \in \mathbb{R}^{999}$ y calcular los ϵ_c^j.

Prueba de Hipótesis

1. Con el C calculado anteriormente, calcular y ordenar los ϵ_c^j de menor a mayor y seleccionar el valor en la posición 950 para obtener $\epsilon_C^{(950)}$.

2. Comparar ϵ_C con $\epsilon_C^{(950)}$:

   - Si $\epsilon_{C(950)} > \epsilon_C^{(950)}$, rechazar la hipótesis inicial.
   - En caso contrario, no hay evidencia suficiente para rechazar la hipótesis.

Una vez explicado el algoritmo, se puede hablar de su complejidad temporal. Se tiene que el coste temporal de hacer un bootstrap es de $O(na)$, donde n es el número de muestras bootstrap y a el tamaño de la muestra inicial. Por tanto, suponiendo que Y_1 e Y_2 tengan ambos a elementos, se tiene que el primer bootstrap tiene una complejidad del orden de $O(nN)$. Para el segundo bootstrap, se tendrá una complejidad de $O(nmN)$, siendo m el numero de muestras bootstrap del segundo bootstrap. También hay que tener en cuenta los cálculos de $\epsilon_c, \epsilon_c^j, \epsilon_c^{j,i}$ y sus ordenaciones. Por ejemplo, si el coste temporal de calcular ϵ_c es $O(\theta)$, el coste de calcular todos los ϵ_c^j y $\epsilon_c^{j,i}$ es $O(\theta n)$ y $O(\theta nm)$ respectivamente, pues simplemente es aplicar el mismo algoritmo de integración de ϵ_c tantas veces como muestras bootstrap se tengan. Si para ordenar los $\epsilon_c^{j,i}$ se usa el algoritmo "Quick Sort", cuya complejidad temporal media es $O(x\log(x))$, entonces obtener todos los $\epsilon_c^{j,(950)}$ equivale a ordenar n vectores de tamaño m, por lo

que la complejidad es de $O(nm\log(m))$. Por lo que evaluar una única vez la función $R(c)$ tiene un coste temporal de $O(\eta) = \text{máx}\{O(\theta n), O(\theta nm), O(nm\log(m))\}$. Por tanto, el algoritmo tendrá una complejidad temporal de $\text{máx}\{O(mna), O(k\eta)\}$, donde k es el número de evaluaciones de funciones que se han necesitado para encontrar el máximo.

Por último, hay que estudiar cual sería un método bueno de optimización para la función R(c), por tanto se deben estudiar las propiedades de esta función. Primeramente, por cómo se define, el dominio de la función es $[0, \infty)$ y su imagen estará en el conjunto $\{z \in \mathbb{Z} : \ 0 \leq z \leq 999\}$.
Luego se puede observar que la función solo podría ser continua si es constante. Si se fijan las muestras iniciales Y_1 e Y_2, se tiene que para un c suficientemente grande, los conjuntos I_c^j y I_c^{ji} van a ser vacíos, por lo que $\epsilon_c^j = \epsilon_c^{ji} = 0$. Entonces si esta función es continua, solamente puede ser la función constante 0. Se tiene que esto no será posible porque aun estando en las condiciones de la hipótesis nula, la función R(c) tomará valores positivos para algún c por causa del error de tipo I del test. Por tanto R(c) no será continua y por tanto tampoco derivable.

Volviendo a la optimización de la función R(c), se puede buscar el máximo, probando valores en una malla y eligiendo el máximo entre esos valores. Este método nos da un máximo entre todos los valores posibles de la malla, pero no garantiza que sea el máximo global real ni tampoco que sea un número cercano cuando la malla tiene pocos valores. A cambio, el método va a encontrar un punto que sea máximo entre esos valores de una forma rápida. Si por el contrario se tiene una malla con muchos valores, este proceso puede ser muy lento. Para que este método sea más rápido, se pueden hacer varias mejoras aprovechando las características de la función que se está optimizando. Una de ellas sería pensar en condiciones de parada.

1. Si dado un punto η de la malla se tiene que $R(\eta) = 999$, se ha encontrado el máximo y el método puede parar.

2. Si dado un punto γ de la malla se tiene que $R(\gamma) = 0$, se ha encontrado un punto donde los conjuntos son vacíos, por tanto los siguientes puntos también tendrán conjuntos vacíos y el método no va a encontrar el máximo para $c > \gamma$. Se puede parar el método.

Esta segunda condición se podría restringir más, por ejemplo parando el método cuando el cociente entre el valor de la función en el c actual y el valor máximo es inferior a una tolerancia dada.
$\frac{R(c_{\max})}{R(c_{\text{actual}})} < \text{tol}$. Esta propuesta solo es útil para reducir el tiempo de ejecución, pero su coste es perder exactitud en la solución, por lo que no se ha usado en los experimentos y solo se nombra. Un algoritmo que optimice de esta manera puede ser:

Algorithm 1 Optimización

m Máximo valor de $R(x)$ y el correspondiente x máximo $\leftarrow -1$
Xmax $\leftarrow -1$

for $x \leftarrow 0$ m k **do** $r \leftarrow R(x)$
  if $r >$ maximo **then** maximo $\leftarrow r$ Xmax $\leftarrow x$
    if $r == 0$ **then parar**
      if $r == 999$ **then parar**

Otro método que puede servir para optimizar es usar un algoritmo genético. Los algoritmos genéticos son una técnica de optimización basada en principios de evolución natural que han demostrado ser altamente efectivos en una variedad de problemas. No obstante, dado que el análisis exhaustivo de esta técnica requeriría una extensión considerable, excediendo los límites de este trabajo, se ha optado por únicamente introducir este tema de forma superficial.

5.1. Algoritmos genéticos

Como se empezó a introducir anteriormente, un algoritmo genético es un algoritmo de búsqueda y optimización basado en la evolución natural. Se basa en la teoría de optimización y probabilidad para buscar el punto óptimo en un espacio de soluciones complejas. Este algoritmo fue introducido por John Holland en el año 1975 [15], con el fin de buscar un método de optimización con espacios de soluciones complejas y de grandes dimensiones. Para seguir con el tema se deben introducir ciertas definiciones generales:

Definición 5.1.1 *Sea $X \in \mathbb{X}^n$, donde el conjunto $\mathbb{X}$ es un grupo. Para un algoritmo genético, se llama a X cromosoma si es una posible solución al problema de optimización. Además, si $X = (x_1, \ldots, x_n)$, entonces cada x_i, $i = 1, \ldots, n$ se llama gen.*
Normalmente el conjunto $\mathbb{X}$ es el conjunto de los reales, números binarios o enteros.

Definición 5.1.2 *Se llama población de la generación t al vector de cromosomas $P(t) = (X_1, \ldots, X_m)$.*

Definición 5.1.3 *Se llama función de fitness a la función $f(X)$ que evalúa la calidad del cromosoma X.*

Definición 5.1.4 *Un cruce de dos cromosomas X e Y es una operación donde se combinan los genes de ambos cromosomas de cierta manera para obtener un cromosoma hijo de X e Y. Algunas maneras de hacer este cruce para $X = (x_1, \ldots, x_n)$ y $Y = (y_1, \ldots, y_n)$ son:*

1. *Cruce en un punto;*
   Se fija un punto i en ambos vectores, el hijo será $(x_1, \ldots, x_i, y_{i+1}, \ldots, y_n)$
2. *Cruce en dos puntos;*
   Se fijan los puntos i,j en ambos vectores. El hijo será entonces: $(x_1, \ldots, x_i, y_{i+1}, \ldots, y_j, x_{j+1}, \ldots, x_n)$
3. *Cruce uniforme: Se eligen los genes del hijo de forma aleatoria con una probabilidad p (normalmente 0.5) entre los genes de los padres. Es decir, el gen i-ésimo del hijo será x_i con probabilidad p*
4. *Cruce en N puntos; Similar al cruce en dos puntos.*

Además de estas maneras propuestas aquí, hay muchas más en la literatura.

Definición 5.1.5 *Se llama mutación a un proceso aleatorio donde al cromosoma X se le aplica una modificación.*

Definición 5.1.6 *El proceso de selección de los padres consiste en una función estocástica basada en la función fitness. Algunos procesos de selección pueden elegir a los individuos de la población con una probabilidad proporcional a su valor en la función. Otros eligen un subconjuntos de la población y toman el cromosoma con un fitness máximo.*

El modelo matemático es el siguiente: Para un número T de iteraciones fijo.

1. Generar $P(0)$
2. Para i desde 0 hasta T hacer:
   a) evaluar cada cromosoma de la generación $P(i)$.
   b) Seleccionar los mejores candidatos $P'(i)$ mediante una función estocástica de selección.
   c) Se aplica cruce y mutación para obtener $P''(i)$

d) Se reemplaza $P(i+1) = P''(i)$

e) Parar si se cumple el criterio de parada

Los test de parada en los algoritmos genéticos están normalmente relacionados con la evolución de las iteraciones o el número máximo de iteraciones. Por tanto, se pueden tener los siguientes test:

- Parar si supera un máximo de iteraciones
- Parar si $\text{máx}\{f(x) : x \in P(t)\} \geq F_{\text{objetivo}}$ (parada por objetivo)
- Parar si $\frac{\sigma_f(t)}{\overline{f(t)}} \leq \epsilon$. Donde $\sigma_f(t)$ es la desviación estándar del fitness, $\overline{f(t)}$ es el fitness promedio y ϵ un valor límite, (generalmente cercano a 0). (Parada por convergencia poblacional).
- Parar si se supera el tiempo de ejecución permitido
- Parar si el fitness máximo de la generación actual es el mismo que el fitness máximo de todas las anteriores k poblaciones. Se tiene que k es un número mmáximo de poblaciones donde se permite este estancamiento. (Parada por estancamiento)

Normalmente estos test de parada se usan combinados, por lo que el algoritmo tiene en cuenta varios test de parada a la vez.
Un teorema importante en este campo es el teorema del esquema. Para comprender este teorema, hace falta conocer lo que es un esquema.

Definición 5.1.7 *Un esquema* $\boldsymbol{B}$ *es un conjunto de cromosomas donde se fijan ciertos genes. Un ejemplo con la notación habitual de un esquema para cromosomas con 5 genes puede ser: (*,0,*,*,1). Esto quiere decir que las posiciones 2 y 5 quedan fijadas en esos valores.*

Definición 5.1.8 *Se define la longitud de un esquema como la diferencia entre la última y la primera posición fijadas.*
Para un esquema $\boldsymbol{B}$*, donde la posición i es la primera posición que se fija y la posición j es la última, esta longitud es* $\delta(\boldsymbol{B}) = j - i$

Teorema 5.1.1 ***TEOREMA DEL ESQUEMA***
Sea H un esquema que cumple una solución dentro del espacio de las soluciones. Sea $m(H,t)$ *el número de individuos que pertenecen al esquema H en la generación t,* $\overline{f(H)}$ *el fitness promedio de los individuos del esquema H,* $\overline{f}$ *el promedio de todos los individuos de la población,* p_c *la probabilidad de cruce,* p_m *la probabilidad de mutación y* $o(H)$ *el número de bits del esquema H. Entonces, se tiene la siguiente desigualdad:*

$$E(m(H,t+1)) \geq m(H,t) \cdot \frac{\overline{f(H)}}{\overline{f}} \cdot \left(1 - p_c \frac{\delta(H)}{l-1}\right) \cdot (1-p_m)^{o(H)}$$

Este resultado muestra que conforme van pasando las generaciones, los esquemas con un alto fitness. Entonces, aunque no se tienen unas condiciones de optimalidad generales, este teorema muestra que en la práctica podrá funcionar bien.
La demostración formal de este teorema, junto con una discusión más extensa está en el libro de Goldberg (1989)[14] en el capítulo 2 o en el libro de Holland [15] en el capítulo 5.

Para finalizar, la complejidad computacional de este algoritmo de optimización depende de la población, la complejidad de evaluar la función fitness y el número de generaciones. Primero, viendo el algoritmo, se tiene que a lo sumo se tienen T generaciones. Para una generación, si se supone que todas las generaciones tienen un tamaño M, hay que evaluar la función fitness M veces. Además hay que tener en cuenta el método de selección y el de cruce y mutación. Normalmente el método de cruce no

supera una complejidad superior a $O(M \log M)$ y el método de cruce y mutación, si los cromosomas tienen L genes, tiene una complejidad de $O(ML)$, ya que se evalúan L posiciones de los M genes. Por tanto, la complejidad será de $O(T(MT_{\text{eval}} + M \log M + ML))$. En la literatura, suele asumirse que la función fitness es sencilla de evaluar y los cromosomas son largos, por lo que suele verse simplificado a $O(TML)$. En esta simplificación, la complejidad solo dependería del número de generaciones, el tamaño de la población y el tamaño de los genes. Otra simplificación que se puede encontrar en la literatura es para el caso donde la complejidad de la evaluación de la función fitness es alta. En este caso, se simplifica la complejidad del algoritmo a $O(TMT_{\text{eval}})$.
En el problema de optimización que hay que resolver para el método bootstrap, la función fitness es la función que se debe optimizar, es decir, la función fitness es R(c). Los cromosomas son los distintos valores de c. Esto quiere decir que la mejor aproximación de la complejidad temporal es $O(TMT_{\text{eval}})$. A priori, este método será mas lento que el anterior método, pero puede que encuentre mejor máximo.

Para aquellos interesados en una introducción más profunda a los algoritmos genéticos, se recomienda consultar la obra de Goldberg (1989)[14], o la de Holland (1975) [15] que proporciona una base sólida en este campo y de la cual se ha obtenido la información básica que se ha expuesto.
Una cosa positiva de este método respecto al anterior es que es un método más robusto y conocido. Aporta más rigurosidad al máximo encontrado, pero con el precio de un mayor número de evaluaciones de función, que aumentan el coste temporal.

5.2. Algoritmo de simulación de recocido

Por último, otro algoritmo de optimización es el algoritmo de simulación de recocido. Fue propuesto por Scott Kirkpatrick, C. Daniel Gelatt, y Mario P. Vecchi en 1983 [16]. Este algoritmo, también conocido como algoritmo de enfriamiento, es un método que está inspirado en el proceso metalúrgico de recocido. Este método de optimización permite que las estructuras internas reconfiguren su estado hacia una mínima energía, similar al proceso de los átomos en un metal durante el recocido. El algoritmo aplica este concepto para encontrar soluciones óptimas en problemas de optimización complejos y de gran escala. Para este algoritmo hay que introducir varios conceptos iniciales típicos de este algoritmo.

Definición 5.2.1 *Se llama función de coste a una función:*

$$f : S \to \mathbb{R}$$

El conjunto S es el conjunto de soluciones posibles.

Definición 5.2.2 $\triangle E = f(x) - f(y)$ *se llama diferencia de energía entre x e y.*

Definición 5.2.3 *La temperatura T es un número real positivo y la tasa de enfriamiento $\alpha \in (0, 1)$*

Definición 5.2.4 *La probabilidad de aceptación de una solución x se expresa como $p = exp(-\frac{\triangle E}{T})$*

Una vez se han definido estos conceptos, se puede escribir el algoritmo:

1. Se inicia una temperatura inicial T_0, una tasa de enfriamiento α y una función de coste (la función a optimizar)
2. Se elige un punto inicial x_0
3. Mientras no se cumpla el criterio de parada, hacer:
   a) Generar una solución x' que pertenezca al vecindario de x
   b) evaluar $\triangle E = f(x') - f(x)$

c) Si $\triangle E \leq 0$, aceptar. En otro caso, aceptar con probabilidad $p = exp(-\frac{\triangle E}{T})$

d) Actualizar la temperatura como: $T \leftarrow \alpha T$

Los criterios de parada habituales para el algoritmo de recocido simulado incluyen un número máximo de iteraciones y un umbral de temperatura. Es decir, el algoritmo se detiene si la temperatura actual cae por debajo de una temperatura específica predefinida o si se alcanza el número máximo de iteraciones permitidas.
En estos algoritmos existen algunos teoremas con los que podemos asegurar una convergencia a un óptimo global, como el teorema de Geman & Geman, que se puede encontrar en Geman, S., & Geman, D. (1984) [17], o el de Lundy y Mees, que se puede encontrar en Lundy, M., & Mees, A. (1986) [18]. Ambos teoremas requieren de una temperatura inicial suficientemente alta y hacen uso de reglas para actualizar la temperatura de una manera controlada, para así poder asegurar una buena convergencia al óptimo. Sin embargo, requieren de un espacio de soluciones finito, por lo que limitan su aplicación al problema actual y por ello no se han presentado formalmente.

El algoritmo de recocido, en una primera aproximación pesimista, finaliza el método por el número de iteraciones. En este caso, la complejidad sería $O(NT_{\text{eval}})$, con N el número máximo de iteraciones y T_{eval} el tiempo de evaluar la función de coste una vez. Pero teniendo en cuenta el enfriamiento, esto es difícil de que ocurra, por lo que se puede reducir la complejidad a $O(\log(\frac{T_0}{T_{\mathbf{min}}})T_{\text{eval}})$. Dependiendo unicamente del tiempo de evaluar la función coste y de la temperatura inicial y la temperatura mínima que se fije. Para el problema que interesa en el método bootstrap, la función R(c) será la función coste. Este algoritmo, en un principio, podría ser algo más rápido que el genético y menos complejo que el primer algoritmo que se presentó. Para más detalles sobre este algoritmo, se recomienda consultar el trabajo de Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983) [16].

Como último comentario sobre el algoritmo bootstrap, cuando las varianzas de las variables aleatorias X e Y son muy grandes o distintas entre sí, puede darse el caso de que el método empleado para su análisis genere fallos o problemas. Esto se debe a que las diferencias significativas en la dispersión de los datos pueden afectar la precisión y confiabilidad de los resultados obtenidos. Para solventar esta situación, se propone modificar el estadístico utilizado en el análisis añadiendo un control de la varianza. Esta modificación consiste en ajustar el estadístico de tal manera que se tenga en cuenta la varianza de cada variable, equilibrando su influencia en el resultado final. Con este control, se busca mejorar la robustez y exactitud del método, reduciendo la probabilidad de errores debidos a varianzas desproporcionadas. Por tanto, se pueden tener las siguientes dos variantes del estadístico:

$$\epsilon_c = \int_{I_c} \frac{\left|\overline{Y_{1k}} - \overline{Y_{2k}}\right|}{\sqrt{\sigma^2_{Y_1} + \sigma^2_{Y_2}}} dk$$

$$\epsilon_c = \int_{I_c} \left| \frac{\overline{Y_{1k}}}{|\sigma_{Y_1}|} - \frac{\overline{Y_{2k}}}{|\sigma_{Y_2}|} \right| dk$$

Siendo σ_{Y_1} y σ_{Y_2} la desviación típica de Y_1 e Y_2 respectivamente.

6 Simulaciones

En esta sección se realizarán simulaciones utilizando las tres propuestas del documento: MANOVA, FANOVA y la nueva propuesta con el uso del bootstrap, junto con las propuestas de optimización presentadas en el capítulo anterior, para también comparar los tiempos de ejecución. Las librerías necesarias para estas simulaciones son las de FdAnova [3] para los test del FANOVA y la librería stats [31] para hacer el MANOVA. Para el método bootstrap, se usan librerías como GA [32] y las funciones que se encuentran en el archivo FuncionBootstrap.R. Para una mayor comprensión del funcionamiento del script, véase el manual de usuario [33].

Para ello, se llevarán a cabo las siguientes pruebas para los tres métodos presentados en este trabajo:

- **Prueba 1:** Evaluación de los tiempos de ejecución y resultados del método bootstrap, con el objetivo de identificar el método más rápido.
- **Prueba 2:** Uso de una distribución normal con media 0 y desviación estándar 1 para generar los datos de las dos matrices.
- **Prueba 3:** Uso de una distribución normal con media 0 y desviación estándar 1 para una de las matrices, mientras que para la otra se utilizará una distribución normal con media $\beta > 0$.
- **Prueba 4:** Comparación de diferentes tamaños de muestra (pequeño, mediano, grande).
- **Prueba 5:** Evaluación del efecto de diferentes distribuciones de datos (uniforme, exponencial, etc.).
- **Prueba 6:** Variación de los parámetros de las distribuciones (cambios en la media, desviación estándar).

La motivación para realizar estas pruebas es la siguiente:

- **Prueba 1:** Identificar el método más rápido evaluando los tiempos de ejecución del método bootstrap. Una vez se encuentre, se usará ese método para las demás pruebas.
- **Pruebas 2 y 3:** Evaluar si el test funciona correctamente bajo las condiciones de la hipótesis nula y alternativa, y determinar la proporción de rechazos cuando los datos tienen medias cercanas.
- **Prueba 4:** Observar el comportamiento de los test con diferentes tamaños muestrales y su efecto.
- **Pruebas 5 y 6:** Verificar el funcionamiento de los test con distintas distribuciones de datos y parámetros, similar a la Prueba 3 pero de manera más general.
  Como ultimo aporte, hay que saber que para los test del FANOVA, se podría necesitar el uso de una base de funciones. Para este caso, se ha optado por el uso de 20 nodos para todas las bases.

6.1. Prueba 1

Como se explicó anteriormente, en esta prueba se probarán los tres métodos de optimización para el caso de dos muestras normales (una con media 0 y otra con media 1). Se harán tres pruebas con el tamaño. Con tamaño pequeño, (se usarán matrices de tamaño 10×10). Se ejecutarán 5 simulaciones del programa para cada método de optimización. Posteriormente, se estudiarán los máximos encontrados por cada método y se observará si existen cambios notables entre ellos o, por el contrario, se encuentran puntos cercanos. Por tanto, se usarán para todas las simulaciones las mismas muestras.
Para tamaño mediano, se usarán matrices de tamaño 200×50 y se volverán a hacer 5 simulaciones y se procederá al estudio de los valores c obtenidos y el tiempo de obtención.

Por último, para el caso de un tamaño grande, se ha usado una matriz de tamaño 1000×40, solamente se ejecutará una vez cada método. El motivo por el cual se opta por no hacer mas simulaciones es principalmente el tiempo de ejecución extremadamente elevado. Hay que tener en cuenta que los valores de c son meramente ilustrativos de los que se podrían obtener. Al hacerse tan pocas simulaciones, no son resultados concluyentes.

Para estos casos, el primer método donde se usaba una malla en un intervalo, se tendrá el intervalo $[0, 3]$ y se dará un paso de tamaño 0.002 para todos los tamaños de muestras. Para el algoritmo genético, se usa la librería de "GA" ([19]) con la configuración de una tasa de mutación de 0.2 y una tasa de cruce de 0.8, un número máximo de iteraciones de 100 y 25 iteraciones máximas donde no hay cambio para detener el algoritmo.
Para el algoritmo de enfriamiento se ha usado la función de optimización de R ([20])) con la configuración adecuada para la optimización con este algoritmo. Los resultados obtenidos son los siguientes:

Método	**Tamaño**	**Tiempo de una ejecución aproximada**	**Valor de c**
Malla	Pequeño	20 minutos	0.3169
	Mediano	16 horas	0.37104
	Grande	más de 3 días	–
Algoritmo Genético	Pequeño	30 minutos	0.4130
	Mediano	1.7 días	0.37513
	Grande	más de 3 días	–
Algoritmo de Enfriamiento	Pequeño	35 minutos	0.4501
	Mediano	1.4 días	0.4101
	Grande	más de 3 días	–

Cuadro 6.1: Comparación de tiempos de los diferentes métodos

Por tanto, se decidió por lo visto en estos resultados usar el algoritmo de malla. Porque aunque los otros dos algoritmos son más sofisticados, son también más lentos porque requieren un mayor costo computacional. Los máximos encontrados con estos métodos pueden ser de mejor calidad, pero no se diferencian tanto de los encontrados por el algoritmo que hace uso de la malla. Por lo que se ha decidido prescindir de calidad en los resultados en favor de una mayor velocidad de ejecución. Por último, para el caso de un tamaño grande, no se pudo encontrar un valor de c porque el tiempo computacional excedía los 3 días y se paró el programa.
Como último comentario de la prueba, estos resultados tan poco favorables han obligado a requerir de la potencia de cómputo del superordenador del IFCA, el Altamira. Esto ha logrado que se reduzca de manera muy significativa el tiempo de cómputo. En conclusión, para poder lograr hacer las simulaciones, se deberá usar una malla tosca y el supercomputador Altamira, con el fin de que el tiempo para hacer los experimentos se reduzca considerablemente.

6.2. Prueba 2

Como se ha expuesto, en esta prueba inicial únicamente se hará para una primera comprobación de que el test funciona correctamente en el caso de que se cumpla la hipótesis nula. En los test de hipótesis hay una probabilidad α (el nivel de significación) de cometer un error de tipo I, es decir, un falso rechazo. En este caso, $\alpha = 0{,}05$, por lo que lo esperado es que se rechace falsamente un 5 % de las veces. Para cada test de cada modelo, se ejecutarán 1000 simulaciones y en los casos donde se obtenga un p-valor, se mostrará un resumen de esos p-valores. En todos los test se han usado 400 muestras en $\mathbb{R}^{40}$, para ambos conjuntos de datos, Y_1 e Y_2.

6.2.1. MANOVA

	Media	Desviación estándar	proporción de rechazos
Wilks	0.4982966	0.2894941	0.0512
Hotelling-Lawley	0.4982966	0.2894941	0.0512
Roy	0.4982966	0.2894941	0.0512
Pillai	0.4983033	0.2894804	0.05119488

Cuadro 6.2: Resumen de los p-valorespara la prueba 2 con MANOVA

En este caso, tenemos los resultados esperados y, además, vemos que los cuatro test dan resultados parecidos.

6.2.2. FANOVA

Test	**Proporción de rechazos**	**Media de p-valores**	**Desviación estándar**
FP	0.058	0.5044340	0.3021575
CH	0.046	0.5020878	0.2958862
CS	0.050	0.4990012	0.2969175
L2N	0.052	0.5000730	0.2942773
L2B	0.061	0.5016862	0.2984682
L2b	0.052	0.4988807	0.2967507
FN	0.051	0.5003977	0.2938576
FB	0.056	0.5026161	0.2979975
Fb	0.047	0.5053808	0.2958759
GPF	0.051	0.4977767	0.2868301
Fmaxb	0.043	0.4916419	0.2842790
TRP_anova	0.037	0.6058528	0.3220415
TRP_ATS	0.040	0.6002095	0.3236287
TRP_WTPS	0.037	0.6055274	0.3223910

Cuadro 6.3: Resumen de los p-valores usando FANOVA en la prueba 2

Nuevamente, los resultados son los esperados o incluso mejores de lo esperado, como en los últimos test. Pero en general todos los test mantienen resultados cercanos al 5 % de rechazos.

6.2.3. Bootstrap

Caso	Proporción de rechazos	valor promedio de c
Caso 1	0.047	0.37718

Cuadro 6.4: Proporción de rechazos y valor promedio de c

Los resultados del método bootstrap también son cercanos al 5 % de rechazos. Por lo que se puede decir que es un método que cumple unos requisitos mínimos de calidad.
Como conclusión de esta prueba, se puede observar que todos los métodos tienen la misma proporción de rechazos y son proporciones cercanas al 0.05. Es lo esperado para esta prueba por lo que se puede concluir que el método nuevo así como todos los demás métodos conocidos se comportan con normalidad en este caso.

6.3. Prueba 3

En esta prueba, se empezará a probar como afecta a la proporción de rechazos un cambio en la media. Como se mencionó anteriormente, una muestra tendrá la media fijada en 0 mientras que la otra muestra tendrá una media β. Este β irá variando, tomando unos valores iniciales $\{1, 0{,}5, 0{,}3, 0{,}2, 0{,}1, 0{,}01\}$; con el fin de encontrar un punto en ese conjunto donde el test empiece a fallar. Una vez se encuentre ese punto entre los beta, se irá depurando para encontrar uno mejor. En cualquier caso, para los conjuntos Y_1 e Y_2 se tendrán 400 muestras de $\mathbb{R}^{40}$. En el caso de $\beta = 1$, también habrá un resumen de los p-valores obtenidos.

6.3.1. MANOVA

	Proporción de rechazos	Media	Desviación estándar
Wilks	1	$1{,}941544 \times 10^{-60}$	$7{,}855328 \times 10^{-59}$
Hotelling-Lawley	1	$1{,}941544 \times 10^{-60}$	$7{,}855328 \times 10^{-59}$
Roy	1	$1{,}941544 \times 10^{-60}$	$7{,}855328 \times 10^{-59}$
Pillai	1	$1{,}94135 \times 10^{-60}$	$7{,}854936 \times 10^{-59}$

Cuadro 6.5: Resumen de los p-valores con MANOVA en la prueba 3 con $\beta = 1$

	$\beta = 1$	$\beta = 0{,}5$	$\beta = 0{,}3$	$\beta = 0{,}2$	$\beta = 0{,}1$	$\beta = 0{,}01$
Wilks	1	1	1	1	1	0.06
Hotelling-Lawley	1	1	1	1	1	0.06
Roy	1	1	1	1	1	0.06
Pillai	1	1	1	1	1	0.05994006

Cuadro 6.6: Resumen de la proporción de rechazos de MANOVA para la prueba 3 con diferentes valores de β

Aquí teóricamente se debería de rechazar siempre, independientemente del β, pero se ve que no sucede. Esto es lógico porque se dispone de una muestra limitada. Cuando las diferencias entre medias son sutiles, los modelos tienden a fallar y rechazar menos, por lo que podemos afirmar que son resultados muy correctos los que aporta el MANOVA. Sobre la media y la desviación estándar de los p-valores, vemos que es algo esperado. Todos los p-valores obtenidos están concentrados muy cerca del 0.

6.3.2. FANOVA

Test	**Proporción de rechazos**	**Media de p-valores**	**Desviación estándar**
FP	1	0	0
CH	1	0	0
CS	1	0	0
L2N	1	0	0
L2B	1	0	0
L2b	1	0	0
FN	1	0	0
FB	1	0	0
Fb	1	0	0
GPF	1	0	0
Fmaxb	1	0	0
TRP_anova	1	9.193368e-14	2.215164e-12
TRP_ATS	1	7.406964e-14	1.791091e-12
TRP_WTPS	1	0	0

Cuadro 6.7: Resumen de los p-valores con FANOVA en la prueba 3 con $\beta = 1$

Test	$\beta = 1$	$\beta = 0{,}5$	$\beta = 0{,}3$	$\beta = 0{,}2$	$\beta = 0{,}1$	$\beta = 0{,}01$
FP	1	1	1	1	1	0.086
CH	1	1	1	1	1	0.08
CS	1	1	1	1	1	0.079
L2N	1	1	1	1	1	0.083
L2B	1	1	1	1	1	0.084
L2b	1	1	1	1	1	0.079
FN	1	1	1	1	1	0.083
FB	1	1	1	1	1	0.083
Fb	1	1	1	1	1	0.079
GPF	1	1	1	1	1	0.06
Fmaxb	1	1	1	1	1	0.069
TRP_anova	1	1	1	1	1	0.063
TRP_ATS	1	1	1	1	1	0.063
TRP_WTPS	1	1	1	1	1	0.061

Cuadro 6.8: Resumen de los p-valores con FANOVA en la prueba 3 con diferentes valores de β

En esta prueba se ven resultados esperados. Como en el MANOVA, es lógico encontrar una tasa de rechazo muy pequeña cuando las diferencias de media son pequeñas. Como un resultado a analizar es la media y la desviación típica de los p-valores obtenidos. En la mayoría aparece media 0 y desviación estándar 0. Esto realmente es por el redondeo. Los p-valores están tan cercanos a 0 que R lo redondea como 0, por lo que el resto de cálculos de media o desviación típica los hace con todos los p-valores redondeados a 0.

6.3.3. Bootstrap

Test	$\beta = 1$	$\beta = 0{,}5$	$\beta = 0{,}3$	$\beta = 0{,}2$	$\beta = 0{,}1$	$\beta = 0{,}01$
Proporción de rechazos	1	1	1	1	1	0.071
Valor de c	0.41301	0.303	0.303	0.303	0.049	0.049

Cuadro 6.9: Resumen de los p-valores con el método bootstrap en la prueba 3 con diferentes valores de β

El resultado también es el esperado. Los resultados de rechazos son muy positivos en todas las pruebas. Como en los otros casos, no es preocupante que al tener medias muy parecidas la tasa de rechazos sea inferior.
Como conclusión de estos resultados, vemos que el MANOVA arroja una proporción de rechazos algo inferior que el FANOVA para el caso de $\beta = 0{,}01$, lo que nos puede indicar que los test del FANOVA son algo más precisos. Estos dos métodos aportan las mismas proporciones de rechazos en el resto de casos. Para el método bootstrap se puede observar que tiene el mismo nivel de exactitud que los métodos del FANOVA.

6.4. Prueba 4

En este caso, se procederá a repetir las simulaciones de la prueba 3 con $\beta = 0{,}01$ con 3 tamaños de muestras. Para la muestra pequeña, se usarán 40 muestras en $\mathbb{R}^{40}$, para el tamaño mediano, se dejarán las 400 muestras obtenidas anteriormente y para el tamaño grande, se usarán 1000 muestras. El objetivo principal será observar en un caso muy desfavorable el comportamiento de los test.

6.4.1. MANOVA

	Tamaño pequeño	Tamaño mediano	Tamaño grande
Wilks	0.055	0.06	0.082
Hotelling-Lawley	0.055	0.06	0.082
Roy	0.055	0.06	0.082
Pillai	0.05494505	0.05994006	0.08291708

Cuadro 6.10: Resumen de la proporción de rechazos de MANOVA para la prueba 4 con diferentes tamaños con $\beta = 0{,}01$

Se puede observar una tasa de rechazos normal. Las diferencias entre medias son casi imperceptibles. Igualmente, se ve la tendencia de que si se aumenta la muestra, la tasa de rechazos aumenta ligeramente. Esto es normal porque aumentar la muestra mejora la precisión del test reduciendo el error de tipo II.

6.4.2. FANOVA

Test	**tamaño pequeño**	**tamaño mediano**	**tamaño grande**
FP	0.066	0.086	0.13
CH	0.048	0.08	0.122
CS	0.056	0.079	0.127
L2N	0.055	0.083	0.128
L2B	0.066	0.084	0.128
L2b	0.058	0.079	0.126
FN	0.055	0.083	0.128
Fb	0.064	0.079	0.128
FB	0.045	0.083	0.125
GPF	0.053	0.06	0.097
Fmaxb	0.064	0.069	0.112
TRP_anova	0.045	0.063	0.088
TRP_ATS	0.056	0.063	0.089
TRP_WTPS	0.05	0.061	0.093

Cuadro 6.11: Resumen de los p-valores con el método FANOVA en la prueba 4 con diferentes tamaños de muestra para $\beta = 0{,}01$

Son resultados normales, vemos nuevamente que la tasa de rechazo aumenta ligeramente al aumentar la muestra. En este caso la mejora en la tasa de rechazo es superior a la que presentó MANOVA. Es esperable que la tasa de rechazo sea así de baja porque el valor β es muy cercano al 0. Igualmente se puede observar la potencia de este test para lograr una tasa de rechazos alrededor del 10 % en este caso tan desfavorable.

6.4.3. Bootstrap

	Tamaño pequeño	Tamaño mediano	Tamaño grande
proporción de rechazos	0.059	0.081	0.114
Valor de c	0.048912	0.049564	0.049136

Cuadro 6.12: Resumen de la proporción de rechazos con método bootstrap para la prueba 4 con diferentes tamaños con $\beta = 0{,}01$

El resultado es el esperado como en las tablas anteriores y los cambios en las tasas de rechazo son similares que en FANOVA. Como algo más remarcable, se puede ver que el valor de c no se modifica notablemente al aumentar la muestra o la tasa de rechazos. Finalmente, se puede observar que el método MANOVA es el que tiene una tasa de rechazos de la hipótesis nula más baja en los tres tamaños. Mientras, los test del FANOVA y el nuevo test tienen proporciones de rechazos muy similares entre si.

6.5. Prueba 5

En esta prueba, con 400 muestras de $\mathbb{R}^{40}$ para cada conjunto, se cambiarán las distribuciones y se usará en cada caso una distinta:

- **Caso 1:** 2 distribuciones uniformes U(0,1).
- **Caso 2:** 2 distribuciones exponenciales con $\lambda = 1$
- **Caso 3:** 1 distribución uniforme en el intervalo (0,1) y una distribución exponencial con $\lambda = 1$.
- **Caso 4:** 1 distribución exponencial con $\lambda = 1$ y otra exponencial con $\lambda = 1{,}01$
- **Caso 5:** 1 distribución exponencial con $\lambda = 1$ y otra exponencial con $\lambda = 1{,}05$

6.5.1. MANOVA

	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
Wilks	0.057	0.054	1	0.058	0.567
Hotelling-Lawley	0.057	0.054	1	0.058	0.567
Roy	0.057	0.054	1	0.058	0.567
Pillai	0.057	0.05394605	1	0.05794206	0.5674326

Cuadro 6.13: Resumen de la proporción de rechazos de MANOVA para la prueba 5 con diferentes casos

Se puede observar que si bien los datos no cumplen con los requistos de normalidad que se exige en las hipótesis del MANOVA, en los casos extremos si funciona con una cierta normalidad. Pero se reduce mucho la precisión cuando es un caso donde las diferencias de medias no son tan extremas, que es lo que pasa en el caso 5. En el caso 4 los resultados obtenidos muestran una incapacidad para ver las diferencias.

6.5.2. FANOVA

Test	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
FP	0.049	0.055	1	0.082	0.827
CH	0.054	0.055	1	0.086	0.903
CS	0.055	0.052	1	0.086	0.902
L2N	0.055	0.053	1	0.087	0.905
L2B	0.055	0.054	1	0.087	0.905
L2b	0.055	0.052	1	0.087	0.902
FN	0.055	0.053	1	0.087	0.904
Fb	0.055	0.054	1	0.087	0.905
FB	0.054	0.052	1	0.086	0.902
GPF	0.051	0.044	1	0.057	0.758
Fmaxb	0.051	0.055	1	0.065	0.913
TRP_anova	0.05	0.035	1	0.068	0.758
TRP_ATS	0.05	0.035	1	0.068	0.763
TRP_WTPS	0.053	0.036	1	0.072	0.760

Cuadro 6.14: Resumen de los p-valores con el método FANOVA en la prueba 5 con diferentes casos

Se puede observar que en casos extremos donde las diferencias de medias son grandes o son mismas distribuciones da los resultados esperados. Cuando las diferencias de medias no son tan grandes como en el caso 5, se puede observar que los resultados en su mayoría son muy positivos. Todos los test tienen proporciones de rechazo superiores al 70 % y 9 de ellos superiores al 90 %, pero en el caso donde las diferencias son sutiles, el test da resultados poco favorables.

6.5.3. Bootstrap

	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5
proporción de rechazos	0.043	0.049	1	0.06l7	0.917
valor de c	0.073666	0.260534	0.184	0.223719	0.2425769

Cuadro 6.15: Resumen de los p-valores con el método bootstrap en la prueba 5 con diferentes casos

Se observan datos positivos. Las tasas de rechazo son similares a las de los mejores test del FANOVA en esta prueba y se pueden analizar de una forma similar.
La conclusión de estas pruebas es que se observa que el MANOVA tiene dificultades a la hora de rechazar en los casos 4 y 5. Destaca el caso 5 sobre todo, pues la diferencia entre esos resultados y los obtenidos con el FANOVA y el nuevo método son grandes. De rechazar en torno a un 56,7 % de las veces, pasa a mínimo un 76 % con el FANOVA y a más del 90 % de las veces con el método nuevo.

6.6. Prueba 6

En esta prueba se optará nuevamente por 400 muestras de $\mathbb{R}^{40}$ con distribuciones normales de media 0 como en la prueba 2, pero cambiando la desviación típica. Se procederá con estos casos:

- **Caso 1**: Desviación típica 10 en ambos conjuntos.
- **Caso 2**: Desviación típica 1 en un conjunto y 10 en otro.

- **Caso 3**: Desviación típica 1 en un conjunto y 100 en otro.
- **Caso 4**: Desviación típica 100 en ambos conjuntos.

Para el caso con media 0.5 y los valores de β límite de la prueba 3, se repetirán los casos anteriormente descritos.

6.6.1. MANOVA

	Caso 1	**Caso 2**	**Caso 3**	**Caso 4**
Wilks	0.043	0.075	0.093	0.084
Hotelling-Lawley	0.043	0.075	0.093	0.084
Roy	0.043	0.075	0.093	0.084
Pillai	0.04295704	0.075	0.09290709	0.08391608

Cuadro 6.16: Resumen de la proporción de rechazos de MANOVA para la prueba 6 con diferentes casos para cuando la hipótesis nula es cierta

	Caso 1	**Caso 2**	**Caso 3**	**Caso 4**
Wilks	0.585	0.97	0.097	0.091
Hotelling-Lawley	0.585	0.97	0.097	0.091
Roy	0.585	0.97	0.097	0.091
Pillai	0.5844156	0.97003	0.0969031	0.09090909

Cuadro 6.17: Resumen de la proporción de rechazos de MANOVA para la prueba 6 con diferentes casos para cuando la hipótesis nula es falsa

Se puede dividir el análisis entre el caso donde no se debería rechazar y el que sí.
En el caso donde no se debe rechazar, los resultados son malos ya que aumenta la tasa de rechazos cuando no debería. En este caso lo que más ha afectado es la diferencia en las varianzas. Es normal porque una hipótesis inicial para el MANOVA es igualdad en las matrices de covarianza. Al tener desviaciones típicas distintas en cada coordenada, esta igualdad en las matrices de covarianzas no se cumple y por eso el método falla. Un problema que también presenta es el aumento de la varianza. Aunque las matrices de covarianza sean iguales, el método igualmente presenta un rechazo mayor.

En el caso donde se debería rechazar, se puede obsvar que el aumento de la varianza ya es un problema para MANOVA. En este caso se observa que es mejor tener varianzas distintas aunque signifique distintas matrices de covarianza, que varianzas iguales pero grandes. Esto se refleja en el caso 1 y 2. Igualmente, cuando las desviaciones típicas son muy grandes, el hecho de que cumpla o no las hipótesis iniciales completamente no afecta demasiado al resultado, que es muy negativo.

6.6.2. FANOVA

Test	**Caso 1**	**Caso 2**	**Caso 3**	**Caso 4**
FP	0.039	0.041	0.041	0.054
CH	0.032	0.042	0.042	0.048
CS	0.034	0.042	0.042	0.046
L2N	0.034	0.042	0.042	0.048
L2B	0.035	0.043	0.043	0.049
L2b	0.034	0.042	0.042	0.046
FN	0.034	0.042	0.042	0.047
Fb	0.035	0.042	0.042	0.048
FB	0.033	0.043	0.043	0.047
GPF	0.052	0.36	0.036	0.058
Fmaxb	0.047	0.053	0.053	0.054
TRP_anova	0.043	0.035	0.035	0.048
TRP_ATS	0.044	0.035	0.035	0.046
TRP_WTPS	0.046	0.04	0.04	0.048

Cuadro 6.18: Resumen de los p-valores con el método FANOVA en la prueba 6 con diferentes casos para cuando la hipótesis nula es cierta

Test	**Caso 1**	**Caso 2**	**Caso 3**	**Caso 4**
FP	0.897	1	0.999	0.07
CH	0.914	1	1	0.072
CS	0.911	1	1	0.073
L2N	0.915	1	1	0.072
L2B	0.917	1	1	0.074
L2b	0.915	1	1	0.07
FN	0.915	1	1	0.072
Fb	0.916	1	1	0.074
FB	0.912	1	1	0.07
GPF	0.746	1	1	0.062
Fmaxb	0.926	0.981	0.981	0.06
TRP_anova	0.774	0.9811	1	0.059
TRP_ATS	0.775	0.982	0.983	0.057
TRP_WTPS	0.775	0.992	0.983	0.063

Cuadro 6.19: Resumen de los p-valores con el método FANOVA en la prueba 6 con diferentes casos para cuando la hipótesis nula es falsa

Nuevamente se puede dividir el análisis entre el caso donde no se debería rechazar, que presenta resultados correctos y muy favorables. Todos estos resultados tienen proporciones de rechazo cercanas al 0.05. Quiere decir que en los casos donde se cumple la hipótesis nula, al test no le afecta significativamente la modificación de la varianza.

Por otro lado, en el caso donde se debería rechazar, se observa que los resultados son buenos independientemente de que no se tenga igualdad en las matrices de covarianza, que se puede ver en el caso 2 y 3. También se observa que un aumento en la desviación típica reduce la precisión del test. Esto se puede ver en el caso 1,2 y 4. En el caso 2 hay mejores proporciones de rechazo que en el caso 1 y el

caso 1 mejores proporciones que el caso 4.

6.6.3. Bootstrap

Test	**Caso 1**	**Caso 2**
proporción de rechazos con la hipótesis nula cierta	0.056	0.058
valor medio de c	0.45061	0.02023
proporción de rechazos con la hipótesis nula falsa	0.551/0.859/0.860	1/1/1
valor medio de c	0.18012/0.1795/0.18195	0.11805/0.12501/0.1236

Cuadro 6.20: Resumen de los p-valores con el método bootstrap en la prueba 6 (Caso 1 y Caso 2)

Test	**Caso 3**	**Caso 4**
proporción de rechazos con la hipótesis nula cierta	0.046	0.046
valor medio de c	0.3367805	0.87430
proporción de rechazos con la hipótesis nula falsa	1/1/1	0.073/0.775/0.764
valor medio de c	0.33507/0.34452/0.33845	0.01248/0.02459/0.01738

Cuadro 6.21: Resumen de los p-valores con el método bootstrap en la prueba 6 (Caso 3 y Caso 4)

En el método bootstrap, se habló en capítulos anteriores sobre las modificaciones que tenía este método para tratar la varianza:

$$\epsilon_c = \int_{I_c} \frac{|\overline{Y_{1k}} - \overline{Y_{2k}}|}{\sqrt{\sigma^2_{Y_1} + \sigma^2_{Y_2}}}\, dk$$

$$\epsilon_c = \int_{I_c} \left| \frac{\overline{Y_{1k}}}{|\sigma_{Y_1}|} - \frac{\overline{Y_{2k}}}{|\sigma_{Y_2}|} \right| dk$$

Siendo σ_{Y_1} y σ_{Y_2} la desviación típica de Y_1 e Y_2 respectivamente. Estas modificaciones se aplicaron únicamente en el caso donde se rechaza la hipótesis nula, porque los resultados cuando la hipótesis nula es cierta ya son correctos cuando no se aplica ninguna modificación. Como se podía observar en la tabla, hay 3 valores, el primer valor es la proporción de rechazos sin ninguna modificación, el segundo valor corresponde con la tasa de rechazos usando la primera modificación y el tercer valor, corresponde con la segunda modificación. De la misma manera se pueden interpretar los tres valores de c, que corresponden con cada caso mencionado anteriormente.

Sabiendo esto, como se ha dicho anteriormente, los datos en los casos donde no se debería rechazar, arrojan datos correctos. Proporciones de rechazo aceptables y cercanas al 5 %. Para el caso donde se debe rechazar, se observan diferencias significativas en las tasas de rechazo entre aplicar una modificación y no. En el caso de aplicar una modificación, no hay diferencias entre las dos disponibles que se han aplicado. Por otro lado, el valor medio de c no cambia demasiado entre usar una modificación o no. Cuando no se usa una modificación, la tasa de rechazo es buena cuando hay diferencia entre matrices de covarianza pero es gracias a que la desviación típica de uno de los grupos es baja (casos 2 y 3). Cuando se aumenta la varianza (casos 1 y 4), el error aumenta y baja el número de rechazos obtenido de forma drástica. Usando estas modificaciones mantienen la calidad en los casos donde el método funcionaba muy bien sin modificaciones y mejora considerablemente la calidad en los casos

donde aumenta la varianza (casos 1 y 4). Además el descenso de calidad del resultado conforme aumenta la varianza es ligero y poco preocupante.

Como conclusión de estas pruebas, tenemos que el método que mejor responde a un aumento grande de la varianza es el método bootstrap con cualquier modificación. Se tiene que la tasa de rechazos en el MANOVA es peor en todos los casos en comparación al FANOVA y el método bootstrap. Entre el método FANOVA y el bootstrap, se tiene que cuando la desviación típica aumenta poco, los test del FANOVA son correctos, pero si aumenta considerablemente, los test del FANOVA dejan de funcionar correctamente. Por otro lado, en el método bootstrap si bien se reduce la proporción de rechazos, la reducción es ligera y mantiene una calidad muy aceptable.

7 Conclusiones

En este documento se han presentado tres enfoques distintos para el análisis de la varianza en el caso multivariante: el MANOVA, FANOVA y un método basado en bootstrap. De estos tres métodos se han hecho simulaciones con las que ahora se pueden ver las ventajas y desventajas de cada método para obtener unas conclusiones.

El MANOVA es una herramienta clásica y ampliamente utilizada, reconocida por su rapidez y su capacidad para controlar el error de tipo I de manera eficiente. Sin embargo, es importante señalar que sus resultados, aunque robustos, son esperables, ya que se trata de un método más antiguo y consolidado en la literatura. Esto explica por qué sigue siendo efectivo en contextos donde se cumplen los supuestos estrictos de homogeneidad de las matrices de covarianza y normalidad de los datos. No obstante, es relevante mencionar que estos requisitos pueden relajarse cuando se trabaja con un gran número de muestras o al aplicar versiones robustas del test.

Por otro lado, el FANOVA, una opción más moderna, ofrece una mayor flexibilidad con varios tests disponibles, muchos de ellos no paramétricos, lo que le permite ser utilizado en situaciones donde los supuestos de normalidad no se cumplen. Las simulaciones realizadas en este estudio han mostrado resultados positivos en estos casos, lo que lo convierte en una opción recomendable para una variedad más amplia de escenarios, especialmente cuando no se puede asumir normalidad en los datos.

Finalmente, el método bootstrap se destaca por no requerir grandes hipótesis iniciales, lo que lo hace aplicable a un abanico más amplio de problemas. Las simulaciones realizadas han evidenciado que cuando las varianzas son altas o existen diferencias significativas entre ellas, los resultados obtenidos con MANOVA y FANOVA pueden ser insuficientes. En contraste, el método bootstrap ha ofrecido resultados mucho más favorables en estos casos. En los demás escenarios, los resultados del bootstrap han sido similares, e incluso superiores, en algunos aspectos.
Sin embargo, una desventaja considerable del método bootstrap es su complejidad tanto temporal como espacial. Aunque se podrían realizar optimizaciones en el código para reducir la carga computacional, sigue siendo un enfoque poco práctico en problemas de gran escala. Incluso para problemas de tamaño medio, aunque el método bootstrap ofrece mejores resultados, su uso puede no ser recomendable en comparación con los tests del FANOVA, debido a la eficiencia de estos últimos.

En conclusión, este estudio ha mostrado que aplicar un modelo funcional de análisis de varianza es una excelente opción para estudios con varianzas reducidas. Mientras tanto, el método bootstrap es superior para analizar datos con varianzas elevadas. La relativa simplicidad y rapidez del MANOVA son atributos esperables dado su antigüedad, pero aunque en un principio puede verse como una debilidad, en la mayoría de problemas sencillos será una excelente herramienta. Sin embargo, los problemas y desafíos contemporáneos a veces exigen un mayor esfuerzo. Es ahí donde los métodos del FANOVA y el nuevo método bootstrap son una alternativa a tener muy en cuenta para tener mejores resultados.

Bibliografía

[1] Vidakovic, Brani. *Wavelets for Kids: A Tutorial Introduction*. Disponible en: https://www2.isye.gatech.edu/~brani/Wavelets/02-02.pdf

[2] Aneiros Pérez, Gemma, et al. *Análisis de la Varianza Funcional*. ICM.

[3] Tomasz Górecki, Lukasz Smaga. *fdANOVA: An R Package for ANOVA for Functions*. CRAN. Disponible en: https://cran.r-project.org/web/packages/fdANOVA/vignettes/fdANOVA.pdf

[4] Johnson, Richard A., y Wichern, Dean W. (2002). *Applied Multivariate Statistical Analysis*. Quinta edición. Prentice Hall, Upper Saddle River, New Jersey. ISBN 0-13-092553-5.

[5] Alcaraz Ramos, Rafael. *Capítulo 10: Lambda de Wilks*. Universidad de Alicante. Disponible en: https://rua.ua.es/dspace/bitstream/10045/12080/1/Capitulo10.pdf

[6] Penn State Eberly College of Science. *MANOVA: Multivariate Analysis of Variance*. Disponible en: https://online.stat.psu.edu/stat505/lesson/8/8.3

[7] Penn State Eberly College of Science. *MANOVA*. Disponible en: https://online.stat.psu.edu/stat505/book/export/html/762

[8] SAS Institute Inc. *Approximating the Distributions of the MANOVA Test Statistics to the F Distribution*. Disponible en: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_introreg_sect012.htm

[9] Statology. *Pillai's Trace*. Disponible en: https://www.statology.org/pillais-trace/

[10] Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*. Probability and Mathematical Statistics. Academic Press Inc. Disponible en: https://statisticalsupportandresearch.wordpress.com/wp-content/uploads/2017/06/k-v-mardia-j-t-kent-j-m-bibby-multivariate-analysis-probability-and-mathematical-statistics-academic-press-inc-1979.pdfhttps://statisticalsupportandresearch.wordpress.com/wp-content/uploads/2017/06/k-v-mardia-j-t-kent-j-m-bibby-multivariate-analysis-probability-and-mathematical-statistics-academic-press-inc-1979.pdf

[11] Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson Prentice Hall. ISBN: 978-0131877153.

[12] Cuevas, A., Febrero, M., & Fraiman, R. (2004)."An ANOVA test for functional dat". In *Computational Statistics & Data Analysis*, 47(1), 111-122. doi:10.1016/j.csda.2003.10.021.

[13] Horváth, L., & Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

[14] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

[15] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

[16] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671-680. Disponible en: https://doi.org/10.1126/science.220.4598.671

[17] Geman, S., & Geman, D. (1984). *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6), 721-741.

[18] Lundy, M., & Mees, A. (1986). *Convergence of an Annealing Algorithm*. Mathematical Programming, 34(1), 111-124.

[19] Scrucca, L. (2024). *A quick tour of GA*. R package vignette, version 3.2. `https://cran.r-project.org/web/packages/GA/vignettes/GA.html`

[20] R Core Team. (2024). *optim: General-purpose Optimization*. R documentation. `https://stat.ethz.ch/R-manual/R-devel/library/stats/html/optim.html`

[21] Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.

[22] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley.

[23] Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis* (2nd ed.). Springer.

[24] Ferraty, F., & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

[25] Antoniadis, M. (2007). Wavelet methods in statistics: Recent developments and applications. In *Recent Advances in Nonparametric Functional Data Analysis* (pp. 1-24). Springer.

[26] Galeano, P., Cuevas, A. (2006). Functional analysis of variance: A review. *Computational Statistics Data Analysis*, 51(1), 453-465. DOI: `10.1016/j.csda.2005.06.007`

[27] Cuevas, A., Galeano, P. (2008). Functional ANOVA: A review. *Statistical Science*, 23(1), 41-62. DOI: `10.1214/07-STS213`

[28] Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics*, 2, 360-378.

[29] Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.

[30] J. A. Cuesta Albertos, A Tuero Díaz. Inferencia estadística

[31] R Core Team. (2019). Package 'stats', version 3.6.2. *https://www.rdocumentation.org/packages/stats/versions/3.6.2*.

[32] Luca Scrucca. (2013). GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1-37. Disponible en https://cran.r-project.org/package=GA.

[33] Miguel Pérez. (2024). *Manual de usuario*.

8 Apéndice

En esta sección se mostrarán algunos teoremas con su respectiva demostración, que son necesarios para algunos apartados de la memoria, pero no son lo suficiente relevantes para el trabajo como para añadirlos en alguna sección del trabajo; como el teorema de descomposición espectral que se ha usado varias veces durante la memoria y que se vio en clase.
A continuación se nombrarán dos enunciados del teorema de descomposición espectral, junto con su demostración que se han extraído de la página web de Leonardo Martínez, profesor de ciencias de la UNAM

Teorema 8.0.1 *Teorema de descomposición espectral 1 Sea V un espacio euclideo y $T : V \leftarrow V$ una transformación simétrica. Entonces existe una base ortonormal de V de los vectores propios de T.*

Demostración: Se tiene que el espacio euclídeo tiene dimensión finita n. Se va a probar por inducción fuerte. Suponiendo que n = 1, se tiene que para un valor propio λ hay un único vector propio v. Se tiene que $\frac{v}{||v||}$ también será vector propio y además forma una base ortonormal de V. Suponiendo que se cumple para todo espacio de dimensión menor que n, veamos si se cumple para uno de dimensión n.
Por el teorema fundamental del álgebra, se tiene que existirá una raíz compleja λ de polinomio característico de T. Cualquier matriz A que represente a la transformación también tendrá esa misma raíz λ de su polinomio característico. Al ser A simétrica, se tiene que esta raíz compleja será real. Considerando el núcleo W de la transformación $\lambda I - T$. Si W es de dimensión n, entonces W = V y por tanto $T(v) = \lambda v$ para todo vector v en V, por lo que todo vector no nulo es vector propio de T y cualquier base ortonormal satisface lo buscado. Se puede suponer entonces que la dimensión de W es menor que n. Por lo que $V = W \oplus W^{\perp}$ Sea B una base ortonormal de W que estará formada por los vectores propios de T con valor propio λ. Como la restricción T_1 de T a $W^{\perp}$ es una transformación simétrica, podemos aplicar la hipótesis inductiva y encontrar una base ortonormal B' de vectores propios de T_1 y por tanto de T para $W^{\perp}$.
Volviendo a que $V = W \oplus W^{\perp}$. Se tiene que una base de V es $B \cup B'$ formada por vectores propios de T. Cada base por separada es ortonormal, por tanto el producto interno de dos vectores distintos de B o de B' es 0 y el producto de un vecor de B y otro de B' también es 0 porque un elemento está en W y otro en $W^{\perp}$. Por último, se tiene que todos los elementos de $B \cup B'$ tienen norma 1 y por tanto esta base de V es ortonormal y está compuesta por los vectores propios de T.

Teorema 8.0.2 *Teorema de descomposición espectral 2 Sea A una matriz simétrica en $\mathbb{R}^n$, entonces existe una matriz ortogonal P y una matriz diagonal D, ambas en $\mathbb{R}^n$, tal que:*

$$A = P^{-1}DP$$

Demostración: Como A es simétrica, la transformación $T : \mathbb{F}^n \leftarrow \mathbb{F}^n$ dada por $T(X) = AX$ es simétrica, entonces por la primera versión de este teorema existe una base ortonormal de $\mathbb{F}^n$ que consiste en los vectores propios de T. Estos vectores propios, tal como se define la aplicación, son también vectores propios de A. Se puede demostrar que si los vectores propios de A son $v_1, \ldots, v_n$ y sus valores propios son $\lambda_1, \ldots, \lambda_n$, se tiene que $A = BDB^{-1}$, donde B es una matriz donde cada columna i es el vector propio v_i de A y D es una matriz diagonal de los valores propios de A. Sabiendo esto, se puede ver que B es ortogonal. Se tiene que es ortogonal si $B^T B = I$. La entrada (i,j) de la matriz $B^T B$, es el producto de $v_i v_j$, que tal como se han definido ortonormales se tiene que $v_i v_j = 0$ si $i \neq j$ y $v_i v_j = 1$ si $i = j$. Se tiene también que la matriz B^{-1} también es ortogonal. Por último, se define $P = B^{-1}$.