UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Test para igualdad de medias entre un gran número de grupos.

Antonella Raffa

Directora: Daniela Rodriguez

14/05/2024

Agradecimientos

A mi Mamá, mi Papá, y mis abuelos. Gracias por su esfuerzo y sacrificio.

A mi hermano. Gracias por aportarme la confianza que a veces no tengo.

A mi directora, Daniela. Gracias por orientarme y acompañarme a lo largo de este trabajo. Espero algún día tener tu capacidad y conocimiento.

A Mariela y Marina, por aceptar ser jurados de esta tesis.

A mis compañeros de la facultad: Tamara, Lau, Mariela, Manu, Yami, Kari, Rocío y tantos otros que no nombro porque, de otra forma, la dedicatoria sería tan larga como el desarrollo de este trabajo. Un agradecimiento especial a Lu, Maru, Sole, Ceci, Meli y Noe. Si no las hubiese conocido, es probable que no hubiera llegado a esta instancia. Las admiro.

A las que están desde el principio: Vane, Caro y Estefi. Gracias por pensar siempre en mí.

A mis docentes y colegas. Si hay algo que me llevo más allá de conocimiento, es su entusiasmo por estar frente a un aula.

Por último, quisiera agradecer a quienes sostienen económicamente a la universidad. Su esfuerzo ha permitido mi formación de tal manera que si en alguna ocasión necesité más tiempo para comprender algunos conceptos, no tuve que preocuparme por cómo costearlo. Han sido la mano en el hombro que de alguna forma nos indica que si queremos estudiar, van a estar con nosotros, invitándonos a seguir adelante.

Índice general

Introducción **IV**

1. ANOVA Clásico **1**
1.1. Construcción del Estadístico del Test de ANOVA 2
1.2. Test 4
1.3. Limitaciones del test 6

2. Análisis de ANOVA para k grande **8**
2.1. Estadístico del Test 8
2.2. Demostración del Teorema 2.1.1 14
2.3. Construcción del Test 18

3. Estudio de Simulación **20**
3.1. Estudio de error de tipo I 20
3.1.1. Conclusiones 26
3.2. Estudio de la potencia 27
3.2.1. Conclusiones 32

Conclusiones del trabajo de tesis **34**

Bibliografía **36**

Resumen

El método de Análisis de Varianza (ANOVA) es una técnica estadística utilizada para comparar las medias de tres o más grupos para determinar si al menos uno de los grupos es significativamente diferente de los demás. Es decir, el ANOVA se utiliza para responder preguntas sobre si hay diferencias significativas entre grupos.

A diferencia del enfoque clásico, donde el número de grupos se mantiene fijo y el tamaño de muestra de cada población aumenta, en esta tesis exploraremos que sucede cuando se asume que el número de grupos (k) es grande y el tamaño de cada muestra está limitado o es pequeño en comparación con k. Estudiaremos como se obtiene la distribución asintótica del estadístico del test bajo la hipótesis nula de igualdad de las k medias. Específicamente, mostraremos que el estadístico del test es asintóticamente normal bajo la hipótesis nula.

También evaluaremos el rendimiento en muestras finitas del test basado en la distribución asintótica nula mediante un estudio de simulación en la que estudiaremos el comportamiento tanto del nivel como de la potencia del test.

Capítulo 1

ANOVA Clásico

El ANOVA es una herramienta ampliamente difundida en la investigación científica y se utiliza en una variedad de campos, desde la biología hasta la economía. Se aplica comúnmente en experimentos donde se prueban diferentes tratamientos, condiciones o niveles de un factor para determinar su efecto en una variable de interés. El método de ANOVA es una herramienta estadística fundamental para analizar y comprender las diferencias entre grupos en un conjunto de datos.

La idea general de esta herramienta se basa en examinar la variabilidad dentro de cada grupo y la variabilidad entre los grupos para determinar si las diferencias observadas son estadísticamente significativas o simplemente producto del azar. Esto se logra comparando la dispersión entre los grupos con la dispersión dentro de los grupos. Si la dispersión entre los grupos es significativamente mayor que la dispersión dentro de los grupos, entonces se concluye que al menos un grupo es diferente de los demás.

Mas precisamente, supongamos que tenemos k muestras independientes con varianzas iguales y que cada muestra tiene n observaciones. Llamemos $X_{11} \dots X_{1n}$, $X_{21} \dots X_{2n}, \dots,$ $X_{k1}, \dots, X_{kn}$ a las $N = kn$ variables aleatorias independientes, con varianza σ^2 que asumiremos finita y notemos $\mu_1, \dots \mu_k$ a las esperanzas del grupo 1 a k, respectivamente.

La hipótesis que se estudia en el problema de ANOVA es si las medias de estas k muestras son iguales. Para ello se diseñan tests de hipótesis que nos permitan saber si hay evidencia significativa como para rechazar nuestra afirmación. Es decir, el problema consiste en saber si tenemos evidencia significativa en contra de la hipótesis nula (H_0) siendo

$$H_0 : \mu_1 = \mu_2 \cdots = \mu_k$$

versus la hípotesis alternativa que supone que algún par de medias es diferente entre sí.

1.1. Construcción del Estadístico del Test de ANOVA

El término "Análisis de Varianza "fue acuñado por Fisher en su libro de 1925 "Statistical Methods for Research Workers"[2]. Fisher desarrolló el ANOVA como una extensión de la prueba t de Student, que había sido introducida por Gosset (1908), conocido bajo el seudónimo "Student"[8].

Fisher estaba interesado en comparar las medias de más de dos grupos y buscaba una forma de hacerlo de manera eficiente y efectiva. Observó que la dispersión de los datos podía dividirse en dos componentes: la variación debida a las diferencias entre los grupos y la variación debida a las diferencias dentro de los grupos. Fisher propuso una prueba estadística que comparaba estas dos fuentes de variación para determinar si las diferencias entre las medias de los grupos eran significativas.

El enfoque de Fisher en el ANOVA fue revolucionario en su época y tuvo un gran impacto en la investigación científica. Desde entonces, el ANOVA se ha convertido en una herramienta fundamental en muchas áreas de la ciencia, la ingeniería y la investigación social, utilizada para analizar datos experimentales y determinar si existen diferencias significativas entre grupos de datos. A lo largo de los años, el ANOVA ha seguido evolucionando con la introducción de diferentes variantes y extensiones, pero su concepto básico de comparar la variabilidad entre grupos con la variabilidad dentro de los grupos sigue siendo fundamental en su aplicación.

El estadístico F se calcula dividiendo la variabilidad entre grupos por la variabilidad dentro de los grupos. Si hay diferencias en las medias, esperaríamos que la variabilidad entre los grupos fuera mayor que la variabilidad dentro de los grupos. Por lo tanto, el valor del estadístico F sería grande.

Más precisamente, el ANOVA clásico utiliza el siguiente estadístico:

$$F = \frac{1}{k-1} \sum_{i=1}^{k} \frac{n \left(\overline{X}_{i.} - \overline{X}_{..} \right)^2}{{S_p}^2}$$

donde $\overline{X}_{i.} = n^{-1} \sum_{j=1}^{n} X_{ij}$; $\quad \overline{X}_{..} = k^{-1} \sum_{i=1}^{k} \overline{X}_{i.}$, y ${S_p}^2 = k^{-1} \sum_{i=1}^{k} s_i^2$ con

$s_i^2 = (n-1)^{-1} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_{i.} \right)^2$. A continuación tratemos de entender cuál es la intuición de la propuesta de este estadístico F.

Notemos que tanto cada grupo a analizar, representado por una muestra, como el sistema total, es decir, todos los datos observados, tienen su propia dispersión. Esta última puede ser explicada por la dispersión que hay **entre** cada grupo, y **dentro** de cada uno de ellos.

La dispersión dentro de los grupos se puede entender como la comparación de cada dato observado con la media muestral del grupo al que pertenece. Es decir,

$$\begin{aligned}\text{dispersión dentro grupos} &= \frac{\text{suma cuadrados dentro de grupos}}{\text{grados de libertad}} \\ &= \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_{i.}\right)^2}{k(n-1)} \\ &= k^{-1}\sum_{i=1}^{k} s_i^2 = S_p^2.\end{aligned}$$

La dispersión entre los grupos se puede entender como la comparación de cada media muestral con la media del sistema. Es decir,

$$\begin{aligned}\text{dispersión entre grupos} &= \frac{\text{suma cuadrados entre grupos}}{\text{grados de libertad}} \\ &= \frac{\sum_{i=1}^{k} n\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2}{k-1}.\end{aligned}$$

El ANOVA compara las dos dispersiones a través del cociente de las mismas. Si asumimos H_0 verdadera, este cociente debería tomar valores cercanos a 1.

$$\frac{\text{dispersión entre grupos}}{\text{dispersión dentro grupos}} = \frac{1}{k-1}\sum_{i=1}^{k}\frac{n\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2}{S_p^2} = F.$$

Si asumimos adicionalmente que las muestras en cada grupo son normales, se puede probar que el estadístico F tiene una distribución $F_{k-1,k(n-1)}$.

Veamos que bajo la hipótesis normalidad, F puede escribirse como el cociente de dos variables que siguen una distribución χ^2. Notemos que

$$F = \frac{\frac{1}{k-1}\sum_{i=1}^{k}\frac{n\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2}{\sigma^2}}{\frac{1}{k(n-1)}\sum_{i=1}^{k}\frac{\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_{i.}\right)^2}{\sigma^2}}$$

bajo las hipótesis de independencia y normalidad, se puede probar que $s_i^2 = \frac{\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_{i.}\right)^2}{\sigma^2}$ tiene distribución χ^2_{n-1} para $1 \le i \le k$ y $s_1^2, \ldots, s_k^2$ son independientes e independientes de $\overline{X}_{i.}$ luego

$$\sum_{i=1}^{k}\frac{\sum_{j=1}^{n}\left(X_{ij}-\overline{X}_{i.}\right)^2}{\sigma^2}$$

tiene distribución $\chi^2_{k(n-1)}$.

Por otro lado, la suma del numerador puede entenderse como si hubiésemos observado la i-ésima media muestral de cada grupo que bajo normalidad es también normal, pero con varianza σ^2/n y bajo la hipótesis nula con igual media Por lo tanto,

$$\sum_{i=1}^{k} \frac{\left(\overline{X}_{i.} - \overline{X}_{..}\right)^2}{\sigma^2/n},$$

tiene distribución χ^2_{k-1}, bajo la hipótesis nula.

Finalmente, del hecho que s_i^2 es independiente de $\overline{X}_{i.}$ también lo será el denominador del numerador de F, por lo tanto $F \sim \frac{\frac{\chi^2_{k-1}}{k-1}}{\frac{\chi^2_{k(n-1)}}{k(n-1)}} \sim F_{k-1,k(n-1)}$.

Es importante destacar a esta altura que esta distribución requiere fuertemente que los grupos sigan una distribución normal. Esta hipótesis fue esencial para obtener las distribuciones χ^2 en los párrafos anteriores. Diversos autores han discutido el caso en donde este supuesto no se cumple (véanse por ejemplo [4],[3],[5], y [6]).

En particular, en [6] se explica que si las muestras no siguen una distribución normal, pero tenemos un número alto de observaciones, por el Teorema Central del Límite podemos asumir que las medias muestrales siguen una distribución normal, y, por lo tanto podemos usar el ANOVA clásico sin preocuparnos por el supuesto de normalidad.

Resumiendo en términos más simples, podríamos imaginar el ANOVA como una herramienta que nos permite responder a la pregunta: "¿Las diferencias entre los grupos son más grandes que las diferencias dentro de los grupos?". Si las diferencias entre los grupos son grandes en comparación con las diferencias dentro de los grupos, entonces podríamos concluir que al menos uno de los grupos es estadísticamente diferente de los demás.

1.2. Test

En la sección anterior mostramos como se construye el estadístico del ANOVA, y la distribución que tiene bajo normalidad y asumiendo H_0 verdadera. Veamos ahora cómo diseñar un test de hipótesis para nuestro problema.

En general, los test de hipótesis consisten en definir un nivel de significación α, que es la probabilidad de cometer un error de tipo I y en base al cálculo del p-valor p, construir una región de rechazo.

Recordemos que el p-valor es la probabilidad de haber observado lo que se observó, o algo más extremo, asumiendo H_0 verdadera. La definición de extremo depende del contexto del problema a resolver (por ejemplo, si estamos diseñando un test unilateral o bilateral). Básicamente, rechazaremos H_0 si $\alpha \in [p, +\infty]$. Ahora bien, para este test

en particular, el estadístico del mismo es $F = \frac{1}{k-1}\sum_{i=1}^{k} \frac{n\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2}{{S_p}^2}$ que sigue una distribución $F_{k-1,k(n-1)}$ bajo H_0, y notaremos F_{obs} al valor del estadístico observado para la muestra. Como explicamos en 1.1, si las diferencias entre las medias de los grupos son significativas, esperamos notar que la variabilidad entre los grupos sea mayor que la variabilidad dentro de los grupos y por lo tanto, que el valor de F_{obs} sea grande. Entonces buscamos rechazar H_0 si el valor F_{obs} es grande. Así que en este contexto definimos al p-valor como:

$$p = P(F \geq F_{obs}).$$

Como ya mencionaos en la sección anterior, el ANOVA se basa en el supuesto de normalidad de los datos. Esto es especialmente notable cuando mostramos que el estadístico del test puede pensarse como cociente de distribuciones de la forma χ^2_m/m. Sin embargo, cuando el tamaño de la muestra es grande, podemos asumir normalidad de las medias muestrales debido al Teorema Central del Límite.

Otra alternativa posible cuando no sabemos si se cumple el supuesto de normalidad, pero sí se cumplen los supuestos de homogeneidad de las varianzas y la independencia de las observaciones, es usar el siguiente resultado publicado en Boos y Brownie (1995):

Proposición 1.2.1. Supongamos que tenemos k muestras independientes con varianzas iguales, y que cada muestra tiene n observaciones. Llamemos $X_{11} \dots X_{1n}, X_{21} \dots X_{2n}, \dots, X_{k1}, \dots, X_{kn}$ a las $N = kn$ variables. con $X_{ij} = \mu_i + \varepsilon_{ij}$ y ε_{ij} variables aleatorias iid, con media 0 y varianza $\sigma^2 < \infty$ Entonces bajo H_0 y para k fijo, se cumple que $(k-1)F \xrightarrow{d} \chi^2_{k-1}$, cuando $n \to \infty$.

Con esto, cuando el tamaño de las muestras es grande, podemos diseñar un test asintótico que nos ayude a decidir si hay evidencia suficiente para rechazar la hipótesis nula.

Como ya mencionamos anteriormente, para realizar un test de hipótesis necesitamos definir el nivel de significación del mismo (α), hallar el pvalor y con este último construir la región de rechazo. Ahora bien, para este test en particular, el estadístico es $T = (k-1)F$, que, para n grandes podemos asumir que tiene una distribución χ^2_{k-1}, y notaremos $T_{obs} = (k-1)F_{obs}$ al valor del estadístico observado para la muestra. Como explicamos en 1.1, si las diferencias entre las medias de los grupos son significativas, esperamos notar que la variabilidad entre los grupos sea mayor que la variabilidad dentro de los grupos y por lo tanto, que el valor de F observado sea grande. Entonces, como $T_{obs} = (k-1)F_{obs}$, buscamos rechazar H_0 si el valor T_{obs} es grande. Así que en este contexto definimos al pvalor como aquel que cumple lo siguiente:

$$p = P(T \geq T_{obs}).$$

Por todo lo desarrollado en esta sección, podemos notar que el no-cumplimiento de uno de los supuestos del ANOVA puede llevar a que el estudio de nuestro problema se

dificulte. Sin embargo, con algunas modificaciones se pueden crear otras herramientas que nos permitan resolverlo. En el caso anterior, con multiplicar al estadístico F por k-1 y diseñar un test asintótico podemos sortear el obstáculo de no tener normalidad en los datos. En la siguiente sección mencionaremos otras limitaciones, y posibles modificaciones para poder estudiar nuestro problema original, a pesar de que no se respeten todos los supuestos.

1.3. Limitaciones del test

Aunque el ANOVA es una herramienta estadística poderosa y ampliamente utilizada, también tiene algunas limitaciones importantes que deben tenerse en cuenta. En primer lugar, requiere supuestos específicos. ANOVA se basa en ciertos supuestos, como la normalidad de los datos, la homogeneidad de las varianzas y la independencia de las observaciones. Si estos supuestos no se cumplen, los resultados de ANOVA pueden ser poco confiables. Ya discutimos anteriormente que si carecemos de normalidad de las observaciones, podemos de cierto modo compensarlo si contamos con un número grande de observaciones. La validación y chequeo de supuestos es un paso importante antes de aplicar esta técnica.

Otra limitación es su sensibilidad a valores atípicos. ANOVA puede ser sensible a valores atípicos en los datos, especialmente cuando los tamaños de muestra son pequeños. Los valores atípicos pueden afectar significativamente la estimación de las medias y las varianzas, lo que a su vez puede influir en las conclusiones del análisis. Este es un tema que ha sido ampliamente tratado en la literatura a partir de la implementación de métodos de detección de outliers y alternativas robustas.

Otro punto débil al que se enfrenta ANOVA es la dificultad para manejar datos no balanceados. ANOVA asume que todas las muestras tienen el mismo tamaño, lo que puede ser problemático cuando los tamaños de muestra son desiguales. Existen variantes de ANOVA que pueden manejar datos no balanceados, introduciendo correcciones como la de Huynh-Feldt o Greenhouse-Geisser. Estas variantes ajustan los grados de libertad y las estimaciones de la varianza para tener en cuenta los desequilibrios en los tamaños de muestra.

También es importante señalar que ANOVA no proporciona información sobre las relaciones individuales. ANOVA se centra en la comparación de las medias de grupos, pero no proporciona información sobre las relaciones individuales entre las variables. Esto puede ser problemático si se desea comprender las relaciones específicas entre variables en detalle. Sin embargo, existen, por ejemplo, lo que se denominan comparaciones múltiples que permiten detectar en caso de rechazar la igualdad de medias cuáles de ellas difieren.

Finalmente, notemos que la propuesta clásica de ANOVA se basa en el hecho de que el número de grupos es fijo, es decir, no crece. Sin embargo, se puede considerar

un escenario donde el número de grupos puede ser muy grande, especialmente en estudios de ciertas disciplinas como la genética, la bioinformática, la clasificación de datos, entre otras. En particular, en estudios genéticos, el número de genotipos diferentes puede ser elevado, especialmente cuando se consideran variaciones genéticas en poblaciones grandes o diversificadas. En análisis de secuencias biológicas, como ADN o proteínas, el número de secuencias únicas puede ser enorme. Por ejemplo, en estudios de metagenómica que analizan comunidades microbianas, se pueden encontrar miles o incluso millones de secuencias únicas. En minería de texto o análisis de grandes conjuntos de documentos, el número de categorías o temas diferentes puede ser vasto, especialmente si se consideran todas las palabras o frases únicas en los documentos. En estudios de mercado o marketing, se pueden identificar una gran cantidad de segmentos de mercado diferentes basados en diversas características demográficas, geográficas o de comportamiento.

En los capítulos siguientes analizaremos si esta situación es una limitación para ANOVA.

Capítulo 2

Análisis de ANOVA para k grande

Como presentamos en la sección anterior, en muchos campos se hacen experimentos para comparar diferentes tratamientos y detectar si los tratamientos en media son iguales o difieren. Y para ello la técnica por excelencia es ANOVA. Este método funciona bien cuando se tienen muchos datos para cada tratamiento. Pero a veces, en ciertos contextos, es usual tener muchos tratamientos, pero pocos datos para cada uno de ellos.

En esta tesis se estudiará la propuesta introducida por Boos y Brownie (1995) en la que discuten cómo funcionan el test clásico de ANOVA cuando hay muchos tratamientos y pocos datos para cada uno. En este capítulo comentaremos cuál es el problema que nos interesa analizar, y describiremos las ideas que nos van a ayudar a resolverlo.

2.1. Estadístico del Test

Consideramos que tenemos el diseño de un solo factor o tratamiento, y que la asignación de los tratamientos a los grupos se realiza de manera completamente aleatoria. Llamemos X_{ij} a la j-ésima observación del grupo i, donde $1 \leq i \leq k$ y $1 \leq j \leq n$, es decir, supondremos que los tamaños de muestra son iguales en cada grupo. Además, supondremos que $E(X_{ij}) = \mu_i$ y $V(X_{ij}) = \sigma^2$ para $1 \leq i \leq k$ y $1 \leq j \leq n$, esto es, todos los grupos tienen igual varianza. Es decir, consideraremos la siguiente hipótesis:

S1 $X_{11} \dots X_{1n}, X_{21} \dots X_{2n} \dots X_{k1} \dots X_{kn}$ son $N = kn$ con $X_{ij} = \mu_i + \varepsilon_{ij}$ y ε_{ij} variables aleatorias iid, con media 0 y varianza $\sigma^2 < \infty$.

Estamos interesados en realizar un test para la siguiente hipótesis:

$$H_0 : \mu_1 = \mu_2 \cdots = \mu_k \text{ vs } H_1 : \exists\, r \neq s \text{ tal que } \mu_r \neq \mu_s$$

Consideremos el estadístico F de ANOVA definido en el capítulo anterior,

$$F = \frac{1}{k-1} \sum_{i=1}^{k} \frac{n\left(\overline{X}_{i.} - \overline{X}_{..}\right)^2}{{S_p}^2}$$

donde $\overline{X}_{i.} = n^{-1} \sum_{j=1}^{n} X_{ij}$; $\quad \overline{X}_{..} = k^{-1} \sum_{i=1}^{k} \overline{X}_{i.}$, y ${S_p}^2 = k^{-1} \sum_{i=1}^{k} s_i^2$ con $s_i^2 = (n-1)^{-1} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_{i.}\right)^2$.

En lo que sigue del capítulo desarrollaremos cuidadosamente la demostración del siguiente Teorema que fue estudiado en el trabajo de Boos y Brownie (1995).

Teorema 2.1.1. Bajo la hipótesis S1, si $E((X_{i1} - \mu_i)^4) \leq M$ para todo $1 \leq i \leq k$ y suponiendo que H_0 es cierta,

$$k^{1/2}(F-1) \xrightarrow{d} \mathcal{N}\left(0, \frac{2n}{n-1}\right),$$

si $k \to \infty$ y n está fijo.

Notemos que el Teorema 2.1.1 requiere una condición extra que es $E((X_{i1} - \mu_i)^4) \leq M$ para todo $1 \leq i \leq k$. Esta condición es necesaria para controlar la convergencia de ${S_p}^2$. Como veremos más adelante ${S_p}^2$ converge en probabilidad a σ^2 pero a diferencia de cuando el número de grupos está fijo, en este caso necesitamos controlar su varianza, lo que lleva a suponer una cota para el momento cuarto de las variables.

Antes de demostrar del teorema 2.1.1, introduciremos las siguientes identidades algebraicas, que nos serán útiles más adelante.

Proposición 2.1.2. Sean $Y_1, \ldots, Y_n$ una muestra de tamaño n, y sean μ y σ^2 constantes arbitrarias. Luego

a)

$$s^2 - \sigma^2 - \frac{1}{n} \sum_{i=1}^{n} \left[(Y_i - \mu)^2 - \sigma^2\right] = -\frac{2}{n(n-1)} \sum_{i<j} (Y_i - \mu)(Y_j - \mu). \qquad \text{(IA1)}$$

donde $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$.

b)

$$\left(\overline{Y} - \mu\right)^2 - \frac{s^2}{n} = \frac{2}{n(n-1)} \sum_{i<j} (Y_i - \mu)(Y_j - \mu). \qquad \text{(IA2)}$$

Notemos que para que estas igualdades tengan sentido, necesitamos que $n \geq 2$ pero esto no constituye una restricción cuando se trabaja con una muestra aleatoria.

Antes de demostrar esta proposición, observemos que si $Y_1, \ldots, Y_n$ son iid con media μ y varianza σ^2, el lado derecho de las igualdades IA1 y IA2

$$\frac{2}{n(n-1)} \sum_{i<j} (Y_i - \mu)(Y_j - \mu)$$

tiene esperanza 0 y varianza $\frac{2\sigma^4}{n(n-1)}$.

Usando la independencia de Y_i e Y_j para $i \neq j$ tenemos que

$$\mathbb{E}\left(\frac{2}{n(n-1)} \sum_{i<j} (Y_i - \mu)(Y_j - \mu)\right) = \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}\left[(Y_i - \mu)(Y_j - \mu)\right]$$

$$= \frac{2}{n(n-1)} \sum_{i<j} \mathbb{E}(Y_i - \mu)\,\mathbb{E}(Y_j - \mu) = 0$$

Ahora calculemos la varianza,

$$Var\left(-\frac{2}{n(n-1)} \sum_{i<j} (Y_i - \mu)(Y_j - \mu)\right) = \frac{1}{n^2(n-1)^2} Var\left(2\sum_{i<j} (Y_i - \mu)(Y_j - \mu)\right)$$

$$= \frac{1}{n^2(n-1)^2} Var\left(\sum_{i\neq j} (Y_i - \mu)(Y_j - \mu)\right)$$

$$= \frac{1}{n^2(n-1)^2} cov\left(\sum_{i\neq j} (Y_i - \mu)(Y_j - \mu), \sum_{s\neq t} (Y_s - \mu)(Y_t - \mu)\right)$$

$$= \frac{1}{n^2(n-1)^2} \sum_{i\neq j} \sum_{s\neq t} cov\left[(Y_i - \mu)(Y_j - \mu), (Y_s - \mu)(Y_t - \mu)\right].$$

Calculemos entonces $cov\left[(Y_i - \mu)(Y_j - \mu), (Y_s - \mu)(Y_t - \mu)\right]$ teniendo en cuenta que $i \neq j$ y que $s \neq t$. Notemos que

$$cov\left[(Y_i - \mu)(Y_j - \mu), (Y_s - \mu)(Y_t - \mu)\right]$$

$$= \mathbb{E}\left[(Y_i - \mu)(Y_j - \mu)(Y_s - \mu)(Y_t - \mu)\right] - \mathbb{E}\left[(Y_i - \mu)(Y_j - \mu)\right]\mathbb{E}\left[(Y_s - \mu)(Y_t - \mu)\right]$$

$$\underset{i\neq j, s\neq t}{=} \mathbb{E}\left[(Y_i - \mu)(Y_j - \mu)(Y_s - \mu)(Y_t - \mu)\right]$$

Veamos entonces las diferentes combinaciones de índices:

- Si i, j, s, t son todas distintas, $cov\left[(Y_i - \mu)(Y_j - \mu), (Y_s - \mu)(Y_t - \mu)\right] = 0$, porque las variables son iid.

- Sí $i = s$, y $j \neq t$, $cov\left[\left(Y_i - \mu\right)\left(Y_j - \mu\right), \left(Y_s - \mu\right)\left(Y_t - \mu\right)\right] = \mathbb{E}\left[\left(Y_i - \mu\right)^2 \left(Y_s - \mu\right)\left(Y_t - \mu\right)\right] = 0$, porque las variables son iid.
- Si $i = s$ y $j = t$, $cov\left[\left(Y_i - \mu\right)\left(Y_j - \mu\right), \left(Y_s - \mu\right)\left(Y_t - \mu\right)\right] = \mathbb{E}\left[\left(Y_i - \mu\right)^2 \left(Y_j - \mu\right)^2\right] = \left(\sigma^2\right)^2 = \sigma^4$
- Si $i = t$ y $j = s$, es análogo al caso anterior.

Por lo tanto,

$$Var\left(-\frac{2}{n(n-1)}\sum_{i<j}\left(Y_i - \mu\right)\left(Y_j - \mu\right)\right) =$$

$$\frac{1}{n^2(n-1)^2}\sum_{i\neq j}\sum_{s\neq t} cov\left[\left(Y_i - \mu\right)\left(Y_j - \mu\right), \left(Y_s - \mu\right)\left(Y_t - \mu\right)\right]$$

$$= \frac{1}{n^2(n-1)^2}\sigma^4\left(\#\{(i,j) : i = s, j = t, i \neq j\} + \#\{(i,j) : i = t, j = s, i \neq j\}\right)$$

$$= \frac{1}{n^2(n-1)^2}\sigma^4 2n(n-1) = \frac{2\sigma^4}{n(n-1)}.$$

Ahora sí, demostremos la proposición 2.1.2.

Demostración. Proposición 2.1.2:

Veamos primero la demostración de a). Reescribamos el lado izquierdo de la igualdad IA1

$$s^2 - \sigma^2 - \frac{1}{n}\sum_{i=1}^{n}\left[\left(Y_i - \mu\right)^2 - \sigma^2\right] = s^2 - \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \mu\right)^2.$$

Reemplazando s^2 por $\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ la igualdad anterior es equivalente a :

$$s^2 - \sigma^2 - \frac{1}{n}\sum_{i=1}^{n}\left[\left(Y_i - \mu\right)^2 - \sigma^2\right] = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \mu\right)^2$$

$$= \frac{1}{n(n-1)}\left(n\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 - (n-1)\sum_{i=1}^{n}\left(Y_i - \mu\right)^2\right).$$

Para probar IA1 nos alcanza con ver que

$$n\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 - (n-1)\sum_{i=1}^{n}\left(Y_i - \mu\right)^2 = -2\sum_{i<j}\left(Y_i - \mu\right)\left(Y_j - \mu\right). \tag{1}$$

Trabajemos ahora con el lado derecho de esta igualdad. Aplicando propiedad distributiva, y separando la suma:

$$-2\sum_{i<j}(Y_i-\mu)(Y_j-\mu) = -2\sum_{i<j}Y_iY_j + 2\mu\sum_{i<j}(Y_i+Y_j) - 2\mu^2\sum_{i<j}1. \tag{2}$$

Analicemos con más detalle $\sum_{i<j}(Y_i+Y_j)$, es decir, veamos cuántas veces aparece cada Y_m, con $1 \leq m \leq n$ en cada término de la serie:

- Cuando $i = m$, en la serie aparecen los términos $Y_m + Y_j$ con $m+1 \leq j \leq n$ (Y_m aparece $n-m$ veces)
- Cuando $j = m$, en la serie aparecen los términos $Y_i + Y_m$ con $1 \leq i \leq m-1$ (Y_m aparece $m-1$ veces)

Entonces, cada Y_m aparece $n-m+m-1 = n-1$ veces. Así que podemos afirmar que $\sum_{i<j}(Y_i+Y_j) = (n-1)\sum_{i=1}^{n}Y_i$.

Además, si tenemos en cuenta lo siguiente:

- $2\sum_{i<j}Y_iY_j = \sum_{i\neq j}Y_iY_j$.
- $\sum_{i<j}1 = \frac{n(n-1)}{2}$.

Y reemplazamos esta información en el lado derecho de (2):

$$-2\sum_{i<j}(Y_i-\mu)(Y_j-\mu) = -\sum_{i\neq j}Y_iY_j + 2\mu(n-1)\sum_{i=1}^{n}Y_i - 2\mu^2\frac{n(n-1)}{2}$$

$$= -\sum_{i\neq j}Y_iY_j + 2\mu(n-1)\sum_{i=1}^{n}Y_i - \mu^2 n(n-1).$$

Veamos que esta última expresión es exactamente la misma que conseguimos si desarrollamos el lado izquierdo de (1), y con esto terminamos de probar la igualdad IA1

$$n\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2 - (n-1)\sum_{i=1}^{n}(Y_i-\mu)^2 = n\sum_{i=1}^{n}\left(Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2\right) - (n-1)\sum_{i=1}^{n}\left(Y_i^2 - 2\mu Y_i + \mu^2\right)$$

$$= n\left(\sum_{i=1}^{n}Y_i^2 - 2n\overline{Y}^2 + n\overline{Y}^2\right) - (n-1)\left(\sum_{i=1}^{n}Y_i^2 - 2\mu n\overline{Y} + n\mu^2\right)$$

$$= \sum_{i=1}^{n} Y_i^2 - n^2\overline{Y}^2 + 2\mu n(n-1)\overline{Y} - n(n-1)\mu^2$$

$$= \sum_{i=1}^{n} Y_i^2 - \left(\sum_{i=1}^{n} Y_i^2 + \sum_{i \neq j} Y_i Y_j \right) + 2\mu(n-1) \sum_{i=1}^{n} Y_i - n(n-1)\mu^2$$

$$= -\sum_{i \neq j} Y_i Y_j + 2\mu(n-1) \sum_{i=1}^{n} Y_i - n(n-1)\mu^2.$$

Para finalizar esta demostración veamos b). Notemos que anteriormente demostramos que:

$$-2\sum_{i<j} (Y_i - \mu)(Y_j - \mu) = -\sum_{i \neq j} Y_i Y_j + 2\mu(n-1) \sum_{i=1}^{n} Y_i - \mu^2 n(n-1).$$

Usando esto, del lado derecho de IA2 tenemos:

$$\frac{2}{n(n-1)} \sum_{i<j} (Y_i - \mu)(Y_j - \mu) = \frac{1}{n(n-1)} \sum_{i \neq j} Y_i Y_j - \frac{2\mu}{n} \sum_{i=1}^{n} Y_i + \mu^2.$$

Veamos que esta última expresión es exactamente la misma que conseguimos si desarrollamos el lado izquierdo de IA2:

$$\left(\overline{Y} - \mu\right)^2 - \frac{s^2}{n} = \overline{Y}^2 - 2\mu\overline{Y} + \mu^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2$$

$$= \overline{Y}^2 - 2\mu\overline{Y} + \mu^2 - \frac{1}{n(n-1)} \left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 \right)$$

$$= \overline{Y}^2 \left(1 + \frac{1}{n-1} \right) - 2\mu\overline{Y} + \mu^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} Y_i^2$$

$$= \frac{1}{n(n-1)} \left(\sum_{i=1}^{n} Y_i^2 + \sum_{i \neq j} Y_i Y_j \right) - 2\mu\overline{Y} + \mu^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} Y_i^2$$

$$= \frac{1}{n(n-1)} \sum_{i \neq j} Y_i Y_j - \frac{2\mu}{n} \sum_{i=1}^{n} Y_i + \mu^2$$

que es a lo que queríamos llegar. □

Ahora, si estamos en condiciones del Teorema 2.1.1, la demostración se basa en la dada en [1], a la que le completaremos los detalles omitidos.

2.2. Demostración del Teorema 2.1.1

En primer lugar notemos que

$$k^{1/2}\left[F-1\right]=k^{1/2}\left[\frac{n}{{S_p}^2(k-1)}\sum_{i=1}^{k}\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2-\frac{n}{k{S_p}^2}\sum_{i=1}^{k}\frac{s_i^2}{n}\right]$$

$$=\frac{k^{1/2}\left[\frac{1}{k-1}\sum_{i=1}^{k}\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2-\frac{1}{k}\sum_{i=1}^{k}\frac{s_i^2}{n}\right]}{S_p^2/n}$$

$$=\frac{k^{1/2}\left[\frac{1}{k-1}\sum_{i=1}^{k}\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2-\frac{\sigma^2}{n}-\frac{1}{k}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]}{S_p^2/n}.$$

Veamos que

$$k^{1/2}\left[\frac{1}{k-1}\sum_{i=1}^{k}\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2-\frac{\sigma^2}{n}-\frac{1}{k}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]$$

$$=k^{-1/2}\sum_{i=1}^{k}\left[\left(\overline{X}_{i.}-\mu\right)^2-\frac{\sigma^2}{n}-\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]-\frac{2k^{1/2}}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right).$$

Para ello pensemos que $\overline{X}_{1.},\overline{X}_{2.},\ldots\overline{X}_{k.}$, hacen las veces de $Y_1,Y_2,\ldots Y_n$ en la proposición 2.1.2, y veamos que podemos aplicarla a nuestro problema (notar que en este caso tenemos una muestra de k elementos). Para esto, recordemos que $\mathbb{E}\left(\overline{X}_{i.}\right)=\mu$ y que $Var\left(\overline{X}_{i.}\right)=\frac{\sigma^2}{n}$ $\forall 1\leq i\leq k$. Usando la proposición a) tenemos que:

$$s^2-\frac{\sigma^2}{n}-\frac{1}{k}\sum_{i=1}^{k}\left[\left(\overline{X}_{i.}-\mu\right)^2-\frac{\sigma^2}{n}\right]=-\frac{2}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)\qquad(3)$$

donde $s^2=\frac{1}{k-1}\sum_{i=1}^{k}\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2$. Luego, resulta que

$$k^{1/2}\left[\frac{1}{k-1}\sum_{i=1}^{k}\left(\overline{X}_{i.}-\overline{X}_{..}\right)^2-\frac{\sigma^2}{n}-\frac{1}{k}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]=k^{1/2}\left[s^2-\frac{\sigma^2}{n}-\frac{1}{k}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]$$

$$=k^{1/2}\left[s^2-\frac{\sigma^2}{n}\right]-k^{-1/2}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right).$$

Así usando (3) tenemos

$$k^{1/2}\left[s^2-\frac{\sigma^2}{n}\right]-k^{-1/2}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)$$

$$= k^{1/2}\left[\frac{1}{k}\sum_{i=1}^{k}\left[\left(\overline{X}_{i.}-\mu\right)^2-\frac{\sigma^2}{n}\right]-\frac{2}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)\right]-k^{-1/2}\sum_{i=1}^{k}\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)$$

$$= k^{-1/2}\sum_{i=1}^{k}\left[\left(\overline{X}_{i.}-\mu\right)^2-\frac{\sigma^2}{n}-\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]-\frac{2k^{1/2}}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)$$

Que es lo que queríamos mostrar. A partir de la descomposición encontrada, llamemos $U = k^{-1/2}\sum_{i=1}^{k}\left[\left(\overline{X}_{i.}-\mu\right)^2-\frac{\sigma^2}{n}-\left(\frac{s_i^2}{n}-\frac{\sigma^2}{n}\right)\right]$ y $R = -\frac{2k^{1/2}}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)$. Notemos que

$$k^{1/2}(F-1) = \frac{U+R}{S_p^2/n}.$$

Para completar la demostración probaremos que U es asintóticamente normal y encontraremos el límite en probabilidad de R y S_p^2/n. Empecemos probando que $Var(R) = \frac{2\sigma^4}{(k-1)n^2}$.

$$Var(R) = Var\left(-\frac{2k^{1/2}}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)\right)$$

$$= kVar\left(\frac{2}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)\right).$$

Como $\overline{X}_{1.}, \overline{X}_{2.}, \dots \overline{X}_{k.}$ son iid, podemos usar la observación que detallamos luego de la proposición 2.1.2, y entonces podemos afirmar que

$$Var\left(\frac{2}{k(k-1)}\sum_{i<j}\left(\overline{X}_{i.}-\mu\right)\left(\overline{X}_{j.}-\mu\right)\right) = \frac{2\sigma^4}{n^2k(k-1)}.$$

Y, por lo tanto, $Var(R) = \frac{2\sigma^4}{n^2(k-1)}$ entonces cuando $k \to \infty$, $Var(R) = \frac{2\sigma^4}{n^2(k-1)} \to 0$. Por otra parte, notemos que $E(R) = 0$, luego usando la desigualdad de Chebychev tenemos qué $R \xrightarrow{p} 0$.

Estudiamos ahora el comportamiento asintótico de U cuando $k \to \infty$. Si asumimos $X_{ij} = \mu_i + \epsilon_{ij}$ con ϵ_{ij} independientes entonces bajo H0, que es condición del teorema, se tiene que $\overline{X}_{1.}, \overline{X}_{2.}, \dots \overline{X}_{k.}$ son iid.

Para cada i con $1 \le i \le k$, definimos $Z_i := \left(\overline{X}_{i.}-\mu\right)^2 - \frac{s_i^2}{n} = \left(\overline{\varepsilon}_{i.}\right)^2 - \frac{s_i^2}{n}$, donde $\overline{\varepsilon}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\varepsilon_{ij}$ y $\overline{\varepsilon}_{1.}, \dots \overline{\varepsilon}_{k.}$ son i.i.d. Además, como $s_i^2 = \sum_{j=1}^{n}(X_{ij}-\overline{X}_{i.})^2 = \sum_{j=1}^{n}(\varepsilon_{ij}-\overline{\varepsilon}_{i.})^2$ entonces los s_i^2 tambien son iid.

Luego, $U = k^{-1/2} \sum_{i=1}^{k} Z_i$. Si fijamos i, entonces $X_{i1}, X_{i2} \dots X_{in}$ hacen las veces de $Y_1, Y_2 \dots Y_n$ en la proposición 2.1.2. Por la observación mencionada luego de la Proposición 2.1.2 sabemos que:

$$\mathbb{E}(Z_i) = 0$$
$$Var(Z_i) = \frac{2\sigma^4}{n(n-1)}.$$

Además, como $Z_1, Z_1, \dots, Z_k$ también son iid, entonces por el Teorema Central del Límite cuando $k \to \infty$ sabemos que:

$$U = k^{-1/2} \sum_{i=1}^{k} Z_k \xrightarrow{d} \mathcal{N}\left(0, \frac{2\sigma^4}{n(n-1)}\right).$$

Entonces por el Teorema de Slutsky,

$$U + R \xrightarrow{d} \mathcal{N}\left(0, \frac{2\sigma^4}{n(n-1)}\right) \text{ cuando } k \to \infty.$$

Finalmente veamos que ,$\frac{S_p^2}{n} \xrightarrow{p} \frac{\sigma^2}{n}$cuando $k \to \infty$. Para ello veamos que $\mathbb{E}\left(\frac{S_p^2}{n}\right) = \frac{\sigma^2}{n}$ y $Var\left(\frac{S_p^2}{n}\right) \to 0$ cuando $k \to \infty$.

Recordemos que

${S_p}^2 = k^{-1} \sum_{i=1}^{k} s_i^2$, con $s_i^2 = (n-1)^{-1} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_{i.}\right)^2$; $1 \le i \le k$

$$\mathbb{E}\left(\frac{S_p^2}{n}\right) = \frac{1}{nk} \sum_{i=1}^{k} \mathbb{E}\left(s_i^2\right)$$

Re-escribiendo s_i^2:

$$s_i^2 = (n-1)^{-1} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_{i.}\right)^2 = (n-1)^{-1} \sum_{j=1}^{n} \left(X_{ij}^2 - 2X_{ij}\overline{X}_{i.} + \overline{X}_{i.}^2\right)$$
$$= (n-1)^{-1} \left(\sum_{j=1}^{n} X_{ij}^2 - 2n\overline{X}_{i.}^2 + n\overline{X}_{i.}^2\right) = (n-1)^{-1} \left(\sum_{j=1}^{n} X_{ij}^2 - n\overline{X}_{i.}^2\right)$$

Luego,

$$\mathbb{E}\left(s_i^2\right) = \frac{1}{n-1} \left(\sum_{j=1}^{n} \mathbb{E}\left(X_{ij}^2\right) - n\mathbb{E}\left(\overline{X}_{i.}^2\right)\right)$$

$$= \frac{1}{n-1}\left(\sum_{j=1}^{n}\left[Var\left(X_{ij}\right)+\mathbb{E}^2\left(X_{ij}\right)\right]-n\left[Var\left(\overline{X}_{i.}\right)+\mathbb{E}^2\left(\overline{X}_{i.}\right)\right]\right).$$

Como $\mathbb{E}\left(\overline{X}_{i.}\right)=\mu, Var\left(\overline{X}_{i.}\right)=\frac{\sigma^2}{n}$, la igualdad anterior se puede escribir como:

$$\frac{1}{n-1}\left(\sum_{j=1}^{n}\left[Var\left(X_{ij}\right)+\mathbb{E}^2\left(X_{ij}\right)\right]-n\left[Var\left(\overline{X}_{i.}\right)+\mathbb{E}^2\left(\overline{X}_{i.}\right)\right]\right)=$$

$$\frac{1}{n-1}\left(\sum_{j=1}^{n}\sigma^2+\mu^2-n\left[\frac{\sigma^2}{n}+\mu^2\right]\right)=$$

$$\frac{1}{n-1}\left(n\left(\sigma^2+\mu^2\right)-n\left[\frac{\sigma^2}{n}+\mu^2\right]\right)=\frac{n}{n-1}\left(\sigma^2-\frac{1}{n}\sigma^2\right)=\sigma^2.$$

Utilizando el cálculo anterior podemos concluir que:

$$\mathbb{E}\left(\frac{S_p^2}{n}\right)=\frac{1}{nk}\sum_{i=1}^{k}\mathbb{E}\left(s_i^2\right)=\frac{1}{nk}\sum_{i=1}^{k}\sigma^2=\frac{\sigma^2}{n}$$

Veamos ahora qué ocurre con la varianza:

$$Var\left(\frac{S_p^2}{n}\right)=\frac{1}{n^2k^2}\sum_{i=1}^{k}Var\left(s_i^2\right)$$

Se puede ver que $Var\left(s_i^2\right)=\frac{1}{n}(E(X_{i1}-\mu_i)^4-\sigma^4)+\frac{2}{n(n-1)}\sigma^4$ para todo $1\leq i\leq k$. Si denotamos $\nu_i=E(X_{i1}-\mu_i)^4$, $Var\left(s_i^2\right)=\frac{1}{n}\nu_i+\sigma^4\left(\frac{2}{n(n-1)}-\frac{1}{n}\right)$. Entonces

$$Var\left(\frac{S_p^2}{n}\right)=\frac{1}{n^3k^2}\sum_{i=1}^{k}\nu_i+\frac{1}{n^3k}\sigma^4\left(\frac{2}{(n-1)}-1\right).$$

Luego por hipótesis $\nu_i=E(X_{i1}-\mu_i)^4\leq M$ para todo $1\leq i\leq k$ entonces $k^{-1}\sum_{i=1}^{k}\nu_i<M$ y se tiene que

$$Var\left(\frac{S_p^2}{n}\right)\to 0,$$

cuando $k\to\infty$. De este modo, por la desigualdad de Chebychev,

$$\frac{S_p^2}{n}\xrightarrow{p}\frac{\sigma^2}{n}$$

cuando $k \to \infty$. Así usando el Teorema de Slutsky se obtiene

$$k^{1/2}(F-1) \xrightarrow{d} \mathcal{N}\left(0, \frac{2n}{n-1}\right)$$

cuando $k \to \infty$. Como queríamos demostrar.

2.3. Construcción del Test

En la sección anterior demostramos que bajo S1, y asumiendo H_0 cierta,

$$k^{1/2}(F-1) \xrightarrow{d} \mathcal{N}\left(0, \frac{2n}{n-1}\right),$$

si $k \to \infty$ y n está fijo.

Ahora utilizaremos esta información para diseñar un test de hipótesis asintótico que nos ayude a resolver nuestro problema, en caso en el que la cantidad de grupos sea grande.

Como ya desarrollamos en el capítulo anterior, para diseñar un test de hipótesis, necesitamos definir un nivel α (que es la probabilidad de cometer un error de tipo I), hallar el p-valor p, y construir una región de rechazo a partir de este último. Rechazaremos H_0 si $\alpha \in [p, +\infty]$.

Ahora bien, para este test en particular, el estadístico del mismo es $T = k^{1/2}(F-1)$, para el cual asumimos que asintóticamente sigue una distribución $\mathcal{N}\left(0, \frac{2n}{n-1}\right)$, y notaremos $T_{obs} = k^{1/2}(F_{obs}-1)$ al valor del estadístico observado para la muestra.

Como explicamos en 1.1, si hay diferencias entre las medias de lo grupos esperamos notar que la variabilidad entre los grupos sea mayor que la variabilidad dentro de los grupos y, por lo tanto, que el valor de F_{obs} sea grande. Entonces, bajo H_0 podemos pensar que F_{obs} debería tomar valores cercanos a uno y, por lo tanto, los valores de T_{obs} deberían ser cercanos a 0.

Por todo lo anterior, buscamos rechazar H_0 si el valor T_{obs} es muy grande o muy chico. Así que en este contexto definimos al p-valor como aquel que cumple lo siguiente:

$$p = 2P(T \geq |T_{obs}|).$$

donde T es asintitóticamente $\mathcal{N}\left(0, \frac{2n}{n-1}\right)$.

Hasta el momento, hemos presentado 3 tests de hipótesis para resolver nuestro problema: el ANOVA clásico, un test asintótico para el caso en el que la cantidad de muestras sea grande, y la cantidad de grupos sea un número fijo, y otro test asintótico en el que suponemos que la cantidad de grupos es grande, y la cantidad de muestras para

cada uno de ellos es un número fijo. Para el caso de los últimos 2, por la Proposición 1.2.1 y el Teorema 2.1.1 vimos que la distribución de los estadísticos dependen de k y n respectivamente. Es decir que, por ejemplo, no dependen de la distribución original que siguen los distintos grupos. Esto es particularmente interesante, considerando que el ANOVA clásico necesita, entre otros, del supuesto de normalidad de los datos para poder utilizarlo. En 1.3 ya mencionamos algunas limitaciones del ANOVA, y posibles correcciones al mismo, dependiendo de los supuestos que fallen, y de las características de las muestras. Esto nos abre un nuevo panorama: ante un determinado experimento, un analista puede optar por uno o varios de estos tests, dependiendo de los recursos con los que cuente. Por ejemplo, si quiere analizar la igualdad de medias de una muestra con pocos grupos, pero cuenta con la posibilidad de extraer muestras amplias de cada uno de ellos, podría optar por el ANOVA clásico, o por el test asintótico creado a partir de la proposición 1.2.1, siempre y cuando se asegure de que se cumplen los supuestos de independencia y homogeneidad de las varianzas.

Como para este trabajo nos concentramos en contextos en donde la cantidad de grupos sea grande, solo analizaremos dos de los tres test mencionados:

El primero, es el ANOVA clásico que utiliza como estadístico F y el cual sigue una distribución $F_{k-1,k(n-1)}$ y el segundo es el test asintótico que tiene como estadístico a $k^{1/2}(F-1)$, y converge en distribución a $\mathcal{N}\left(0, \frac{2n}{n-1}\right)$, cuando $k \to \infty$

Capítulo 3

Estudio de Simulación

En esta sección estudiaremos el comportamiento de los estimadores propuestos para diferentes conjuntos de datos generados bajo escenarios controlados. En lo que sigue de este trabajo, llamaremos "Test F"al test clásico de ANOVA, y "Test N"al test asintótico que se obtiene cuando el número de grupos aumenta.

El capítulo se divide en dos partes. En la primera nos concentramos en estudiar el error de Tipo I de ambos tests, y en la segunda a estudiar la potencia de los mismos.

3.1. Estudio de error de tipo I

En esta sección resumimos un estudio de simulación que diseñamos para comparar el comportamiento del error de Tipo I en ambos test. Para ello, tomamos k muestras independientes con n observaciones cada una, de tal forma que las muestras tengan la misma media. Luego, aplicamos los test a nivel 0,05 con el fin de saber si, para este nivel, la muestra da información significativa para rechazar la hipótesis nula.

Generamos 6 escenarios diferentes en el que consideramos diferentes distribuciones. En cada escenario replicamos el experimento 1000 o 10000 veces (dependiendo de la simulación), y calculamos el nivel empírico, es decir, las proporciones de rechazo, más precisamente contamos cuántas veces se rechazó H_0 dividido por el total de repeticiones. Si el test está bien calibrado, esperamos observar que el nivel empírico sea aproximadamente igual al nivel de significación (en este caso, 0,05). En cada escenario variamos los valores de k y de n. El principal objetivo de la simulación es ver el comportamiento de ambos tests, especialmente en los casos en que los valores de k sean cada vez más altos.

En las subsecciones siguientes presentaremos las distribuciones utilizadas junto con los resultados obtenidos.

Simulación 1

En este escenario consideramos $k = \{5, 10, 20, 50, 100, 200, 500, 1000\}$ y $n = \{5, 10, 15, 30, 50\}$. Las k muestras se generaron siguiendo una distribución $\mathcal{N}(0, 1)$. Repetimos cada experimento 10000 veces. Las tablas 3.1 y 3.2 resumen los resultados obtenidos.

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0476	0.0485	0.0503	0.053	0.0479	0.0481	0.0551	0.0518
10	0.0509	0.0518	0.0501	0.0509	0.0495	0.0489	0.0532	0.0528
15	0.0487	0.0458	0.0515	0.0488	0.0488	0.0494	0.052	0.0539
30	0.0492	0.0486	0.0469	0.0477	0.0504	0.0497	0.0497	0.049
50	0.0477	0.0521	0.0502	0.051	0.0516	0.0533	0.0498	0.0502

Tabla 3.1: Proporciones de rechazo del Test F con 10000 repeticiones

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0803	0.0658	0.0586	0.0549	0.0501	0.0493	0.0541	0.0509
10	0.0746	0.0593	0.0515	0.0512	0.0488	0.0495	0.0502	0.0532
15	0.0688	0.0516	0.0529	0.049	0.0484	0.0496	0.0533	0.0541
30	0.0653	0.0516	0.0476	0.0487	0.0502	0.0468	0.0503	0.0495
50	0.0613	0.0551	0.0494	0.0493	0.0503	0.0514	0.0477	0.0478

Tabla 3.2: Proporciones de rechazo del Test N con 10000 repeticiones

Simulación 2

En este caso continuamos bajo un modelo normal. Las k muestras siguen una distribución $\mathcal{N}(0, 1)$, pero sumamos más tamaños de muestras. $k = \{5, 10, 20, 50, 100, 200, 500, 1000\}$ y $n = \{5, 10, 15, 30, 50, 100, 200\}$. Aquí repetimos cada experimento 1000 veces. Los resultados obtenidos se encuentran en las tablas 3.3 y 3.4.

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0490	0.0530	0.0450	0.0490	0.0510	0.0430	0.0530	0.0480
10	0.0520	0.0580	0.0480	0.0470	0.0530	0.0550	0.0530	0.0510
15	0.0490	0.0520	0.0520	0.0430	0.0470	0.0480	0.0590	0.0440
30	0.0510	0.0450	0.0490	0.0490	0.0560	0.0550	0.0460	0.0520
50	0.0400	0.0600	0.0480	0.0440	0.0470	0.0370	0.0430	0.0390
100	0.0570	0.0460	0.0490	0.0560	0.0540	0.0510	0.0460	0.0510
200	0.0570	0.0540	0.0430	0.0530	0.0600	0.0420	0.0540	0.0600

Tabla 3.3: Proporciones de rechazo del Test F con 1000 repeticiones y distribución $\mathcal{N}(0, 1)$.

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0750	0.0660	0.0560	0.0470	0.0490	0.0550	0.0390	0.0540
10	0.0760	0.0630	0.0540	0.0460	0.0580	0.0550	0.0490	0.0560
15	0.0730	0.0560	0.0550	0.0490	0.0510	0.0510	0.0560	0.0480
30	0.0630	0.0480	0.0420	0.0510	0.0520	0.0510	0.0450	0.0590
50	0.0530	0.0630	0.0510	0.0460	0.0460	0.0450	0.0520	0.0390
100	0.0720	0.0500	0.0520	0.0560	0.0410	0.0490	0.0470	0.0530
200	0.0720	0.0580	0.0450	0.0550	0.0540	0.0380	0.0570	0.0590

Tabla 3.4: Proporciones de rechazo del Test N con 1000 repeticiones y distribución $\mathcal{N}(0, 1)$.

Simulación 3

En este escenario, trabajaremos con muestras que siguen una distribución exponencial. Consideramos $k = \{5, 10, 20, 50, 100, 200, 500, 1000\}$ y $n = \{5, 10, 15, 30, 50, 100, 200\}$. Las k muestras siguen una distribución $exp(\frac{1}{10})$. Repetimos cada experimento 1000 veces. Los resultados obtenidos se encuentran en las tablas 3.5 y 3.6

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0310	0.0600	0.0510	0.0540	0.0760	0.0590	0.0520	0.0590
10	0.0340	0.0460	0.0450	0.0470	0.0420	0.0620	0.0400	0.0490
15	0.0500	0.0400	0.0480	0.0460	0.0530	0.0450	0.0470	0.0550
30	0.0510	0.0430	0.0510	0.0490	0.0320	0.0510	0.0470	0.0470
50	0.0410	0.0420	0.0550	0.0500	0.0470	0.0400	0.0560	0.0500
100	0.0550	0.0580	0.0470	0.0570	0.0490	0.0450	0.0560	0.0430
200	0.0430	0.0580	0.0500	0.0510	0.0410	0.0320	0.0600	0.0420

Tabla 3.5: Proporciones de rechazo del Test F con 1000 repeticiones y distribución $exp(\frac{1}{10})$

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0720	0.0790	0.0580	0.0640	0.0630	0.0610	0.0540	0.0530
10	0.0530	0.0550	0.0440	0.0450	0.0420	0.0530	0.0430	0.0560
15	0.0670	0.0450	0.0480	0.0440	0.0530	0.0420	0.0490	0.0630
30	0.0660	0.0460	0.0460	0.0470	0.0380	0.0430	0.0440	0.0370
50	0.0530	0.0430	0.0560	0.0540	0.0440	0.0490	0.0520	0.0580
100	0.0730	0.0590	0.0410	0.0490	0.0520	0.0500	0.0600	0.0430
200	0.0500	0.0600	0.0540	0.0440	0.0440	0.0420	0.0590	0.0460

Tabla 3.6: Proporciones de rechazo del Test N con 1000 repeticiones y distribución $exp(\frac{1}{10})$

Simulación 4

A continuación, trabajamos con muestras que siguen una distribución beta. Consideramos $k = \{5, 10, 20, 50, 100, 200, 500, 1000\}$ y $n = \{5, 10, 15, 30, 50, 100, 200\}$. Las k muestras siguen una distribución $beta(2, 2)$. Repetimos cada experimento 1000 veces. Los resultados obtenidos se encuentran en las tablas 3.7 y 3.8

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0440	0.0530	0.0520	0.0590	0.0400	0.0430	0.0510	0.0360
10	0.0530	0.0540	0.0550	0.0420	0.0510	0.0560	0.0560	0.0470
15	0.0460	0.0480	0.0500	0.0400	0.0450	0.0510	0.0550	0.0430
30	0.0500	0.0680	0.0590	0.0510	0.0470	0.0360	0.0530	0.0530
50	0.0570	0.0500	0.0500	0.0570	0.0470	0.0380	0.0600	0.0500
100	0.0570	0.0360	0.0510	0.0490	0.0450	0.0540	0.0590	0.0590
200	0.0420	0.0470	0.0370	0.0680	0.0510	0.0410	0.0610	0.0600

Tabla 3.7: Proporciones de rechazo del Test F con 1000 repeticiones, y distribución $beta(2, 2)$.

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0830	0.0690	0.0600	0.0670	0.0370	0.0420	0.0470	0.0420
10	0.0770	0.0620	0.0550	0.0430	0.0560	0.0560	0.0500	0.0480
15	0.0640	0.0610	0.0520	0.0440	0.0430	0.0550	0.0570	0.0480
30	0.0670	0.0710	0.0640	0.0480	0.0380	0.0330	0.0440	0.0530
50	0.0730	0.0510	0.0490	0.0510	0.0450	0.0310	0.0440	0.0510
100	0.0740	0.0390	0.0510	0.0470	0.0530	0.0450	0.0600	0.0540
200	0.0540	0.0500	0.0500	0.0640	0.0610	0.0380	0.0490	0.0530

Tabla 3.8: Proporciones de rechazo del Test N con 1000 repeticiones, y distribución $beta(2, 2)$.

Simulación 5

En este escenario, trabajaremos con muestras que siguen una distribución normal logarítmica. Consideramos $k = \{5, 10, 20, 50, 100, 200, 500, 1000\}$ y $n = \{5, 10, 15, 30, 50, 100, 200\}$. Las k muestras siguen una distribución $lognormal(0, 1)$. Repetimos cada experimento 1000 veces. Los resultados obtenidos se encuentran en las tablas 3.9 y 3.10

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0310	0.0450	0.0390	0.0460	0.0600	0.0470	0.0540	0.0530
10	0.0400	0.0470	0.0450	0.0490	0.0520	0.0410	0.0660	0.0620
15	0.0390	0.0420	0.0530	0.0550	0.0380	0.0430	0.0630	0.0410
30	0.0330	0.0420	0.0500	0.0540	0.0610	0.0420	0.0440	0.0510
50	0.0450	0.0480	0.0520	0.0500	0.0430	0.0440	0.0470	0.0500
100	0.0530	0.0430	0.0530	0.0420	0.0560	0.0490	0.0650	0.0500
200	0.0510	0.0590	0.0420	0.0540	0.0540	0.0500	0.0600	0.0570

Tabla 3.9: Proporciones de rechazo del Test F con 1000 repeticiones, y distribución $lognormal(0, 1)$

$n \setminus k$	5	10	20	50	100	200	500	1000
5	0.0510	0.0550	0.0420	0.0400	0.0490	0.0420	0.0490	0.0540
10	0.0580	0.0530	0.0450	0.0460	0.0520	0.0400	0.0590	0.0500
15	0.0560	0.0470	0.0510	0.0430	0.0340	0.0430	0.0600	0.0400
30	0.0520	0.0470	0.0480	0.0470	0.0470	0.0390	0.0400	0.0570
50	0.0620	0.0510	0.0510	0.0440	0.0480	0.0290	0.0480	0.0560
100	0.0660	0.0440	0.0550	0.0430	0.0540	0.0520	0.0580	0.0490
200	0.0670	0.0600	0.0360	0.0530	0.0460	0.0420	0.0550	0.0580

Tabla 3.10: Proporciones de rechazo del Test N con 1000 repeticiones, y distribución $lognormal(0, 1)$

Simulación 6

En este caso continuamos bajo un modelo normal logarítmico. Las k muestras siguen una distribución $lognormal(0, 1)$, pero repetimos cada experimento 10000 veces. Consideramos $k = \{5, 10, 50, 100, 200, 500, 1000\}$ y $n = \{5, 10, 15, 30, 50, 100, 200\}$. Los resultados obtenidos se encuentran en las tablas 3.11 y 3.12

$n \setminus k$	5	10	50	100	200	500	1000
5	0.0314	0.0386	0.0488	0.0481	0.0506	0.0545	0.0562
10	0.0347	0.0412	0.0512	0.0500	0.0522	0.0512	0.0505
15	0.0378	0.0428	0.0481	0.0523	0.0507	0.0477	0.0506
30	0.0405	0.0446	0.0523	0.0518	0.0547	0.0477	0.0533
50	0.0391	0.0481	0.0503	0.0521	0.0538	0.0446	0.0502
100	0.0442	0.0431	0.0479	0.0493	0.0494	0.0513	0.0488
200	0.0420	0.0461	0.0487	0.0558	0.0506	0.0507	0.0486

Tabla 3.11: Proporciones de rechazo del Test F con 10000 repeticiones, y distribución $lognormal(0, 1)$

$n \setminus k$	5	10	50	100	200	500	1000
5	0.0556	0.0536	0.0463	0.0435	0.0451	0.0459	0.0499
10	0.0564	0.0474	0.0470	0.0460	0.0458	0.0481	0.0474
15	0.0538	0.0478	0.0441	0.0504	0.0446	0.0479	0.0474
30	0.0563	0.0471	0.0493	0.0499	0.0519	0.0478	0.0519
50	0.0523	0.0502	0.0477	0.0487	0.0521	0.0453	0.0515
100	0.0550	0.0455	0.0497	0.0462	0.0504	0.0512	0.0504
200	0.0552	0.0485	0.0501	0.0534	0.0479	0.0482	0.0471

Tabla 3.12: Proporciones de rechazo del Test N con 10000 repeticiones, y distribución $lognormal(0, 1)$

3.1.1. Conclusiones

A partir de todos los experimentos realizados, podemos notar algunos aspectos en común. Observando las tablas a grandes rasgos notamos que el error empírico de tipo I de ambos tests está alrededor del 5 %, alcanzando en general el valor nominal del test.

En los primeros dos escenarios los resultados tienen un mejor desempeño para el test Clásico de ANOVA, esto se debe a la normalidad de las muestras. En cierto modo no importa cuan grandes o chicos sean k y n, bajo normalidad, el estadístico F tendrá distribución de Fisher y por lo tanto el test de ANOVA resulta satisfactorio. En los otros escenarios, se puede ver que valores de n chicos arrojan errores empíricos bajos para el ANOVA clásico, mientras que para el "Test N "los errores son más altos.

Como comentamos anteriormente, no pareciera que el ANOVA clásico se encontrará con grandes limitaciones frente al aumento del tamaño de grupos. Más bien parece que en la situación que se corresponde a la falta de normalidad, tener alto tamaño de grupos en cierto modo ayuda. Sí se advierte que en distribuciones como la considerada en el

escenario 6 que son altamente asimétricas, el test F presenta un desempeño más pobre; sin embargo, el test N alcanza un comportamiento bastante superior.

En conclusión, si solo nos guiamos por los datos obtenidos por estos experimentos, parecería no haber muchas diferencias entre ambos tests. Para analizar los estadísticos desde otro punto de vista, en la siguiente sección trabajemos con la función de potencia.

3.2. Estudio de la potencia

En esta sección analizaremos para muestras finitas como es el comportamiento de las funciones de potencia en ambos test. Para ello presentamos los estudios de simulación que detallaremos a continuación.

Como en la sección anterior, tomamos k muestras independientes con n observaciones cada una, con la diferencia que para estudiar la potencia, no tomaremos todos los grupos con la misma media, sino que variaremos un 20 % de cada una de ellas, de tal forma que la media cambie progresivamente. Notemos que con estos cambios sabemos que lo correcto sería rechazar H_0. Pero modificando este porcentaje nos permitirá estudiar la sensibilidad del test frente a diferentes hipótesis alternativas.

A modo de ejemplo, podríamos tomar 10 grupos de los cuales 8 grupos siguen una distribución $\mathcal{N}(0,1)$, y los 2 restantes una distribución $\mathcal{N}(0,1)+\delta$, con $\delta \in \{0.1, 0.2, 0.5, 1\}$. Sería de esperar que el test sepa distinguir estas diferencias en las medias, pero seguramente a medida que δ se acerque más al cero será más complicado detectar la diferencia, mientras que para valores grandes de δ el test debería ser más sensible a estos cambios.

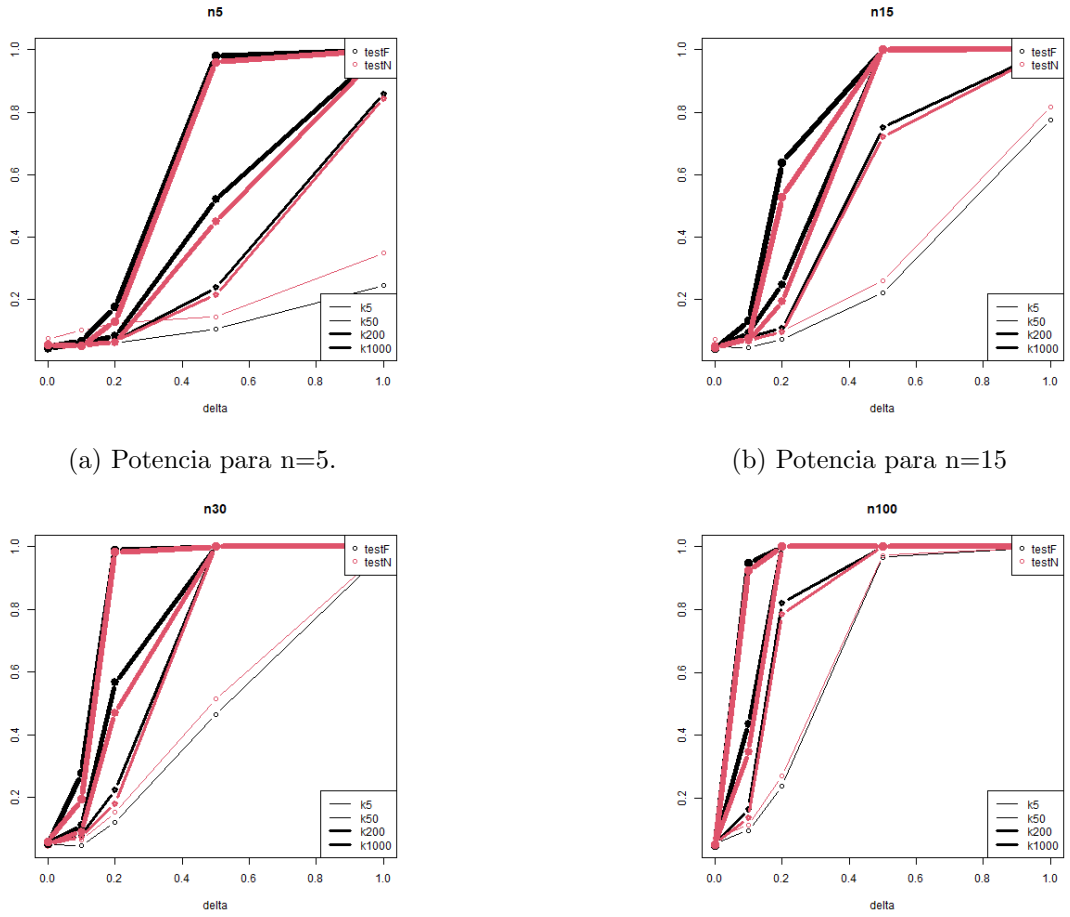
Luego, aplicamos los tests a nivel 0,05 con el fin de saber si, para este nivel, la muestra da información significativa para rechazar la hipótesis nula. Repetimos estos experimentos 1000 veces, y calculamos las proporciones de rechazo del mismo modo que en la sección anterior.

Finalmente, utilizaremos estos datos para ilustrar la función de potencia. Como nos interesa que este gráfico quede lo más completo posible, además de trabajar con muestras que no cumplan con H_0, también agregaremos a la simulación el caso donde $\delta = 0$ (es decir, el caso en donde lo correcto es aceptar H_0). Al igual que en la sección anterior realizaremos este experimento variando los valores de k y n, pero poniendo especial atención en que los valores de k sean cada vez más altos.

En todos los escenarios estudiados observaremos para cada valor de n un gráfico que debe interpretarse de la siguiente forma: Las funciones del test F y test N están representadas por los colores negro y rojo, respectivamente. En el mismo gráfico se verán ambas funciones para varios valores de k. Las que provienen de haber trabajado con el mismo valor de k tendrán el mismo grosor (cuanto mayor es el grosor, mayor es el valor de k seleccionado).

Simulación 1

El primer escenario que estudiaremos corresponde a tomar $k = \{5, 50, 200, 1000\}$ y $n = \{5, 15, 30, 100\}$. Generaremos el 20 % de los k grupos siguiendo una distribución $\mathcal{N}(0, 1) + \delta$, con $\delta \in \{0, 0.1, 0.2, 0.5, 1\}$. Notar que, en este caso, suponemos que H_0 es verdadera si todas las muestras tienen media 0 (caso $\delta = 0$). Repetimos cada experimento 1000 veces, obteniendo los siguientes resultados: Cada uno de los gráficos de la Figura 3.1, resumen un valor de n y varían los valores de k y δ.



(a) Potencia para n=5.

(b) Potencia para n=15

(c) Potencia para n=30

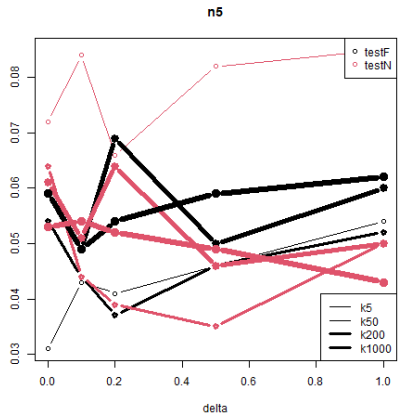
(d) Potencia para n=100

Figura 3.1: Potencias con distribución $\mathcal{N}(0, 1) + \delta$

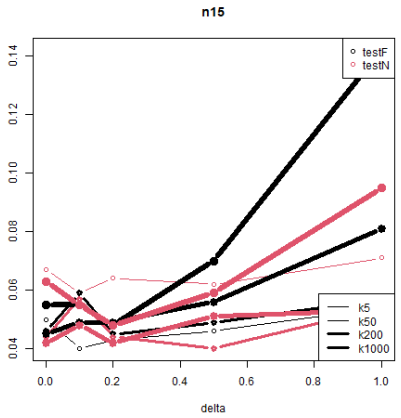
Simulación 2

El escenario que describiremos ahora corresponde a tomar el 20 % de los k grupos generados con una distribución $exp(1/10) + \delta$, con $\delta \in \{0, 0.1, 0.2, 0.5, 1\}$, mientras

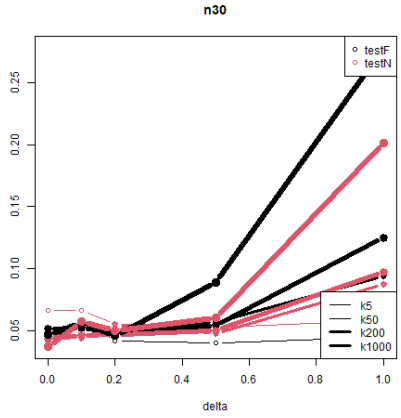
que el 80 % restante sigue una distribución $exp(1/10)$. De este modo, el caso $\delta = 0$ corresponde a H_0 verdadera. Repetimos cada experimento 1000 veces y consideramos $k = \{5, 50, 200, 1000\}$ $n = \{5, 15, 30, 100\}$. Los resultados para cada n variando k y δ se resumen en la Figura 3.2



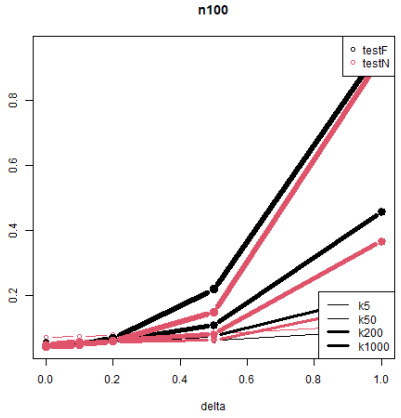
(a) Potencia para n=5.



(b) Potencia para n=15



(c) Potencia para n=30



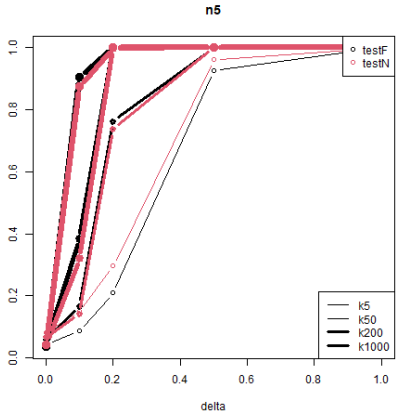
(d) Potencia para n=100

Figura 3.2: Potencias con distribución $exp(1/10) + \delta$

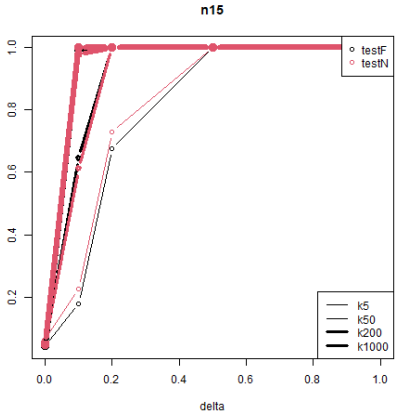
Simulación 3

A continuación, describiremos el escenario correspondiente a tomar el 20 % de los k grupos generados con una distribución $beta(2, 2)+\delta$, con $\delta \in \{0, 0.1, 0.2, 0.5, 1\}$, mientras que el 80 % restante sigue una distribución $beta(2, 2)$. De este modo, el caso $\delta = 0$ corresponde a H_0 verdadera. Repetimos cada experimento 1000 veces y consideramos $k = \{5, 50, 200, 1000\}$ $n = \{5, 15, 30, 100\}$. Los resultados para cada n variando k y δ se

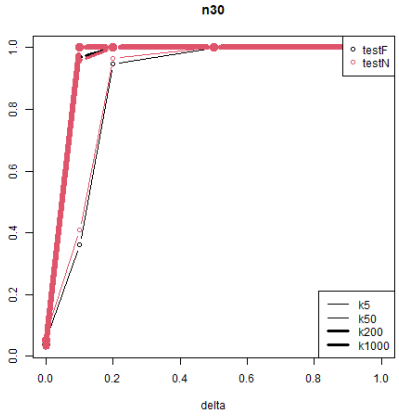
resumen en la Figura 3.3



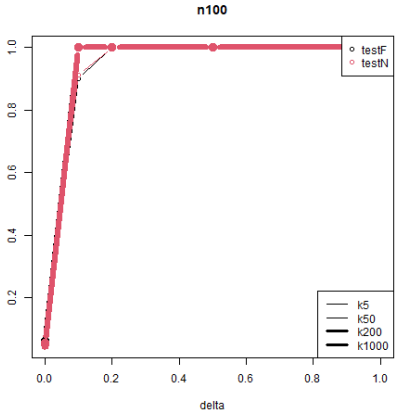
(a) Potencia para n=5.



(b) Potencia para n=15



(c) Potencia para n=30

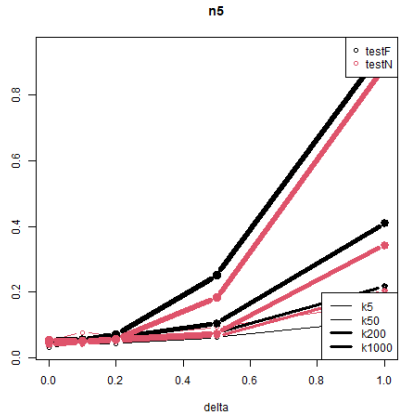


(d) Potencia para n=100

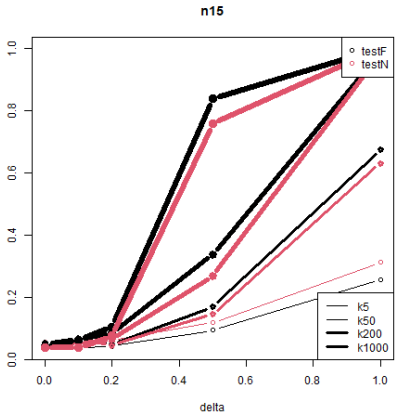
Figura 3.3: Potencias con distribución $beta(2,2)+\delta$

Simulación 4

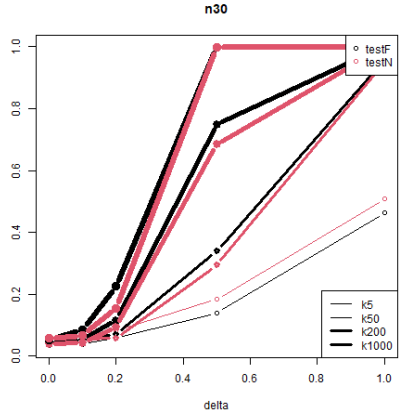
Aquí describiremos el escenario correspondiente a tomar el 20 % de los k grupos generados con una distribución $lognormal(0,1)+\delta$, con $\delta \in \{0, 0.1, 0.2, 0.5, 1\}$, mientras que el 80 % restante sigue una distribución $lognormal(0,1)$. De este modo, el caso $\delta = 0$ corresponde a H_0 verdadera. Repetimos cada experimento 1000 veces y consideramos $k = \{5, 50, 200, 1000\}$ $n = \{5, 15, 30, 100\}$. Los resultados para cada n variando k y δ se resumen en la Figura 3.4



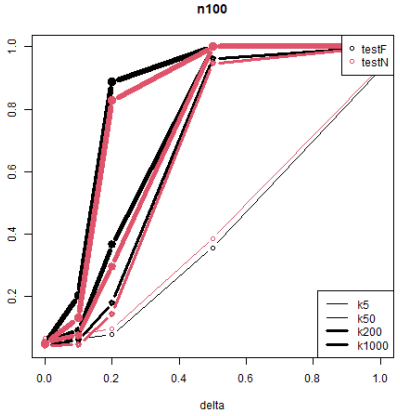
(a) Potencia para n=5.



(b) Potencia para n=15



(c) Potencia para n=30

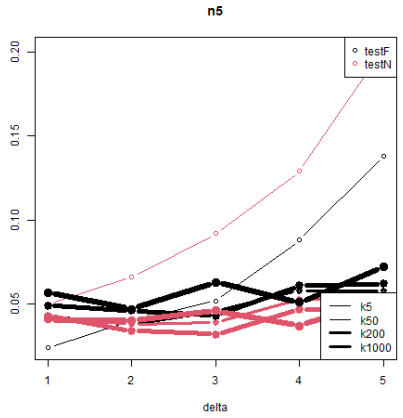


(d) Potencia para n=100

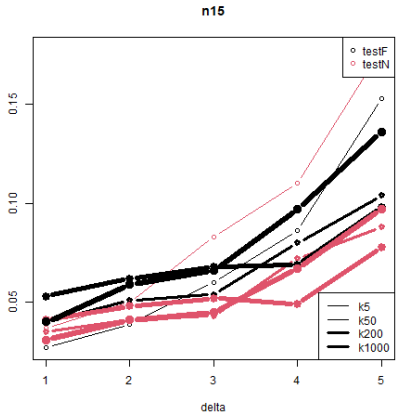
Figura 3.4: Potencias con distribución $lognormal(0,1)+\delta$

Simulación 5

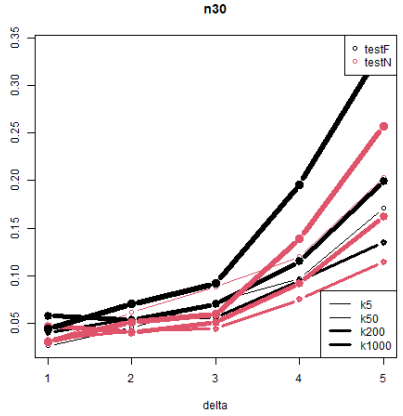
Por último, describiremos el escenario correspondiente a tomar el 20 % de los k grupos generados con una distribución $lognormal(0,2)+\delta$ con $\delta \in \{1,2,3,4,5\}$, mientras que el 80 % restante sigue una distribución $lognormal(0,2)$. Repetimos cada experimento 1000 veces y consideramos $k = \{5,50,200,1000\}$ $n = \{5,15,30,100\}$. Notar que en este caso decidimos tomar valores de δ mayores a los utilizados en las simulaciones anteriores. Los resultados para cada n variando k y δ se resumen en la Figura 3.5



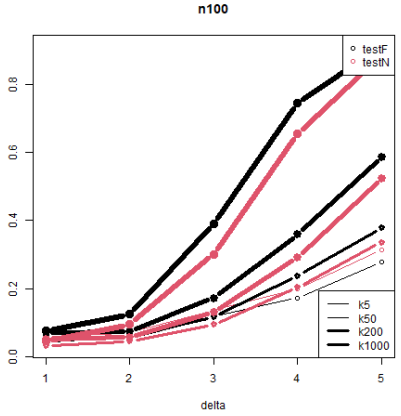
(a) Potencia para n=5.



(b) Potencia para n=15



(c) Potencia para n=30



(d) Potencia para n=100

Figura 3.5: Potencias con distribución $lognormal(0, 2) + \delta$

3.2.1. Conclusiones

Los gráficos de la función de potencia en los diferentes escenarios resumen varios puntos en común, y varias singularidades. Para valores de δ cercanos a 0, y sin importar la cantidad de grupos, las proporciones de rechazo son bajas, y cercanas a la proporción de error de tipo I. Esto parece indicarnos que los test no son sensibles a pequeños cambios. Sin embargo, a medida que los valores de δ crecen y, por lo tanto, la media muestral de los grupos alterados (el 20 % de los grupos), se aleja de la media que entendemos como verdadera, las proporciones de rechazo aumentan significativamente. Vemos que esta proporción mejora aun más para ambos tests a medida que el valor de k es grande.

En el caso particular de la Simulación 5, podemos ver que para $\delta = 1$ las proporciones de rechazo no tienen valores tan altos como en el resto de las simulaciones, por tal motivo

experimentamos con mayores valores de δ para ver mejoras. Igualmente, esta mejoría es notable solo para el caso $n = 100$, lo que nos lleva a pensar que para ver proporciones de rechazo más altas en el resto de los casos, deberíamos generar alternativas con valores de δ mayores a los utilizados en la simulación en cuestión. Es decir que para ver un mejor comportamiento de los test, la media de las muestras perturbadas deben alejarse mucho más de la media verdadera, evidenciando cierta falta de sensibilidad para detectar alternativas cercanas a la hipótesis nula.

Para valores fijos de k y n, ambos test parecen comportarse de forma similar. Las mayores diferencias se notan para valores chicos de n. En el primer escenario bajo el modelo normal, es interesante notar como aumentar el tamaño de los grupos redunda en una mejora de la potencia. En la Simulación 2 con tamaño de muestra chico, el comportamiento de ambos test no es bueno. Aun cuando el tamaño de los grupos es grande, el test no logra discriminar.

En los casos con $k = 5$, el test N pareciera tener un mejor comportamiento que el Test F, en el sentido de que rechaza más casos en donde sabemos que H_0 es falsa. Esta diferencia es muy sutil, y no alcanza para decir que un test es mejor que otro cuando se trabaja con una cantidad pequeña de grupos.

En lineas generales podemos decir que el comportamiento de la potencia es similar al estudio realizado para el error de tipo I. Ambos test parecieran tener comportamientos parecidos salvo algunas excepciones puntuales.

Conclusiones del trabajo de tesis

En esta tesis abordamos el problema ANOVA pero en el contexto que el número de grupos puede ser grande, incluso mayor que el tamaño de muestra en cada grupo. A priori los resultados para ANOVA asumen normalidad y el estadístico tiene una distribción paramétrica conocida. Sin el supuesto de normalidad se puede obtener el comportamiento asintótico también obteniendo una distribución límite conocida. En ambos casos esta distribción depende del número de grupos, por lo tanto resulta interesante estudiar como se comporta el estadístico F bajo este supuesto. Este tema fue abordado por primera vez en el trabajo de Boos y Brownie (1995). Ellos brindan una demostración bastante escueta del comportamiento asintótico y sin considerar una hipótesis necesaria. El trabajo no incluyó experimentos computacionales.

En este trabajo nos concentramos en estudiar, en primer lugar el test de ANOVA que no forma parte de los contenidos de ninguna materia de la carrera. Y en una segunda etapa estudiamos el trabajo de Boos y Brownie (1995). Esto incluyo completar los detalles de la demostración del teorema principal. Por otra parte realizamos un breve estudio de simulación en la que comparamos el estimador clásico de ANOVA con la propuesta considerada en el trabajo de Boos y Brownie.

A modo de resumen y luego de analizar todos los experimentos realizados, explicaremos algunas de las causas de lo observado. En primer lugar, por lo visto en 3.1, podemos ver que ambos test controlan bien el error de tipo I. En cuanto a las funciones de potencia, podemos ver que los tests tienen un mejor comportamiento cuanto mayores son los valores de k. Con esto queremos decir que, dado n fijo, si miramos lo que ocurrió con las funciones de potencia para un valor de δ pequeño, notamos que, a medida que aumenta la cantidad de grupos, el valor de la potencia se acerca a 1. Al menos para el test, N esto puede deberse a que, al ser un test asintótico para k, mejora cuanto más grupos tenemos para testear.

También notamos que para valores de δ cercanos a 0, la función de potencia no parece mejorar a medida que aumentamos el valor de n. Gráficamente notamos que la imagen de la función de potencia para valores de $\delta \leq 0.2$ se mueve en los mismos rangos, sin importar si tomamos 5 o 1000 muestras de cada grupo. Es decir que a los tests les cuesta discernir pequeños apartamientos de H_0.

En general, vemos que sin importar la distribución que siguieron los grupos, en los casos en donde hubieron una gran cantidad de ellos, ambos tests se comportaron de forma similar. Si recordamos que el ANOVA clásico se rige bajo el supuesto de normalidad de los grupos, llama la atención que hayamos visto tanta similitud entre los tests. Sin embargo, esto es coherente con lo que se explica en Boos y Brownie (1995) y que detallamos en 2.1. La demostración del Teorema 2.1.1, no supone ninguna hipótesis sobre qué tipo de distribución siguen los distintos grupos. Solo partimos de suponer que las variables son iid, con la misma media y varianza, con cuarto momento finito, y que las muestras de cada grupo tienen el mismo tamaño. En ese caso, a nivel asintótico, la distriubción del estadístico de lo que llamamos "Test N" $\left(k^{1/2}(F-1)\right)$ solo depende del tamaño de las muestras n.

De aquí podemos deducir que, a nivel asintótico, F tampoco depende de las distribuciones que siga cada grupo, y es por esto que ambos tests terminan por parecerse. No importando demasiado las características de las distribuciones de cada grupo.

Como conclusión final, el trabajo realizado en esta tesis no determina un mal diagnóstico o una limitación clara al usar el test clásico de ANOVA cuando el número de grupos es grande. El test introducido por Boss y Brownie es tan simple como el ANOVA clásico y arroja resultados parecidos. Queda para un futuro trabajo extender el estudio de simulación y analizar los resultados asintóticos cuando tanto k como n aumentan.

Bibliografía

[1] Boos, Dennis D y Brownie, Cavell. ANOVA and rank tests when the number of treatments is large. *Statistics & Probability Letters* (1995).

[2] Fisher, R.A. *Statistical Methods for Research Workers*. Oliver y Boyd, Edinburgh, 1925.

[3] Glass, Gene V, Peckham, Percy D. y Sanders, James R. Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research* 42 (1972), págs. 237-288.

[4] Harwell, Michael R., Rubinstein, Elaine N., Hayes, William S. y Olds, Corley C. Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics* 17.4 (1992), págs. 315-339. URL: `http://www.jstor.org/stable/1165127`.

[5] Lix, Lisa M., Keselman, Joanne C. y Keselman, H. J. Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test. *Review of Educational Research* 66 (1996), págs. 579-619.

[6] Norman, Geoffrey R. y Streiner, David L. *Biostatistics: The Bare Essentials. Third Edition*. B.C. Decker Inc, 2008, pág. 80.

[7] Skidmore, Susan y Thompson, Bruce. Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior research methods* 45 (oct. de 2012). DOI: `10.3758/s13428-012-0257-2`.

[8] Student. The Probable Error of a Mean. *Biometrika* 6.1 (1908), págs. 1-25.