UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE CIENCIAS BÁSICAS

Pruebas de hipótesis para la matriz de covarianza poblacional de datos de dimensión alta

Tesis

Para obtener el título de
Maestro en Ciencias en Matemáticas Aplicadas

Presenta:
L.M. Didier Cortez Elizalde

Directores de tesis:
Dra. Addy Margarita Bolívar Cimé
Dr. Víctor Manuel Pérez-Abreu Carrión

Cunduacán, Tabasco, México. **Marzo 2020**

DIRECCIÓN

28 de febrero de 2020

Lic. Didier Cortez Elizalde
Pasante de la Maestría en Ciencias
en Matemáticas Aplicadas
Presente.

Por medio del presente y de la manera mas cordial, me dirjo a Usted para hacer de su conocimiento que proceda a la impresión del trabajo titulado "**Pruebas de hipótesis para la Matriz de covarianza poblacional de datos de dimensión alta**", en virtu de que reúne los requisitos para el EXAMEN PROFESIONAL para obtener el grado de Maestro en Ciencias en Matemáticas Aplicadas.

Sin otro particular, reciba un cordial saludo.

Atentamente

Dr. Gerardo Delgadillo Piñón
Director

DIVISIÓN ACADÉMICA DE CIENCIAS BÁSICAS

C.c.p.- Archivo
Dr.GDP/Dr'MAVR/emt

CARTA DE AUTORIZACIÓN

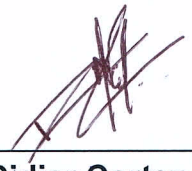
El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de grado denominada "**Pruebas de hipótesis para la matriz de covarianza poblacional de datos de dimensión alta**", de la cual soy autor y titular de los Derechos de Autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa mas no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra red académica con las que la Universidad tenga relación institucional.

Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en este documento.

Se firma la presente autorización en el municipio de Cunduacán, Tabasco a los 28 días del mes de febrero del año 2020

AUTORIZÓ



L.M. Didier Cortez Elizalde

Agradecimientos

Primero quiero agradecer a Dios por darme la oportunidad de vivir y permitirme haber llegado a este momento tan importantes de mi vida.

Quiero expresar mi más profundo agradecimiento a los miembros del jurado revisor por su disposición y voluntad para la pronto revisión de la tesis así como por sus valiosos comentarios que me permitieron mejorar el contenido del mismo.

Quiero agradecer a mi madre la **Sra. Blanca Emma Elizalde Velazquez** por brindarme su amor y apoyo incondicional durante todo este tiempo, es ella a quien dedico todos mis logros. A mis hermanos **Dania Cortez Elizalde, Jorge Cortez Elizalde, Lorena Cortez Elizalde, Angel Eduardo Meneses Elizalde,** gracias por estar presente en cada día de mi vida y a mi familia en general gracias por su apoyo.

A la **División Académica de Ciencias Básicas de la UJAT**, por darme conocimientos para desarrollarme como profesionista y todas las facilidades prestadas para realizar este trabajo.

Agradezco al **Centro de investigación en matemáticas A.C. (Cimat)** por haberme aceptado durante la estancia, el tiempo ahí ayudo mi a mi formación académica y profesional.

A la **Dra. Addy Margarita Bolívar Cimé** y **Dr.Víctor Manuel Pérez Abreu** mis directores de tesis, gracias por toda la motivación, enseñanza y apoyo recibido durante este tiempo.

A mis amigos del posgrado: Veronica, Marcos, Luis Yair, Esther, Carmelo, Dorilian, Saul, Leonardo, gracias por acompañarme en los buenos momentos.

Al **Consejo Nacional de Ciencia y Tecnología (CONACYT)**, por el apoyo económico proporcionado durante estos dos años de estudio en el posgrado.

Índice general

Introducción

Los datos multivariados de dimensión mayor o igual al tamaño de la muestra (datos de dimensión alta) aparecen en diversos campos, algunos de ellos son genética, análisis funcional, finanzas, análisis de imágenes médicas, climatología, reconocimiento de texto, entre otros (ver [9]). Cabe mencionar que en el contexto de datos de dimensión alta la estimación de la matriz de covarianza poblacional no es un problema fácil, ya que se tienen que estimar muchos parámetros con pocos datos, por lo que la estimación de esta matriz y pruebas de hipótesis acerca de ella requieren técnicas estadísticas diferentes a las del caso clásico donde el tamaño de la muestra es mucho mayor que la dimensión de los datos.

Consideremos $X_1, X_2, \ldots, X_N$ un conjunto de vectores aleatorios independientes de la distribución normal multivariada $N_d(\mu, \Sigma)$, donde la media μ y la matriz de covarianza Σ son desconocidas, y estamos interesados en probar

$$H_0 : \Sigma = I_p \quad \text{vs} \quad H_1 : \Sigma \neq I_p, \tag{1}$$

o

$$H_0 : \Sigma = \lambda I_p \quad \text{vs} \quad H_1 : \Sigma \neq \lambda I_p, \tag{2}$$

donde λ es desconocida. La hipótesis nula H_0 de (2) es llamada **hipótesis de esfericidad**. En [1] y [13] es demostrado que la prueba de razón de verosimilitud para contrastar las hipótesis en (2) se basa en el *estadístico de elipticidad* dado por

$$V = \frac{\det(S)}{[\text{tr}(S)/p]^p}, \tag{3}$$

donde S es la matriz de covarianza muestral de los datos. En el caso en que $p \geq N$ (caso de dimensión alta), con probabilidad uno, S no es de rango completo y consecuentemente $\det(S) = 0$. Esto indica que la prueba de razón de verosimilitud para (2) solo existe cuando $p < N$ (caso clásico). Debido a lo anterior, ha habido mucho interés en proponer y analizar pruebas de esfericidad en el contexto de datos normales de dimensión alta.

En esta tesis se presentan un conjunto de pruebas de hipótesis en el contexto de dimensión alta y también el caso clásico, estas pruebas son comparadas mediante simulaciones en término del error Tipo I y la función potencia de las pruebas. El análisis

presentado proporciona una comparación más amplia entre varias pruebas de hipótesis encontradas en la literatura para los contrastes de hipótesis (1) y (2), para el caso clásico y de dimensión alta, siendo este último el de mayor interés en esta tesis.

En el Capítulo 1 se presentan conceptos básicos y resultados importantes de pruebas de hipótesis. En el Capítulo 2 se presenta el conjunto de pruebas de hipótesis que serán consideradas, así como sus propiedades y resultados asintóticos. En el Capítulo 3 se presenta un estudio de simulación para evaluar el comportamiento de las pruebas de hipótesis, además se presentan unos ejemplos de aplicación para datos de dimensión alta. Finalmente se presenta un capítulo de conclusiones y un apéndice de detalles técnicos.

Capítulo 1

Pruebas de hipótesis

En este capítulo se dan algunas definiciones y teoremas importantes de pruebas de hipótesis, las cuales fueron tomadas de [6].

1.1. Conceptos básicos

Definición 1.1. *Una **hipótesis** es una afirmación acerca de un parámetro poblacional.*

El problema de prueba de hipótesis es un procedimiento basado en una muestra poblacional sobre la cual se realizan dos afirmaciones (o hipótesis) en las que se debe de decidir cuál es verdadera. En general estas dos afirmaciones son excluyentes.

Definición 1.2. *Las dos hipótesis en un problema de prueba de hipótesis son llamadas **hipótesis nula** e **hipótesis alternativa**, las cuales son denotadas como H_0 y H_1 respectivamente.*

Si θ denota un parámetro poblacional, la forma general de la hipótesis nula y alternativa es el siguiente

$$H_0 : \theta \in \Theta_0 \quad \text{y} \quad H_1 : \theta \in \Theta_0^c,$$

donde Θ_0 es algún subconjunto del espacio parametral Θ y Θ_0^c es el complemento. En un problema de prueba de hipótesis, después de observar la muestra el experimentador debe decidir no rechazar H_0 y aceptarla como verdadera, o rechazar H_0 y aceptar H_1 como verdadera.

Definición 1.3. *Un **procedimiento de prueba de hipótesis** (o **prueba de hipótesis**) es una regla que especifica*

i) Para que valores de la muestra se toma la decisión de aceptar H_0 como verdadera.

ii) Para que valores de la muestra H_0 es rechazada y H_1 es aceptada como verdadera.

El subconjunto del espacio muestral para el cual H_0 debe ser rechazada es llamado ***región de rechazo*** *o* ***región crítica****. El complemento de la región de rechazo es llamado* ***región de aceptación****.*

En ocasiones nos preocupamos en la distinción entre rechazar H_0 y aceptar H_1. En el primer caso, no hay nada implícito sobre qué declaramos como aceptado, solo que la afirmación dada por H_0 está siendo rechazada. De manera similar se puede hacer la distinción entre aceptar H_0 y no rechazar H_0. En la primera frase el experimentador está dispuesto a aceptar la afirmación especificada en H_0, mientras que la segunda frase implica que realmente no creemos en H_0, pero no tenemos evidencia para rechazarla. No nos preocuparemos por estas cuestiones, en lo que sigue veremos el problema de pruebas de hipótesis como un problema en el cual una de las dos hipótesis es tomada como verdadera.

1.2. Prueba de razón de verosimilitud

La prueba de razón de verosimilitud está relacionada con la función y el estimador de máxima verosimilitud, los cuales se definen a continuación.

Definición 1.4. *Si $X_1, X_2, ..., X_n$ es una muestra aleatoria de una población con función de densidad de probabilidad (fdp) o función masa de probabilidad (fmp) $f(x \mid \theta)$ (θ podría ser un vector), la* ***función de verosimilitud*** *está definida como*

$$L(\theta \mid x_1, x_2, ..., x_n) = L(\theta \mid x) = f(x \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta), \quad \forall \theta \in \Theta,$$

donde $x = (x_1, x_2, \ldots, x_n)$ es una observación de la muestra aleatoria.

Definición 1.5. *Para cada punto muestral $x \in \mathbb{R}^n$, sea $\widehat{\theta}(x)$ el valor del parámetro $\theta = (\theta_1, ..., \theta_k)$ para el cual $L(\theta \mid x)$ alcanza el máximo como una función de θ, con x fijo. El* ***estimador de máxima verosimilitud*** *(EMV) del parametro θ basado en la muestra X es $\widehat{\theta}(X)$.*

La idea es que el EMV de θ debe ser tal que el valor numérico de la muestra aleatoria X tenga probabilidad máxima.

Definición 1.6. *El* ***estadístico razón de verosimilitud*** *para $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$ es*

$$\lambda(x) = \frac{sup_{\theta \in \Theta_0} L(\theta \mid x)}{sup_{\theta \in \Theta} L(\theta \mid x)}, \tag{1.1}$$

donde $L(\theta \mid x)$ es la función de verosimilitud de θ basada en la muestra observada x. Una ***prueba de razón de verosimilitud*** *(LRT), es cualquier prueba que tiene región de rechazo*

$$\{x : \lambda(x) \leqslant c\}, \quad \text{c} \in (0, 1).$$

En el caso en que x es la observación de una muestra aleatoria con distribución discreta, el numerador de $\lambda(x)$ es la probabilidad máxima de la muestra observada, cuando el máximo se calcula sobre los parámetros en la hipótesis nula. El denominador de $\lambda(x)$ es la probabilidad máxima de la muestra observada sobre todos los posibles parámetros. La razón entre estos dos máximos es pequeña si hay puntos parametrales en la hipótesis alternativa para los cuales la muestra observada es mucho más verosímil que para cualquier punto parametral en la hipótesis nula. En este caso el criterio de la prueba de razón de verosimilitud nos dice que H_0 debe ser rechazada y aceptar H_1 como verdadera. Tenemos que $0 \leqslant \lambda(x) \leqslant 1$, entre más cercano a 1 sea $\lambda(x)$, más verosímil es que $\theta \in \Theta_0$, por otro lado mientras más alejado de 1 sea $\lambda(x)$, más creíble será que $\theta \in \Theta_0^c$.

Ejemplo 1.1. *Sean $X_1, X_2, ..., X_n$ variables aleatorias i.i.d. $N(\theta, 1)$. Consideramos el juego de hipótesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, donde θ_0 es un valor fijo. Sea $L(\theta \mid x)$ la función de verosimilitud, entonces*

$$\begin{aligned} L(\theta \mid x) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_i-\theta)^2}{2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[\sum_{i=1}^{n} -\frac{(x_i-\theta)^2}{2}\right]. \end{aligned}$$

Para encontrar el máximo de la función de $L(\theta \mid x)$ aplicamos logaritmo. Note que $L(\theta \mid x)$ se maximiza en el mismo punto donde $\log(L(\theta \mid x))$ lo hace, debido a las propiedades del logaritmo y su monotonía.

$$\log(L(\theta \mid x)) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2.$$

La ecuación $d\log(L(\theta|x))/d\theta = 0$ se reduce a

$$\begin{aligned} \sum_{i=1}^{n}(x_i-\theta) &= 0 \\ \theta &= \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}. \end{aligned}$$

Calculando la segunda derivada se tiene que

$$d^2\log(L(\theta|x))/d\theta^2 = -n.$$

Por el criterio de la segunda derivada tenemos que $d^2\log(L(\theta|x))/d\theta^2 < 0$ y por lo tanto $\theta = \overline{x}$ es un máximo, así el estimador de máxima verosimilitud es $\widehat{\theta} = \overline{X}$.

Ya que solo hay un valor de θ especificado por H_0, el numerador de $\lambda(x)$ es $L(\theta_0 \mid x)$ y el denominador es $L(\overline{x} \mid x)$. Así el estadístico de razón de verosimilitud es

$$\lambda(x) = \frac{(2\pi)^{-n/2}\exp\left[-\sum_{i=1}^{n}\frac{(x_i-\theta_0)^2}{2}\right]}{(2\pi)^{-n/2}\exp\left[-\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{2}\right]}$$
$$= \exp\left[\left(-\sum_{i=1}^{n}(x_i-\theta_0)^2 + \sum_{i=1}^{n}(x_i-\overline{x})^2\right)/2\right].$$

La expresión dada en la exponencial de $\lambda(x)$ puede ser simplificada usando la igualdad

$$\sum_{i=1}^{n}(x_i-\theta_0)^2 = \sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\theta_0)^2.$$

Así el estadístico de la prueba de razón de verosimilitud es

$$\lambda(x) = \exp\left[\frac{-n(\overline{x}-\theta_0)^2}{2}\right].$$

Una prueba de razón de verosimilitud es una prueba que rechaza H_0 para un valor de $\lambda(x)$ pequeño. De la expresión anterior la región de rechazo es $\{x : \lambda(x) \leqslant c\}$ y puede reescribirse como

$$\{x : |\overline{x}-\theta_0| \geqslant \sqrt{-2(\log c)/n}\}.$$

Ya que el rango de c está entre 0 y 1, entonces el rango de $\sqrt{-2(\log c)/n}$ está entre 0 e ∞. Por lo tanto las pruebas de razón de verosimilitud son solo aquellas pruebas que rechazan H_0 si la media muestral difiere del valor hipotético θ_0 en una cantidad mayor que la especificada.

1.3. Tipos de error y función potencia

Definición 1.7. *Consideramos el juego de hipótesis $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$. Si $\theta \in \Theta_0$ pero la prueba de hipótesis decide incorrectamente rechazar H_0, se dice que se ha cometido un **error de tipo I**. Si, por otro lado $\theta \in \Theta_0^c$ pero se decide incorrectamente aceptar la hipótesis H_0, se dice que se ha cometido un **error de tipo II**.*

Denotamos por $\mathbb{P}_\theta$ a la distribución del vector aleatorio X cuyas entradas tienen función de densidad o función masa $f(x \mid \theta)$. Entonces en un juego de hipótesis $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$, dada una prueba de hipótesis con región de rechazo R, si $\theta \in \Theta_0$, la probabilidad del error de tipo I es $\mathbb{P}_\theta(X \in R)$. Si $\theta \in \Theta_0^c$, la probabilidad del error de tipo II es $\mathbb{P}_\theta(X \in R^c) = 1 - \mathbb{P}_\theta(X \in R)$.

Definición 1.8. *Dada una prueba de hipótesis con región de rechazo R, la función $\beta : \Theta \to [0,1]$ dada por $\beta(\theta) = \mathbb{P}_\theta(X \in R)$ se llama* ***función potencia*** *de la prueba.*

Una función potencia ideal sería una que toma el valor de 0 para todo $\theta \in \Theta_0$ y el valor 1 para todo $\theta \in \Theta_0^c$, sin embargo, excepto en situaciones triviales, esta función potencia ideal no se alcanza. De esta manera, una buena prueba será aquella que tome valores cercanos a 0 para $\theta \in \Theta_0$ y valores cercanos a 1 para $\theta \in \Theta_0^c$.

Definición 1.9. *Consideremos el juego de hipótesis de $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$.*

a) Para $0 \leqslant \alpha \leqslant 1$, una prueba con función potencia $\beta(\theta)$ es una ***prueba de tamaño*** *α si $sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$.*

b) Para $0 \leqslant \alpha \leqslant 1$, una prueba con función potencia $\beta(\theta)$ es una ***prueba de nivel*** *α si $sup_{\theta \in \Theta_0} \beta(\theta) \leqslant \alpha$.*

1.4. Pruebas uniformemente más potentes

Definición 1.10. *Sea C una clase de pruebas para contrastar $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$. Se dice que una prueba de la clase C, con función potencia $\beta(\theta)$, es una prueba* ***uniformemente más potente*** *en la clase C si $\beta(\theta) \geqslant \beta'(\theta)$ para todo $\theta \in \Theta_0^c$ y para toda $\beta'(\theta)$ que es una función potencia de una prueba en la clase C.*

Si construimos una prueba de hipótesis con un nivel α dado, entonces se está controlando solo la probabilidad del error de tipo I, y no la del error de tipo II. Generalmente, las hipótesis nula y alternativa se eligen de tal modo que sea más importante controlar la probabilidad de error de tipo I.

En lo que sigue se considera a la clase C como pruebas de nivel α, es decir, el propósito será encontrar la prueba uniformemente más potente de nivel α.

El siguiente teorema describe con claridad cuáles son las pruebas uniformemente más potentes de nivel α cuando H_0 y H_1 son hipótesis simples, es decir, cuando H_0 y H_1 especifican cada una sólo una distribución para la muestra X.

Teorema 1.1 (Lema de Neyman-Pearson). *Consideremos el juego de hipótesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, donde $f(x \mid \theta_i)$ es la función de densidad o función masa correspondiente a θ_i, $i = 1, 2$. Considerar una región de rechazo R tal que*

$$\begin{aligned} x \in R \quad si \quad f(x \mid \theta_1) > k f(x \mid \theta_0) \quad y \\ x \in R^c \quad si \quad f(x \mid \theta_1) < k f(x \mid \theta_0), \end{aligned} \tag{1.2}$$

para algún $k \geqslant 0$, y

$$\alpha = \mathbb{P}_{\theta_0}(X \in R). \tag{1.3}$$

Entonces

a) (Suficiencia) Toda prueba con región de rechazo R que cumple (1.2) y (1.3) es una prueba de nivel α uniformemente más potente.

b) (Necesidad) Si existe una prueba que satisface (1.2) y (1.3) con $k > 0$, entonces toda prueba de nivel α uniformemente más potente es una prueba de tamaño α que satisface (1.3), y toda prueba de nivel α uniformemente más potente satisface (1.2), excepto quizá en un conjunto A que satisface $\mathbb{P}_{\theta_0}(X \in A) = \mathbb{P}_{\theta_1}(X \in A) = 0$.

Demostración. Para probar el teorema consideramos el caso en que $f(x \mid \theta_0)$ y $f(x \mid \theta_1)$ son funciones de densidad de una variable aleatoria; la demostración para el caso en el que las variables aleatorias son discretas es análogo, solo hay que reemplazar las integrales por sumas. Cualquier prueba que satisface (1.3) es una de tamaño α y por lo tanto de nivel α debido a que

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(X \in R) = \mathbb{P}_{\theta_0}(X \in R) = \alpha,$$

ya que Θ_0 tiene solo al punto θ_0.

Para cualquier prueba con región de rechazo R^*, definamos la *función prueba* ϕ^* de la siguiente manera

$$\phi^*(x) = \begin{cases} 1, & si\ x \in R^* \\ 0, & si\ x \in R^{*^c}. \end{cases}$$

Sean $\phi(x)$ la función prueba de una prueba de hipótesis que satisface (1.2) y (1.3) y $\phi'(x)$ otra función prueba de cualquier prueba de nivel α. Sean $\beta(\theta)$ y $\beta'(\theta)$ las funciones potencia correspondientes a $\phi(x)$ y $\phi'(x)$, respectivamente. Veamos que

$$[\phi(x) - \phi'(x)]\,[f(x \mid \theta_1) - k f(x \mid \theta_0)] \geqslant 0, \quad \forall x.$$

Si $x \in R$, tenemos que $\phi(x) = 1$ y $0 \leqslant \phi'(x) \leqslant 1$, entonces $\phi(x) - \phi'(x) \geqslant 0$. Por lo tanto de (1.2) tenemos que

$$[\phi(x) - \phi'(x)]\,[f(x \mid \theta_1) - k f(x \mid \theta_0)] \geqslant 0.$$

Si $x \in R^c$, tenemos que $\phi(x) = 0$ y entonces $\phi(x) - \phi'(x) \leqslant 0$. Por lo tanto de (1.2) tenemos que

$$[\phi(x) - \phi'(x)]\,[f(x \mid \theta_1) - k f(x \mid \theta_0)] \geqslant 0.$$

Si R' es la región de rechazo de ϕ',

$$
\begin{aligned}
0 \leqslant & \int [\phi(x) - \phi'(x)] \, [f(x \mid \theta_1) - k f(x \mid \theta_0)] \, dx \\
= & \int \phi(x) f(x \mid \theta_1) - \phi'(x) f(x \mid \theta_1) - k\phi(x) f(x \mid \theta_0) + k\phi'(x) f(x \mid \theta_0) dx \\
= & \int \phi(x) f(x \mid \theta_1) dx - \int \phi'(x) f(x \mid \theta_1) dx - k \int \phi(x) f(x \mid \theta_0) dx \\
& + k \int \phi'(x) f(x \mid \theta_0) dx.
\end{aligned} \tag{1.4}
$$

Por lo que

$$
\begin{aligned}
0 \leq & \int_R \phi(x) f(x \mid \theta_1) dx + \int_{R^c} \phi(x) f(x \mid \theta_1) dx - \int_{R'} \phi'(x) f(x \mid \theta_1) dx \\
& - \int_{R'^c} \phi'(x) f(x \mid \theta_1) dx - k \int_R \phi(x) f(x \mid \theta_0) dx - k \int_{R^c} \phi(x) f(x \mid \theta_0) dx \\
& + k \int_{R'} \phi'(x) f(x \mid \theta_0) dx + k \int_{R'^c} \phi'(x) f(x \mid \theta_0) dx \\
= & \int_R f(x \mid \theta_1) dx - \int_{R'} f(x \mid \theta_1) dx - k \int_R f(x \mid \theta_0) dx + k \int_{R'} f(x \mid \theta_0) dx \\
= & \beta(\theta_1) - \beta'(\theta_1) - k(\beta(\theta_0) - \beta'(\theta_0)).
\end{aligned} \tag{1.5}
$$

La afirmación a) se prueba al notar que, si ϕ' es una prueba de nivel α y ϕ es una prueba de tamaño α, entonces

$$\beta(\theta_0) - \beta'(\theta_0) = \alpha - \beta'(\theta_0) \geqslant 0.$$

Por lo tanto de (1.5) y para $k \geqslant 0$ tenemos que

$$
\begin{aligned}
0 & \leqslant \beta(\theta_1) - \beta'(\theta_1) - k(\beta(\theta_0) - \beta'(\theta_0) \\
& \leqslant \beta(\theta_1) - \beta'(\theta_1).
\end{aligned}
$$

Así $\beta(\theta_1) \geqslant \beta'(\theta_1)$ y por lo tanto ϕ tiene mayor potencia que ϕ'. Como ϕ' es una prueba de nivel α arbitraria y θ_1 es el único punto de Θ_0^c, se concluye que ϕ es una prueba de nivel α uniformemente más potente.

Para probar la afirmación b), sea ϕ' la función prueba de una prueba de nivel α uniformemente más potente. Por a) si ϕ es la función prueba de una prueba que satisface (1.2) y (1.3), también es una prueba de nivel α uniformemente más potente, de donde $\beta(\theta_1) = \beta'(\theta_1)$. De (1.5) y para $k > 0$, tenemos que

$$\alpha - \beta'(\theta_0) = \beta(\theta_0) - \beta'(\theta_0) \leqslant 0.$$

De lo anterior y el hecho de que $\beta'(\theta_0) \leqslant \alpha$, se tiene que $\beta'(\theta_0) = \alpha$ y por lo tanto ϕ' es una prueba de tamaño α y también implica que (1.4) es una igualdad para este caso, es decir,

$$\int (\phi(x) - \phi'(x))(f(x \mid \theta_1) - kf(x \mid \theta_0))dx = 0.$$

Como el integrando es no negativo, se tiene que

$$(\phi(x) - \phi'(x))(f(x \mid \theta_1) - kf(x \mid \theta_0)) = 0,$$

excepto en un conjunto A con medida de Lebesgue cero, es decir, excepto en un conjunto A tal que

$$\int_A dx = 0.$$

Como $f(x \mid \theta_1) - kf(x \mid \theta_0) \neq 0 \; \forall x \in \mathbb{R}^n$, resulta que

$$\phi(x) = \phi'(x) \quad \forall x \in \mathbb{R}^n \setminus A.$$

De aquí se sigue que las regiones de rechazo de ϕ y ϕ' difieren en un conjunto de medida cero. Así ϕ' satisface (1.2) excepto en A. Por la continuidad de $f(x \mid \theta_i)$, $i = 1, 2$, con respecto a la medida de Lebesgue, se tiene

$$\int_A f(x \mid \theta_i)dx = 0, \quad i = 1, 2,$$

es decir,

$$\mathbb{P}_{\theta_0}(X \in A) = \mathbb{P}_{\theta_1}(X \in A) = 0.$$

□

Definición 1.11. *Sea $X = X_1, X_2, ..., X_n$ una muestra aleatoria de la función de densidad o función masa $f(x \mid \theta)$. Un estadístico $T = T(X)$ es un* ***estadístico suficiente*** *si la distribución condicional de X dado $T = t$ no depende de θ.*

Teorema 1.2 (Teorema de factorización). *Sea $f(x \mid \theta)$ función de densidad o masa de una muestra aleatoria X. $T(X)$ es un estadístico suficiente para θ si y solo si existen funciones $g(t \mid \theta)$ y $h(x)$ tal que*

$$f(x \mid \theta) = g(T(x) \mid \theta)h(x).$$

Corolario 1.1. *Consideremos el juego de hipótesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. Supongamos que $T(X)$ es un estadístico suficiente del parámetro θ, con función de densidad*

o masa $g(t \mid \theta_i)$, $i = 0, 1$. *Entonces cualquier prueba basada en* T *con región de rechazo* S *es una prueba uniformemente más potente de nivel* α *si satisface*

$$\begin{array}{ll} t \in S & si \quad g(t \mid \theta_1) > kg(t \mid \theta_0) \quad y \\ t \in S^c & si \quad g(t \mid \theta_1) < kg(t \mid \theta_0), \end{array} \tag{1.6}$$

para algún $k \geqslant 0$, *donde*

$$\alpha = \mathbb{P}_{\theta_0}(T \in S). \tag{1.7}$$

Demostración. En términos de la muestra original X, la prueba basada en T tiene una región de rechazo $R = \{x : T(x) \in S\}$. Por el teorema de factorización, la función de probabilidad de X puede escribirse como

$$f(x \mid \theta_i) = g(T(x) \mid \theta_i)h(x), \quad i = 0, 1,$$

para alguna función $h(x)$ no negativa. Multiplicando por esta función no negativa a la desigualdad (1.5), podemos observar que R satisface

$$x \in R \quad si \quad f(x \mid \theta_1) = g(T(x) \mid \theta_1)h(x) > kg(T(x) \mid \theta_0)h(x) = kf(x \mid \theta_0)$$

y

$$x \in R^c \quad si \quad f(x \mid \theta_1) = g(T(x) \mid \theta_1)h(x) < kg(T(x) \mid \theta_0)h(x) = kf(x \mid \theta_0).$$

Además, por (1.6) se tiene que

$$\mathbb{P}_{\theta_0}(X \in R) = \mathbb{P}_{\theta_0}(T(X) \in S) = \alpha.$$

Así, por la parte de suficiencia del teorema de Neyman-Pearson, la prueba basada en T es una prueba uniformemente más potente de nivel α. □

Para ilustrar el Lema de Neyman-Pearson consideraremos el siguiente ejemplo.

Ejemplo 1.2. *Sea* $X_1, X_2, ..., X_n$ *una muestra aleatoria de la distribución* $N(\theta, \sigma^2)$, *con* θ *desconocida y* σ^2 *conocida. Supongamos que queremos contrastar las hipótesis*

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta = \theta_1, \quad \theta_0 > \theta_1.$$

Encontraremos la región de rechazo usando el Lema de Neyman-Pearson, el cual hace referencia a la función de verosimilitud de la muestra, en este caso la función de verosimilitud de la distribución normal es

$$L(\theta|x) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right),$$

así

$$\begin{aligned}\frac{L(\theta_1|x)}{L(\theta_0|x)} &= \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\theta_0)^2-(x_i-\theta_1)^2]\right)\\ &= \exp\left(-\frac{1}{2\sigma^2}[-2n\overline{x}(\theta_0-\theta_1)+n(\theta_0^2-\theta_1^2)]\right).\end{aligned}$$

Entonces tenemos que

$$\begin{aligned}\frac{L(\theta_1|x)}{L(\theta_0|x)} > k \quad sii \quad & -\frac{1}{2\sigma^2}[-2n\overline{x}(\theta_0-\theta_1)+n(\theta_0^2-\theta_1^2)] > \log k\\ sii \quad & -2\overline{x}(\theta_0-\theta_1)+(\theta_0^2-\theta_1^2) < \frac{2\sigma^2}{n}\log k\\ sii \quad & \overline{x} < \frac{\frac{2\sigma^2}{n}\log k-(\theta_0^2-\theta_1^2)}{-2(\theta_0-\theta_1)}.\end{aligned}$$

Por lo tanto, la región de rechazo de una prueba uniformemente más potente es

$$\mathcal{C} = \left\{x : \overline{x} < \frac{\frac{2\sigma^2}{n}\log k-(\theta_0^2-\theta_1^2)}{-2(\theta_0-\theta_1)}\right\}.$$

Ahora encontraremos un k de tal manera que $\mathcal{C}$ sea de tamaño α. Tenemos que

$$\begin{aligned}\alpha &= \mathbb{P}_{\theta_0}\left(\overline{X} < \frac{\frac{2\sigma^2}{n}\log k-(\theta_0^2-\theta_1^2)}{-2(\theta_0-\theta_1)}\right)\\ &= \mathbb{P}_{\theta_0}\left(\frac{\overline{X}-\theta_0}{\sigma/\sqrt{n}} < \frac{\frac{2\sigma^2}{n}\log k-(\theta_0^2-\theta_1^2)-\theta_0}{-2(\theta_0-\theta_1)\sigma/\sqrt{n}}\right),\end{aligned}$$

tomando $\dfrac{\frac{2\sigma^2}{n}\log k-(\theta_0^2-\theta_1^2)-\theta_0}{-2(\theta_0-\theta_1)\sigma/\sqrt{n}} = z_\alpha = -z_{1-\alpha}$*, se tiene que esta probabilidad es* α*. Por lo tanto una prueba uniformemente más potente de* H_0 *vs* H_1*, tiene región de rechazo*

$$\left\{x : \frac{\overline{x}-\theta_0}{\sigma/\sqrt{n}} < -z_{1-\alpha}\right\} = \{x : \overline{x}\theta_0 - z_{1-\alpha}\sigma/\sqrt{n}\}.$$

Las hipótesis H_0 y H_1 del lema de Neyman-Pearson que especifican una posible distribución para la muestra X son llamadas hipótesis simples. Sin embargo, en muchos problemas las hipótesis de interés especifican más de una posible distribución para la muestra. Estas hipótesis son llamadas hipótesis compuestas. En particular, las hipótesis que afirman que un parámetro es grande, por ejemplo $H : \theta \geqslant \theta_0$ o más pequeño $H : \theta \leqslant \theta_0$ son llamadas hipótesis de un lado (o unilateral). Las hipótesis que afirman que un parámetro es más grande o más pequeño que un valor, por ejemplo $H : \theta \neq \theta_0$, son llamadas hipótesis de dos lados (o bilaterales).

Para estudiar pruebas uniformemente más potente para hipótesis de un lado requerimos la siguiente definición.

Definición 1.12. *Una familia de funciones de densidad o masa* $\{g(t \mid \theta) : \theta \in \Theta \subseteq \mathbb{R}\}$ *de una variable aleatoria* T*, tiene* ***Razón de verosimilitud monótona*** *si, siempre que* $\theta_1 < \theta_2$*,*

$$\frac{g(t \mid \theta_2)}{g(t \mid \theta_1)},$$

es monótona como función de t *en* $\{t : g(t \mid \theta_1) > 0 \text{ o } g(t \mid \theta_2) > 0\}$*, donde definimos* $\frac{c}{0} = \infty$*, si* $c > 0$*.*

Muchas familias de distribuciones comunes tienen razón de verosimilitud monótona, por ejemplo la normal, la poisson y la binomial. De hecho la familia exponencial cuya fmp o fdp, tienen la forma $g(t|\theta) = h(t)c(\theta)e^{w(\theta)t}$ tiene razón de verosimilitud monótona si $w(\theta)$ es una función no-decreciente.

Ejemplo 1.3. *Consideremos* $X_1, X_2, ..., X_n$ *una muestra aleatoria con distribución* $\exp(\theta)$*. Si* $\{f(x|\theta) : \theta \in \Theta\} = \{\theta e^{-\theta x} I_{(0,\infty)}(x) : \theta > 0\}$ *es la familia de la distribución exponencial*

$$g(x|\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x),$$

tomando $h(x) = I_{(0,\infty)}(x)$*,* $c(\theta) = \theta$ *y* $w(\theta) = -\theta$*, la familia de densidades queda de la forma*

$$g(x|\theta) = h(x)c(\theta)e^{w(\theta)x},$$

es decir, pertenece a una familia exponencial y además notemos que $w(\theta)$ *es una función decreciente. Luego consideramos la razón de verosimilitud*

$$\begin{aligned}\frac{L(\theta')}{L(\theta'')} &= \frac{\theta' \exp(-\theta' \sum_{i=1}^n x_i)}{\theta'' \exp(-\theta'' \sum_{i=1}^n x_i)} \\ &= \left(\frac{\theta'}{\theta''}\right) \exp\left(-(\theta' - \theta'') \sum_{i=1}^n x_i\right).\end{aligned}$$

Si $\theta' > \theta''$, entonces esta razón es decreciente como función de $t = \sum_{i=1}^{n} x_i$. Por lo tanto esta familia tiene razón de verosimilitud monótona.

Proposicion 1.1. *Sea T una variable aleatoria cuya familia de densidad o masa de probabilidad es $\{g(t \mid \theta) : \theta \in \Theta\}$, tal que para $\theta_1 < \theta_2$,*

$$\frac{g(t \mid \theta_2)}{g(t \mid \theta_1)}$$

es creciente como función de t. Entonces

$$\mathbb{P}_{\theta_2}(T \leqslant t) \leqslant \mathbb{P}_{\theta_1}(T \leqslant t).$$

Demostración. Sean $t_1 < t < t_2$, así

$$\frac{g(t_1 \mid \theta_2)}{g(t_1 \mid \theta_1)} \leqslant \frac{g(t_2 \mid \theta_2)}{g(t_2 \mid \theta_1)},$$

entonces,

$$g(t_1 \mid \theta_2) g(t_2 \mid \theta_1) \leqslant g(t_2 \mid \theta_2) g(t_1 \mid \theta_1),$$

integrando ambos lados de la desigualdad de $-\infty$ a t y después de t a ∞, tenemos

$$\int_t^\infty \int_{-\infty}^t g(t_1 \mid \theta_2) dt_1 g(t_2 \mid \theta_1) dt_2 \leqslant \int_t^\infty \int_{-\infty}^t g(t_2 \mid \theta_2) dt_2 g(t_1 \mid \theta_1) dt_1,$$

de aquí se tienen las siguientes desigualdades

$$\begin{aligned}
&\mathbb{P}_{\theta_2}(T \leqslant t) \int_t^\infty g(t_2 \mid \theta_1) dt_2 \leqslant \mathbb{P}_{\theta_1}(T \leqslant t) \int_t^\infty g(t_2 \mid \theta_2) dt_2, \\
&\mathbb{P}_{\theta_2}(T \leqslant t) \mathbb{P}_{\theta_1}(T > t) \leqslant \mathbb{P}_{\theta_1}(T \leqslant t) \mathbb{P}_{\theta_2}(T > t), \\
&\mathbb{P}_{\theta_2}(T \leqslant t) \left[1 - \mathbb{P}_{\theta_1}(T \leqslant t)\right] \leqslant \mathbb{P}_{\theta_1}(T \leqslant t) \left[1 - \mathbb{P}_{\theta_2}(T \leqslant t)\right], \\
&\mathbb{P}_{\theta_2}(T \leqslant t) \leqslant \mathbb{P}_{\theta_1}(T \leqslant t).
\end{aligned}$$

□

Teorema 1.3 (Teorema de Karlin-Rubin). *Considerar el juego de hipótesis $H_0 : \theta \leqslant \theta_0$ vs $H_1 : \theta > \theta_0$. Supongamos que T es un estadístico suficiente de θ, y que la familia de funciones de densidad o masa de T, $\{g(t \mid \theta) : \theta \in \Theta\}$, tiene razón de verosimilitud monótona. Entonces para cualquier t_0, la prueba que rechaza H_0 si y sólo si $T > t_0$ es una prueba uniformemente más potente de nivel α, donde $\alpha = \mathbb{P}_{\theta_0}(T > t_0)$.*

Demostración. Sea $\beta = \mathbb{P}_\theta(T > t_0)$, la función potencia de la prueba. Fijar $\theta' > \theta_0$ y considerar las pruebas $H_0' : \theta = \theta_0$ vs $H_1' : \theta = \theta'$. Supongamos que la familia de funciones de densidad o masa de probabilidad de T tiene razón de verosimilitud monótona creciente, de la proposición 1.1, tenemos que si

$$\theta_0 < \theta',$$

entonces

$$\mathbb{P}_{\theta'}(T \leqslant t) \leqslant \mathbb{P}_{\theta_0}(T \leqslant t),$$

de aquí se tienen las siguientes desigualdades

$$\begin{aligned} &1 - \mathbb{P}_{\theta_0}(T \leqslant t) \leqslant 1 - \mathbb{P}_{\theta'}(T \leqslant t), \\ &\mathbb{P}_{\theta_0}(T > t) \leqslant \mathbb{P}_{\theta'}(T > t), \\ &\beta(\theta_0) \leqslant \beta(\theta'), \quad si \quad \theta_0 < \theta'. \end{aligned}$$

Por lo tanto $\beta(\theta)$ es creciente. Entonces

i) $\sup_{\theta \leqslant \theta_0} \beta(\theta) = \beta(\theta_0) = \alpha$, por lo que es una prueba de nivel α.

ii) Definimos

$$k' = \inf_{t \in \mathcal{T}} \frac{g(t \mid \theta')}{g(t \mid \theta_0)},$$

donde $\mathcal{T} = \{t : t > t_0 \text{ y } g(t \mid \theta') > 0 \text{ o } g(t \mid \theta_0) > 0\}$, de donde se sigue que $t > t_0$ si y solo si

$$\frac{g(t \mid \theta')}{g(t \mid \theta_0)} > k'.$$

El Corolario 1.1, *i*) y *ii*) implican que $\beta(\theta') \geqslant \beta^*(\theta')$, donde $\beta^*(\theta)$ es la función potencia de cualquier otra prueba de nivel α de H_0', es decir, cualquier prueba que satisface $\beta^*(\theta_0) \leqslant \alpha$. Sin embargo, toda prueba de nivel α de H_0 satisface

$$\beta^*(\theta_0) \leqslant \sup_{\theta \in \Theta_0} \beta^*(\theta) \leqslant \alpha.$$

Así, $\beta(\theta') \geqslant \beta^*(\theta')$ para toda prueba de nivel α de H_0. Ya que θ' fue arbitrario, la prueba es una prueba uniformemente más potente de nivel α.

□

Para ilustrar el teorema consideraremos el siguiente ejemplo.

Ejemplo 1.4. *Sea $X_1, X_2, ..., X_n$ una muestra aleatoria con distribución $unif(0, \theta)$, donde $\theta > 0$. Consideremos el juego de hipótesis $H_0 : \theta \leqslant \theta_0$ vs $H_1 : \theta > \theta_0$. Es conocido que $T = X_{(n)} = \max\{x_1, x_2, ..., x_n\}$ es un estadístico suficiente para θ cuya función de densidad es*

$$f(t|\theta) = \frac{nt^{n-1}}{\theta^n} I_{\{0<t<\theta\}}.$$

Si $\theta_1 < \theta_2$ se tiene que

$$\frac{f(t|\theta_2)}{f(t|\theta_1)} = \left(\frac{\theta_2}{\theta_1}\right)^n \frac{I_{\{0<t<\theta_2\}}}{I_{\{0<t<\theta_1\}}}.$$

Como $\theta_1 < \theta$ tenemos los siguientes casos

$$\frac{f(t|\theta_2)}{f(t|\theta_1)} = \begin{cases} \left(\dfrac{\theta_2}{\theta_1}\right)^n, & si\ t < \theta_1, \\ \infty, & si\ \theta_1 \leqslant t \leqslant \theta_2. \end{cases}$$

Por lo tanto la razón de verosimilitud es monótona creciente. Así T tiene razón de verosimilitud monótona. Aplicando el teorema de Karlin-Rubin tenemos que una prueba uniformemente más potente de nivel α tiene región de rechazo $\{x : T > t_0\}$ con

$$\begin{aligned} \alpha &= \mathbb{P}_{\theta_0}(T > t_0) \\ &= \int_{t_0}^{\theta_0} \frac{nt^{n-1}}{\theta_0^n} dt = 1 - \left(\frac{t_0}{\theta_0}\right)^n, \end{aligned}$$

despejando tenemos que $t_0 = \theta_0 \sqrt[n]{1-\alpha}$ y por lo tanto la región de rechazo es

$$\{x : T > \theta_0 \sqrt[n]{1-\alpha}\}.$$

De forma análoga se puede mostrar usando el lema de Neyman-Pearson que la prueba que rechaza $H_0 : \theta \geqslant \theta_0$ a favor de $H_1 : \theta < \theta_0$ si y solo si $T < t_0$, es una prueba uniformemente más potente de nivel $\alpha = \mathbb{P}_{\theta_0}(T < t_0)$.

Capítulo 2

Pruebas de hipótesis para la matriz de covarianza

En este capítulo se presentan pruebas de hipótesis para la matriz de covarianza poblacional, considerando muestras aleatorias normales para el caso clásico y el de dimensión alta. Para cada prueba se conserva en la medida de lo posible la notación de la fuente de donde se extrajo para evitar imprecisiones. Para cada prueba se señala la fuente de la cual se extrajo.

La transpuesta de una matriz B será denotada como B'. Sean $X_1, X_2, ..., X_N$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$, donde p nos indica la dimensión del espacio. La media muestral $\overline{X}$ y la varianza muestral S_n están dadas por

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad S_n = \frac{1}{n}A, \quad donde \quad A = \sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})', \quad n = N - 1. \tag{2.1}$$

2.1. Pruebas de hipótesis para $H_0 : \Sigma = I_p$

Estamos interesados en contrastar

$$H_0 : \Sigma = I_p \quad vs \quad H_1 : \Sigma \neq I_p. \tag{2.2}$$

Si queremos realizar pruebas de hipótesis de $H_0 : \Sigma = \Sigma_0$, donde Σ_0 es una matriz de covarianza positiva definida y especificada, esto es equivalente a probar (2.2), ya que siempre se puede hacer una transformación de los datos $Y_i = \Sigma_0^{-1/2} X_i$, $i = 1, 2, ..., N$, los cuales son vectores i.i.d. con ditribucion $N_p(\Sigma_0^{-1/2}\mu, \Sigma_0^{-1/2}\Sigma\Sigma_0^{-1/2})$ (ver [7], [12]). Observemos que bajo la hipótesis nula las Y_i son i.i.d. con distribución $N(\Sigma_0^{-1/2}\mu, I_p)$.

2.1.1. Prueba de razón de verosimilitud (LRT_1)

Teorema 2.1. *Sean $X_1, ..., X_N$ vectores aleatorios independientes con distribución $N_p(\mu, \Sigma)$ y $N > p$, entonces los estimadores de máxima verosimilitud de μ y Σ son $\widehat{\mu} = \overline{X}$ y $\widehat{\Sigma} = (1/N)A = (n/N)S_n$ respectivamente.*

Demostración. La función de verosimilitud de la distribución normal multivariada es

$$L(\mu, \Sigma) = (2\pi)^{-Np/2}(\det\Sigma)^{-N/2}\text{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right)\exp\left[-\frac{1}{2}N(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)\right],$$

donde $\text{etr}(\cdot) = \exp[\text{tr}(\cdot)]$. Luego

$$L(\mu, \Sigma) \leqslant (\det\Sigma)^{-N/2}\text{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right),$$

esta desigualdad se da ya que como Σ es positiva definida, entonces Σ^{-1} también lo es y entonces

$$\exp\left[-\frac{1}{2}N(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)\right] < 1.$$

Si $\mu = \overline{X}$, entonces $(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu) = 0$, por lo que la función de verosimilitud alcanza su máximo en $\widehat{\mu} = \overline{X}$. Ahora solo queda maximizar la función

$$L(\overline{X}, \Sigma) = (\det\Sigma)^{-N/2}\text{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right),$$

o, equivalente, la función

$$\begin{aligned} g(\Sigma) = \log L(\overline{X}, \Sigma) &= -\frac{N}{2}\log(\det\Sigma) - \frac{1}{2}\text{tr}(\Sigma^{-1}A) \\ &= -\frac{N}{2}\log\left(\det\Sigma\frac{\Sigma^{-1}A}{\Sigma^{-1}A}\right) - \frac{1}{2}\text{tr}(\Sigma^{-1}A) \\ &= -\frac{N}{2}\log[\det(\Sigma\Sigma^{-1}A)(\Sigma^{-1}A)^{-1}] - \frac{1}{2}\text{tr}(\Sigma^{-1}A) \\ &= -\frac{N}{2}[\log(\det(A) - \log(\Sigma^{-1}A)] - \frac{1}{2}\text{tr}(\Sigma^{-1}A) \\ &= \frac{N}{2}\log(\Sigma^{-1}A) - \frac{1}{2}\text{tr}(\Sigma^{-1}A) - \frac{N}{2}\log(\det(A)), \end{aligned}$$

como A es una matriz no negativa definida, entonces existe una matriz no negativa definida $A^{1/2}$, tal que $A = A^{1/2}A^{1/2}$, de la última igualdad tenemos

$$\frac{N}{2}\log(A^{1/2}\Sigma^{-1}A^{1/2}) - \frac{1}{2}\mathrm{tr}(A^{1/2}\Sigma^{-1}A^{1/2}) - \frac{N}{2}\log(\det(A))$$
$$= \frac{1}{2}\sum_{i=1}^{p}(N\log\lambda_i - \lambda_i) - \frac{N}{2}\log(\det(A)),$$

donde $\lambda_1, ..., \lambda_p$ son los eigenvalores de $A^{1/2}\Sigma^{-1}A^{1/2}$, es decir de $\Sigma^{-1}A$. Consideremos la función

$$f(\lambda) = N\log\lambda - \lambda,$$

el cual tiene un máximo en $\lambda = N$, se sigue que

$$g(\Sigma) \leqslant \frac{Np}{2}\log N - \frac{Np}{2} - \frac{N}{2}\log(\det(A)).$$

Por lo que

$$L(\overline{X}, \Sigma) \leqslant N^{Np/2} - \exp(-Np/2)(\det(A))^{-N/2},$$

la igualdad se cumple si y solo si $\lambda_i = N$, $i = 1, ..., p$. Esta última condición es equivalente a $A^{1/2}\Sigma^{-1}A^{1/2} = NI_p$ y por lo tanto $\Sigma = (1/N)A$. Por lo tanto concluimos que

$$L(\mu, \Sigma) \leqslant N^{Np/2} - \exp(-Np/2)(\det(A))^{-N/2},$$

y la igualdad se cumple si y solo si $\mu = \overline{X}$ y $\Sigma = (1/N)A$. □

El siguiente resultado nos da la prueba de razón de verosimilitud (LRT_1) para $H_0 : \Sigma = I_p$ (ver [13]).

Teorema 2.2. *Sean $X_1, ..., X_N$ vectores aleatorios independiente con distribución $N_p(\mu, \Sigma)$ y sea*

$$A = nS_n = \sum_{i=1}^{N}(X_i - \overline{X})(X_i - \overline{X})', \quad (n = N - 1).$$

La prueba de razón de verosimilitud *de tamaño α para $H_0 : \Sigma = I_p$, rechaza H_0 si $\Lambda \leqslant c_\alpha$, donde*

$$\Lambda = \left(\frac{\mathrm{e}}{N}\right)^{pN/2}\mathrm{etr}(-A/2)(\det A)^{N/2}.$$

Demostración. De la Definición 1.6, el estadístico de razón de verosimilitud es

$$\Lambda = \frac{\text{Sup}_{\mu\in\mathbb{R}^p} L(\mu, I_p)}{\text{Sup}_{\mu\in\mathbb{R}^p,\Sigma>0} L(\mu,\Sigma)},$$

donde

$$L(\mu,\Sigma) = (\text{det}\Sigma)^{-N/2}\text{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right)\exp\left[-\frac{1}{2}N(\overline{X}-\mu)'\Sigma^{-1}(\overline{X}-\mu)\right].$$

El numerador puede ser calculado sustituyendo $\mu = \overline{X}$ y el denominador por $\mu = \overline{X}$, $\Sigma = N^{-1}A$, donde la sustitución de estos valores da el resultado deseado. □

El siguiente resultado nos muestra que la prueba de razón de verosimilitud es sesgada. Notemos que rechazar H_0 para valores pequeños de Λ es equivalente a rechazar H_0 para valores pequeños de

$$V = \text{etr}(-A/2)(\text{det}A)^{N/2}.$$

Teorema 2.3. *Para el contraste de hipótesis $H_0 : \Sigma = I_p$ vs $H_1 : \Sigma \neq I_p$, la prueba de razón de verosimilitud con región crítica $V \leqslant c$ es sesgada.*

Haciendo una ligera modificación al estadístico de razón de verosimilitud, la prueba es insesgada. Tal modificación del estadístico es

$$\Lambda^* = \left(\frac{\text{e}}{n}\right)^{pn/2}\text{etr}(-A/2)(\text{det}A)^{n/2},$$

y es obtenido por Λ reemplazando en tamaño de muestra N por los grados de libertad n. Por lo tanto, la prueba de razón de verosimilitud rechaza $H_0 = I_p$ para valores muy pequeños de Λ^*, o equivalente de

$$V^* = \text{etr}(-A/2)(\text{det}A)^{n/2}.$$

El siguiente resultado que puede ser consultado en [13], proporciona la aproximación de la distribución de $-2\rho\text{log}V^*$ a través de la distribución Ji-cuadrada.

Teorema 2.4. *Cuando la hipótesis $H_0 : \Sigma = I_p$ es verdadera y n es grande, la distribución de $-2\rho\text{log}\Lambda^*$, donde $\rho = 1 - (2p^2 + 3p - 1)/(6n(p + 1))$, sigue aproximadamente una distribución Ji-cuadrada con $f = p(p + 1)/2$, es decir,*

$$\mathbb{P}(-2\rho\text{log}\Lambda^* \leq x) \approx \mathbb{P}(\chi^2_f \leqslant x), \quad \forall x \in \mathbb{R}.$$

Utilizando esta aproximación, una prueba de hipótesis de nivel α, es rechazar H_0 si $-2\rho \log \Lambda^* > \chi^2_f(\alpha)$, donde $\chi^2_f(\alpha)$ es el punto porcentual superior α de la distribución Ji-cuadrada con f grados libertad. Esta prueba será llamada **prueba de razón de verosimilitud (LRT_1)**.

2.1.2. Prueba de razón de verosimilitud corregida (CLRT)

Sean $X_1, X_2, ..., X_n$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$, con $p < n$, y queremos probar las hipótesis (2.2). Sean

$$\widehat{S}_n = \frac{1}{n}A, \quad A = \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})' \quad y$$

$$L^* = \log\Lambda^* = \mathrm{tr}\widehat{S}_n - \ln(\det\widehat{S}_n) - p. \tag{2.3}$$

Definiendo

$$T_n = n * L^*,$$

por el Teorema 2.4, T_n converge a la distribución Ji-cuadrada con $p(p+1)/2$ grados de libertad bajo H_0, cuando p está fija y $n \to \infty$.

El problema del estadístico de razón de verosimilitud es que si hacemos tender a $n, p \to \infty$ la prueba comienza a ser mala, para ello Bai et al. [2] hacen una corrección al estadístico usando algunos resultados de teoría de matrices aleatorias que se presentan a continuación.

2.1.3. Algunos resultados de la teoría de matrices aleatorias

Definición 2.1. *Supongamos que A es una matriz cuadrada de $p \times p$ con eigenvalores λ_i^A, $i = 1, 2, ..., p$ reales.* ***La distribución Empírica Espectral (DEE)*** *de A, denotada como F_n^A, está dada por*

$$F_n^A(x) = \frac{1}{p}\sum_{i=1}^{p} 1_{\lambda_i^A \leqslant x}, \quad x \in \mathbb{R}.$$

Consideraremos a la DEE F_n^A de una matriz aleatoria A la cual converge a la distribución límite espectral F^A. Para hacer inferencia estadística acerca del parámetro $\theta = \int f(x)dF^A(x)$, se usa el siguiente estimador

$$\widehat{\theta} = \int f(x)dF_n^A(x) = \frac{1}{p}\sum_{i=1}^{p} f(\lambda_i^A),$$

el cual recibe el nombre de **estadístico espectral lineal (ESL)** de una matriz aleatoria A.

Sean $a(\theta) = (1 - \sqrt{\theta})^2$ y $b(\theta) = (1 + \sqrt{\theta})$, con $0 < \theta \leqslant 1$. La **distribución de Marchenko-Pastur** de índice θ, denotada como F^θ, es la distribución en $[a(\theta), b(\theta)]$, cuya función de densidad es

$$g_\theta = \frac{1}{2\pi\theta x}\sqrt{(b(\theta)-x)(x-a(\theta))}, \quad a(\theta) \leqslant x \leqslant b(\theta).$$

Supongamos que

$$y_n = \frac{p}{n} \to y \in (0,1), \quad n, p \to \infty,$$

y sea F^y, F^{y_n} la ley de Marchenko-Pastur de índices y y y_n, respectivamente.

Sean $\{X_{ij}\}$ variables aleatorias i.i.d. con media 0 y varianza 1. Sea $X_i = (X_{1i}, X_{2i}, \ldots, X_{pi})$, $i = 1, 2, \ldots, n$. Los vectores $X_1, X_2, ..., X_n$ son vectores aleatorios i.i.d. de una distribución p-dimensional con media $\mathbf{0}_p$ y matriz de covarianza I_p.

Sea $\mathcal{U}$ un conjunto abierto del plano complejo, que contiene a $[I_{(0,1)}(y)a(y), b(y)]$, y sea $\mathcal{A}$ el conjunto de funciones analíticas $f : \mathcal{U} \mapsto \mathbb{C}$. Consideremos el proceso empírico $G_n := \{G_n(f)\}$ indexado por $\mathcal{A}$,

$$G_n(f) = p \cdot \int_{-\infty}^{+\infty} f(x)[F_n - F^{y_n}](dx), \quad f \in \mathcal{A},$$

donde F_n es la DEE de la matriz de covarianza muestral de las X_i's. El siguiente resultado es muy importante para poder obtener el estadístico de prueba y tener así la convergencia asintótica a una distribución normal estándar.

Teorema 2.5. *Sean $f_1, f_2, ..., f_k \in \mathcal{A}$, y sean $\{X_{ij}\}$ variables aleatorias i.i.d. con media cero, varianza igual a uno y $E(X_{ij}^4)$=3. Supongamos que $p/n \to y \in (0,1)$, cuando $n, p \to \infty$. Entonces $(G_n(f_1), G_n(f_2), ..., G_n(f_k))$ converge débilmente a un vector Gausiano k-dimensional con vector de medias*

$$m(f_j) = \frac{f_j(a(y)) + f_j(b(y))}{4} - \frac{1}{2\pi}\int_{a(y)}^{b(y)} \frac{f_j(x)}{\sqrt{4y-(x-1-y)^2}}, \quad j = 1, ..., k,$$

y función covarianza

$$v(f_j, f_l) = -\frac{1}{2\pi^2}\oint\oint \frac{f_j(z_1)f_l(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2} d\underline{m}(z_1) d\underline{m}(z_2),$$

donde $\underline{m}(z) \equiv m_{\underline{F}^y}(z)$ es la transformada de Stieltjes de $\underline{F}^y \equiv (1-y)I_{[0,\infty)} + yF^y$.

Como consecuencia del resultado anterior, Bai et al. [2] demostraron el siguiente resultado.

Teorema 2.6 (Bai, et al., (2009)). *Sean $X_1, X_2, ..., X_n$ vectores aleatorios independientes con distribución $N_p(\mu, \Sigma)$. Suponer además $p/n \longrightarrow y \in (0,1)$ cuando $n, p \longrightarrow \infty$. Sean L^* como en la ecuación (2.3) y $g(x) = x - \log x - 1$. Entonces bajo H_0 y cuando $n \longrightarrow \infty$*

$$\widetilde{T}_n = v(g)^{-1/2}[L^* - p \cdot F^{y_n}(g) - m(g)] \longrightarrow N(0,1), \tag{2.4}$$

donde

$$m(g) = -\frac{log(1-y)}{2}, \tag{2.5}$$

$$v(g) = -2\log(1-y) - 2y, \tag{2.6}$$

$$F^{y_n}(g) = 1 - \frac{y_n - 1}{y_n}\log(1-y_n). \tag{2.7}$$

*El estadístico $\widetilde{T}_n$ será llamado **estadístico de razón de verosimilitud corregido**.*

La demostración del resultado anterior se presenta en el Apéndice A. Una prueba de hipótesis para (2.2) basada en el estadístico (2.4), consiste en rechazar H_0 con un nivel de significancia α si $\widetilde{T}_n > z_\alpha$, donde z_α es el punto porcentual superior α de la distribución normal estándar $N(0,1)$. Esta prueba de hipótesis será llamada **prueba de razón de verosimilitud corregida (CLRT)** de nivel α.

2.1.4. Prueba de Ledoit & Wolf (LW)

La siguiente prueba fue propuesta por Ledoit y Wolf en [10].

Definición 2.2. *Sean $X_1, X_2, ..., X_N$ vectores aleatorios i.i.d. $N_p(\mu, \Sigma)$ y estamos interesados en probar el juego de hipótesis (2.2). El estadístico*

$$W = \frac{1}{p}\text{tr}[(S_n - I_p)^2] - \frac{p}{n}\left[\frac{1}{p}\text{tr}S_n\right]^2 + \frac{p}{n}, \quad \textit{con } n = N - 1,$$

*es llamado **estadístico de Ledoit & Wolf**.*

Bajo H_0, cuando $n \to \infty$ mientras p permanece fija, se tiene que

$$nW - p \longrightarrow \frac{2}{p}\chi^2_{p(p+1)/2-1} - p.$$

Teorema 2.7 (Ledoit & Wolf (2002)). *Suponer que $p/n \to y \in (0, \infty)$, cuando $n, p \to \infty$, entonces bajo H_0*

$$nW - p \longrightarrow N(1,4).$$

Una prueba de hipótesis para (2.2) basada en este estadístico, consiste en rechazar H_0 con un nivel de significancia α si $(nW - p - 1)/2 > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0, 1)$. Esta prueba será llamada **prueba de Ledoit & Wolf** de nivel α.

2.1.5. Prueba de Tracy-Widom (T-W)

La distribución de Wishart, dada por John Wishart en 1928, fue la primera distribución propuesta para matrices aleatorias. A continuación se presenta la definición de esta distribución, la cual puede ser consultada en [1] y [13].

Definición 2.3. *Sea $A = Z'Z$, donde las filas de la matriz Z de $n \times p$ son i.i.d. $N_p(\mathbf{0}, \Sigma)$, entonces se dice que A tiene* ***distribución Wishart*** *con n grados de libertad y matriz de covarianza Σ. Notación: $A \sim W_p(n, \Sigma)$.*

El siguiente teorema propuesto por Johnstone [9], muestra cuál es la distribución asintótica del eigenvalor más grande de una matriz Wishart. Este resultado será utilizado para realizar pruebas de hipótesis acerca de la matriz de covarianza poblacional de datos normales multivariados de dimensión alta.

Teorema 2.8 (Johnstone (2001)). *Sea $A \sim W_p(n, I_p)$ y sea l_1 el eigenvalor más grande de A. Si $p/n \to \gamma > 0$, cuando $n, p \to \infty$, entonces*

$$\frac{l_1 - \mu_{np}}{\sigma_{np}} \xrightarrow[dist]{} F_1,$$

donde las constantes de centralización y escala son

$$\mu_{np} = \left(\sqrt{n-1} + \sqrt{p}\right)^2$$
$$\sigma_{np} = \left(\sqrt{n-1} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3},$$

y F_1 es la función de distribución de la ***ley de Tracy-Widom*** *definida como*

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty \left[q(x) + (x-s)q^2(x)\right] dx\right), \quad s \in \mathbb{R},$$

donde q es la solución de la ecuación diferencial de Painlevé II

$$q''(x) = xq(x) + 2q^3(x), \quad q(x) \sim Ai(x), \quad \text{cuando} \quad x \to +\infty,$$

y $Ai(x)$ es la función Airy.

Si tenemos las observaciones de una muestra aleatoria de tamaño n de una normal multivariada $N_p(\mu, \Sigma)$ y l_1 es el eigenvalor más grande de la matriz de covarianza muestral S_n, con $n = N - 1$, una prueba de hipótesis de nivel α rechaza H_0 si

$$\frac{nl_1 - \mu_{np}}{\sigma_{np}}$$

es mayor que el punto porcentual superior α de la distribución Tracy-Widom F_1, denotado por $F_1(\alpha)$, donde μ_{np} y σ_{np} son las constantes de centralización y escala del teorema anterior. Esta prueba será llamada **prueba de Tracy-Widom** de nivel α.

2.1.6. Prueba de Cai

Como ya se mencionó antes, estamos interesados en probar

$$H_0 : \Sigma = I_p.$$

En [5] consideran probar la hipótesis anterior considerando una hipótesis alternativa

$$H_1 : \Sigma \in \Theta, \quad \Theta = \{\Sigma : \| \Sigma - I_p \|_F \geqslant \epsilon_n\}, \tag{2.8}$$

donde $\epsilon > 0$,

$$\| A \|_F = \left(\sum_{i,j} a_{ij}^2 \right)^{1/2}$$

es la norma de Frobenius de una matriz $A = (a_{ij})$. La dificultad de la prueba entre H_0 y H_1 depende del valor de ϵ_n; entre más pequeño es, más difícil es distinguir entre las dos hipótesis.

Sean $X_1, ..., X_n$ vectores aleatorios *i.i.d.* con distribución $N_p(0, \Sigma)$. Definimos el estadístico

$$T_n = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} h(X_i, X_j),$$

donde

$$h(X_i, X_j) = (X_i' X_j)^2 - (X_i' X_i + X_j' X_j) + p.$$

Para contrastar H_0 y H_1 primero se estima la norma de Frobenius al cuadrado

$$\| \Sigma - I_p \|_F^2 = \operatorname{tr}(\Sigma - I_p)^2$$

mediante el estadístico $T_n = T_n(X_1, ..., X_n)$, y rechazamos H_0 si T_n es lo suficientemente grande. Note que $\mathbb{E}_\Sigma h(X_i, X_j) = \operatorname{tr}(\Sigma - I_p)^2$. El siguiente resultado proporciona la distribución asintótica de T_n.

Teorema 2.9 (Cai, et. al (2013)). *Supongamos que $p \to \infty$ cuando $n \to \infty$. Si una sucesión de matrices de covarianza satisface*

$$\mathrm{tr}(\Sigma^2) \to \infty \ y \ \mathrm{tr}(\Sigma^4)/\mathrm{tr}^2(\Sigma^2) \to 0,$$

cuando $n \to \infty$, entonces

$$\frac{T_n - \mu_n(\Sigma)}{\sigma_n(\Sigma}) \to N(0,1)$$

donde

$$\mu_n(\Sigma) = \mathbb{E}_\Sigma(T_n) = \mathrm{tr}(\Sigma - I_p)^2$$
$$\sigma_n^2(\Sigma) = \mathrm{var}_\Sigma(T_n) = \frac{4}{n(n-1)}(\mathrm{tr}^2(\Sigma^2) + \mathrm{tr}(\Sigma^4)) + \frac{8}{n}\mathrm{tr}(\Sigma^2(\Sigma - I_p)^2).$$

Notemos que si $p \to \infty$, la matriz identidad I_p satisface el teorema anterior, además $\mu_n(I_p) = 0$ y $\sigma_n^2(I_p) = 4p(p+1)/n(n-1)$.

El teorema anterior proporciona el comportamiento de T_n bajo H_0, por lo que una prueba de nivel α basada en el estadístico T_n está dada por

$$\psi = I\left(T_n > z_{1-\alpha}2\sqrt{\frac{p(p+1)}{n(n-1)}}\right), \tag{2.9}$$

donde $I(\cdot)$ es la función indicadora, $z_{1-\alpha}$ es el cuantil superior α de la distribución normal $N(0,1)$. Además de especificar la región de rechazo en (2.9), el teorema anterior se puede usar para estudiar la potencia de la prueba sobre una sucesión de alternativas simples.

Para estudiar la potencia de la prueba ψ en (2.9) sobre $H_1 : \Sigma \in \Theta(b)$, donde $\Theta(b) = \{\Sigma : \| \Sigma - I_p \|_F \geqslant b\sqrt{p/n}\}$, $b > 0$, tenemos el siguiente resultado.

Teorema 2.10. *Supongamos que $p \to \infty$ cuando $n \to \infty$. Para cualquier nivel de significancia $\alpha \in (0,1)$ y $\Theta(b)$, la potencia de la prueba en (2.9) satisface*

$$\lim_{n\to\infty} \inf_{\Theta(b)} \mathbb{E}_\Sigma \psi = 1 - \Phi\left(z_{1-\alpha} - \frac{b^2}{2}\right) > \alpha.$$

Además, para $b_n \to \infty$, $\lim_{n\to\infty} \inf_{\Theta(b_n)} \mathbb{E}_\Sigma \psi = 1$.

Este resultado muestra que la prueba ψ puede distinguir entre la hipótesis nula $H_0 : \Sigma = I_p$ y la alternativa de (2.8) con potencia tendiendo a uno cuando $b = b_n \to \infty$.

2.1.7. Prueba de Srivastava (T_{2s}, T_2)

Las pruebas que se presentan en esta sección fueron propuestas por Srivastava [15] y Srivastava et al. [16]. Antes de mencionar en que consisten estas pruebas consideramos las siguientes definiciones.

Sean $X_1, ..., X_N$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$, sean $\overline{X}$ y S_n la media muestral y la matriz covarianza muestral respectivamente, dadas por (2.1), y sea

$$a_i = \frac{\mathrm{tr}\Sigma^i}{p}, \quad i = 1, ..., 8.$$

Considerar los siguientes supuestos

$$\begin{aligned} &a)\ Si\ p \to \infty,\ a_i \to a_i^0,\ 0 < a_i^0 < \infty, \quad i = 1, ..., 8 \\ &b)\ n = \mathcal{O}(p^\delta), \quad 0 < \delta \leqslant 1. \end{aligned} \tag{2.10}$$

Lema 2.1. *Bajo el supuesto a), y haciendo $n \to \infty$, un estimador insesgado y consistente de a_1 y a_2 están dados respectivamente por*

$$\widehat{a}_1 = \frac{\mathrm{tr}S_n}{p}, \quad \widehat{a}_{2s} = \frac{n^2}{(n-1)(n+2)p}\left[\mathrm{tr}S_n^2 - \frac{1}{n}(\mathrm{tr}S_n)^2\right].$$

La demostración de este lema puede verse en [15].

Teorema 2.11 (Srivastava (2005)). *Sean $X_1, ..., X_N$ vectores aleatorios i.i.d. $N_p(\mu, \Sigma)$, considerar los supuestos (2.10). Bajo $H_0 : \Sigma = I_p$, cuando $n, p \to \infty$, se tiene que*

$$T_{2s} = \frac{n}{2}\left(\widehat{a}_{2s} - 2\widehat{a}_1 + 1\right) \to N(0, 1).$$

Por lo que una prueba de hipótesis para (2.2) basada en T_{2s}, consiste en rechazar H_0 con un nivel de significancia α si $T_{2s} > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0, 1)$. Esta prueba será llamada **prueba de Srivastava** T_{2s} de nivel α.

En [16] proponen otra prueba, pero ahora basados en un nuevo estimador insesgado a_2 cuyo cálculo computacional es fácil, el cuál está dado por

$$\widehat{a}_2 = \frac{1}{f}\left[(N-1)n\mathrm{tr}(M^2) - Nn\mathrm{tr}(D^2) + (\mathrm{tr}D^2)\right],$$

donde $f = pN(N-1)(N-2)(N-3)$, $M = X'X$, $X = (x_1, ..., x_N)$, $Y = (y_1, ..., y_N)$, $D = \mathrm{diag}(y_1'y_1, ..., y_N'y_N)$, con $y_i = x_i - \overline{x}$.

Teorema 2.12 (Srivastava, et al. (2014)). *Sean $X_1, ..., X_N$ vectores aleatorios i.i.d. $N_p(\mu, \Sigma)$, considerar el supuesto $N = \mathcal{O}(p^\delta)$, $1/2 < \delta < 1$. Bajo $H_0 : \Sigma = I_p$, cuando $n, p \to \infty$, se tiene que*

$$T_2 = \frac{n}{2}\left(\widehat{a}_2 - 2\widehat{a}_1 + 1\right) \to N(0, 1).$$

Por lo que una prueba de hipótesis para (2.2) basada en T_2, consiste en rechazar H_0 con un nivel de significancia α si $T_2 > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0, 1)$. Esta prueba será llamada **prueba de Srivastava** T_2 de nivel α.

2.2. Pruebas de hipótesis para $H_0 : \Sigma = \lambda I_p$

A continuación se presentan pruebas de esfericidad, es decir, pruebas para la hipótesis de que la matriz de covarianza poblacional sea un múltiplo de la identidad.

Sean $X_1, X_2, ..., X_N$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$. Supongamos ahora que estamos interesados en el juego de hipótesis

$$H_0 : \Sigma = \lambda I_p \quad \text{vs} \quad H_1 : \Sigma \neq \lambda I_p, \tag{2.11}$$

donde λ es desconocida.

2.2.1. Prueba de razón de verosimilitud (LRT_2)

Una prueba muy utilizada en el Análisis Multivariado clásico para contrastar (2.11) es la basada en el estadístico de elipticidad (ver [13]), el cual se define a continuación.

Teorema 2.13. *Sean $X_1, ..., X_N$ vectores aleatorios independiente con distribución N_p (μ, Σ) y sea $A = nS_n$, donde S_n es la matriz de covarianza muestral dada por (2.1). La **prueba de razón de verosimilitud** de tamaño α de $H_0 : \Sigma = \lambda I_p$, donde λ es desconocido, rechaza H_0 si*

$$V \equiv \frac{\det A}{\left(\frac{1}{p}\mathrm{tr}A\right)^p} = \frac{\det S_n}{\left(\frac{1}{p}\mathrm{tr}S_n\right)^p} \leqslant k_\alpha,$$

donde k_α es elegido de tal forma que el tamaño de la prueba es α.

Demostración. La función de verosimilitud de la distribución normal multivariada es

$$L(\mu, \Sigma) = (2\pi)^{-Np/2}(\det\Sigma)^{-N/2}\mathrm{etr}\left(-\frac{1}{2}\Sigma^{-1}A\right)\exp\left[-\frac{1}{2}N(\overline{X} - \mu)'\Sigma^{-1}(\overline{X} - \mu)\right].$$

De la Definición 1.6, el estadístico de razón de verosimilitud es

$$\Lambda = \frac{\text{Sup}_{\mu\in\mathbb{R}^p,\lambda>0} L(\mu, \lambda I_p)}{\text{Sup}_{\mu\in\mathbb{R}^p\Sigma>0} L(\mu,\Sigma)}.$$

Para el denominador tenemos que

$$\text{Sup}_{\mu\in\mathbb{R}^p\Sigma>0} L(\mu,\Sigma) = L(\overline{X}, \widehat{\Sigma}),$$

donde $\overline{X}$ y $\widehat{\Sigma}$ son los EMV dados por el Teorema 2.1. Tomando $\widehat{\Sigma}^{-1} = N^{-1}A$, tenemos lo siguiente

$$\begin{aligned}
L(\overline{X}, \widehat{\Sigma}) &= (\det\widehat{\Sigma})^{-N/2}\text{etr}\left(-\frac{1}{2}\widehat{\Sigma}^{-1}A\right)\exp\left[-\frac{1}{2}N(\overline{X}-\overline{X})'\widehat{\Sigma}^{-1}(\overline{X}-\overline{X})\right] \\
&= (\det N^{-1}A)^{-N/2}\text{etr}\left(-\frac{1}{2}IN\right) \\
&= (\det N^{-1}A)^{-N/2}\exp\left(-\frac{N}{2}\text{tr}(I)\right) \\
&= (\det N^{-1}A)^{-N/2}\text{etr}\left(-\frac{Np}{2}\right) \\
&= N^{(Np)/2}(\det A)^{-N/2}\exp\left(-\frac{Np}{2}\right).
\end{aligned}$$

Mientras que el numerador es

$$\begin{aligned}
L(\overline{X}, \lambda I_p) &= (\det\lambda I_p)^{-N/2}\text{etr}\left(-\frac{1}{2}(\lambda I_p)^{-1}A\right)\exp\left[-\frac{1}{2}N(\overline{X}-\overline{X})'(\lambda I_p)^{-1}(\overline{X}-\overline{X})\right] \\
&= \lambda^{-Np/2}\exp\left(-\frac{1}{2\lambda}\text{tr}(A)\right).
\end{aligned}$$

Notemos que la traza de A es una constante, es decir $\text{tr}(A) = c$, entonces tenemos

$$L(\overline{X}, \lambda I) = \lambda^{-Np/2}\exp\left(-\frac{c}{2\lambda}\right),$$

aplicando logaritmo

$$l(\lambda) = \log(L(\overline{X}, \lambda I)) = -(pn)/2\log(\lambda) - c/2\lambda.$$

Resolviendo la ecuación $(d/d\lambda)l(\lambda) = 0$ se tiene que $\lambda = c/Np$ y por lo tanto

$$L(\overline{X}, \lambda I_p) = \left(\frac{\mathrm{tr}(A)}{Np}\right)^{-Np/2} \exp\left(-\frac{Np}{2}\right).$$

Así, el estadístico de razón de verosimilitud es

$$\begin{aligned}\Lambda &= \frac{\left(\frac{\mathrm{tr}(A)}{Np}\right)^{-Np/2} \exp\left(-\frac{Np}{2}\right)}{N^{(Np)/2}(\det A)^{-N/2}\exp\left(-\frac{Np}{2}\right)} \\ &= \frac{(\det A)^{N/2} p^{-Np/2} N^{(Np)/2}}{(\mathrm{tr}(A))^{-Np/2} N^{(Np)/2}} \\ &= \left(\frac{\det A}{\left(\frac{1}{p}\mathrm{tr}(A)\right)^p}\right)^{N/2}.\end{aligned}$$

La prueba de razón de verosimilitud rechaza H_0 si el estadístico de razón de verosimilitud Λ es pequeño; notemos que esto es equivalente a rechazar H_0 si

$$V = \Lambda^{2/N} = \frac{\det A}{\left(\frac{1}{p}\mathrm{tr}(A)\right)^p}$$

es pequeño y así la prueba esta completa. □

El estadístico que aparece en la prueba de razón de verosimilitud

$$V = \frac{\det S_n}{(\mathrm{tr}S_n/p)^p},$$

es llamado **estadístico de elipticidad**.

La prueba de hipótesis del teorema anterior consiste en rechazar la hipótesis nula con un nivel de significancia α si $V \leq k_\alpha$, donde k_α es el punto porcentual inferior α de la distribución de V (ver [1] y [13]). Si la hipótesis nula es verdadera es claro que V debe ser cercano a uno.

Teorema 2.14. *Cuando la hipótesis $H_0 : \Sigma = \lambda I_p$ es verdadera, la distribución de $-n\rho \log V$, donde $\rho = 1 - (2p^2 + p + 2)/6np$, sigue aproximadamente una distribución chi-cuadrada con $f = (p+2)(p-1)/2$ grados de libertad, cuando n es grande, es decir,*

$$\mathbb{P}(-n\rho \log V \leqslant x) \approx \mathbb{P}(\chi^2_f \leqslant x), \quad \forall x \in \mathbb{R}.$$

Utilizando esta aproximación, una prueba de hipótesis de nivel α, es rechazar H_0 si $-n\rho \log V > \chi^2_f(\alpha)$, donde $\chi^2_f(\alpha)$ es el punto porcentual superior α de la distribución Ji-cuadrada con f grados libertad. Esta prueba será llamada **prueba de razón de verosimilitud (LRT_2)**.

2.2.2. Prueba de John (J)

La siguiente prueba fue propuesta por John [10]. Sean $X_1, X_2, ..., X_N$ vectores aleatorios i.i.d. con distribución $N_p(\mu, \Sigma)$ y estamos interesados en el juego de hipótesis (2.11). El estadístico

$$U = \frac{1}{p}\text{tr}\left[\left(\frac{p}{\text{tr}S_n}\right)S_n^2 - I_p\right], \tag{2.12}$$

es llamado **estadístico de John**. Bajo H_0 cuando $n \to \infty$ mientras p permanece fija, se tiene que

$$nU - p \longrightarrow \frac{2}{p}\chi^2_{p(p+1)/2-1} - p, \quad (N = n-1).$$

Teorema 2.15 (John (2002)). *Suponer que $p/n \to y \in (0, \infty)$, cuando $n, p \to \infty$, entonces bajo H_0*

$$nU - p \longrightarrow N(1, 4).$$

Una prueba de hipótesis para (2.11) basada en este estadístico, consiste en rechazar H_0 con un nivel de significancia α si $(nU - p - 1)/2 > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0, 1)$. Esta prueba será llamada **prueba de John** de nivel α.

2.2.3. Prueba cuasi-razón de verosimilitud (QLRT)

Como estamos interesados en contrastar el juego de hipótesis (2.11), esta prueba puede ser vista como una extensión de la prueba de razón de verosimilitud (ver [11]). La LRT requiere que $p \leqslant n$, sin embargo la QLRT se puede usar cuando $p > n$.

Definición 2.4. *Sean $X_1, X_2, ..., X_n$ vectores aleatorios i.i.d. $N_p(0, \Sigma)$. El estadístico de* ***cuasi-razón de verosimilitud*** *es definido como*

$$L_n = \frac{p}{n}\log\left(\frac{\left(\frac{1}{n}\sum_{i=1}^n \tilde{\lambda}_i\right)^n}{\prod_{i=1}^n \tilde{\lambda}_i}\right),$$

donde $\tilde{\lambda}_{i1 \leqslant i \leqslant n}$ son los eigenvalores de la matriz $\frac{1}{p}X'X$, donde $X = (X_1, \ldots, X_n)$ es la matriz de $p \times n$ cuyas columnas son los vectores X_i. Notar que la matriz $X'X$ tiene los mismos eigenvalores no cero que XX'.

Por los resultados de [11] tenemos el siguiente teorema.

Teorema 2.16. *Suponer que $p/n \to \infty$ cuando $n \to \infty$, entonces bajo H_0*

$$L_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{v_4 - 2}{2} \longrightarrow N(0,1),$$

donde v_4 es el cuarto momento de la distribución normal estándar.

Por lo tanto, una prueba de hipótesis para (2.11) basada en este estadístico, consiste en rechazar H_0 con un nivel de significancia α si $L_n - \frac{n}{2} - \frac{n^2}{6p} - \frac{v_4 - 2}{2} > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0,1)$. Esta prueba será llamada **prueba cuasi-razón de verosimilitud** de nivel α.

2.2.4. Prueba de Srivastava (T_{1s}, T_1)

La siguientes pruebas son proporcionadas por Srivastava [15] y Srivastava et al. [16]. Considerar los estimadores insesgados de a_1 y a_2 del Lema 2.1. Definimos

$$T_{1s} = \frac{n}{2}\left(\frac{\widehat{a}_{2s}}{\widehat{a}_1^2} - 1\right).$$

Teorema 2.17 (Srivastava (2005)). *Sean $X_1, ..., X_N$ vectores aleatorios i.i.d. $N_p(\mu, \Sigma)$, considerar los supuestos (2.10). Bajo $H_0 : \Sigma = \lambda I$, cuando $n, p \to \infty$, se tiene que*

$$T_{1s} \to N(0,1).$$

Una prueba de hipótesis para (2.11) basada en T_{1s}, consiste en rechazar H_0 con un nivel de significancia α si $T_{1s} > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0,1)$. Esta prueba será llamada **prueba de Srivastava** T_{1s} de nivel α.

En [16] proponen otra prueba, pero ahora basados en un nuevo estimador insesgado de a_2 cuyo cálculo computacional es fácil, el cual es

$$\widehat{a}_2 = \frac{1}{f}\left[(N-1)n\mathrm{tr}(M^2) - Nn\mathrm{tr}(D^2) + (\mathrm{tr}D^2)\right],$$

donde $f = pN(N-1)(N-2)(N-3)$, $M = X'X$, $X = (x_1, \ldots, x_N)$, $Y = (y_1, ..., y_N)$ $D = \mathrm{diag}(y_1'y_1, ..., y_N'y_N)$, con $y_i = x_i - \overline{x}$.

Sustituyendo $\widehat{a}_2$ por $\widehat{a}_{2s}$ de T_{1s} tenemos

$$T_1 = \frac{n}{2}\left(\frac{\widehat{a}_2}{\widehat{a}_1^2} - 1\right).$$

Teorema 2.18 (Srivastava, et al. (2014)). *Sean $X_1, ..., X_N$ vectores aleatorios i.i.d. $N_p(\mu, \Sigma)$, considerar $N = \mathcal{O}(p^\delta)$, $1/2 < \delta < 1$. Bajo $H_0 : \Sigma = \lambda I$, cuando $n, p \to \infty$, se tiene que*

$$T_1 \to N(0, 1).$$

Una prueba de hipótesis para (2.11) basada en T_1, consiste en rechazar H_0 con un nivel de significancia α si $T_1 > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución normal estándar $N(0, 1)$. Esta prueba será llamada **prueba de Srivastava** T_1 de nivel α.

2.2.5. Prueba de Zou

La prueba que se presenta a continuación fue propuesta Zou et al. [17]. Sean $X_1, ...X_n$ vectores aleatorios de una **distribución elíptica** p-variada con función de densidad

$$\det(\Sigma_p)^{-1/2} g_p(\|\Sigma_p^{-1/2}(X - \theta_p)\|),$$

donde $\|X\| = (X'X)^{1/2}$ es la norma Euclidiana del vector X, θ_p es el centro simétrico de la distribución y Σ_p es una matriz positiva definida. La matriz Σ_p es la covarianza entre las p-variables, que puede expresarse como $\Sigma_p = \sigma_p \Lambda_p$, donde $\sigma_p = \sigma(\Sigma_p)$ es el parámetro escala y $\Lambda_p = \sigma_p^{-1}\Sigma_p$. El parámetro escala satisface $\sigma(I_p) = 1$.

Estamos interesados en probar $H_0 : \Sigma_p = \lambda I_p$, lo cual es equivalente a $\Lambda_p = I_p$.

Definición 2.5. *La **función signo** multivariada se define como*

$$U(X) = \|X\|^{-1} X.$$

Los signos observados para las X_i's son

$$U_i = U(X_i - \theta_p).$$

Como la distribución normal multivariada es una distribución elíptica, consideraremos el siguiente estadístico para observaciones con distribución normal multivariada donde se tiene que $\theta_p = \mu$ y su estimador está dado por $\widehat{\theta}_{n,p} = \overline{x}$.

Sean $X_1, ..., X_n$ vectores aleatorios *i.i.d.* con distribución $N_p(\mu, \Sigma)$. Considerar el estadístico

$$\tilde{Q} = \frac{p}{n(n-1)} \sum_{i \neq j} (\widehat{U}'\widehat{U}')^2 - 1,$$

donde $\widehat{U}_i = U(X_i - \overline{x})$.

El siguiente teorema nos da la distribución asintótica de $\tilde{Q}$, el cual puede consultarse en [17].

Teorema 2.19 (Zou, et al. (2014)). *Sea* $R_i = \| X_i - \mu \|$. *Bajo* $H_0 : \Sigma = \lambda I_p$, *si* $p = \mathcal{O}(n^2)$, *entonces*

$$\frac{\tilde{Q} - p\delta_{n,p}}{\tilde{\sigma}_0} \longrightarrow N(0,1)$$

en distribución con $n.p \longrightarrow \infty$, *donde* $\tilde{\sigma}_0^2 = 4(p-1)/[n(n-1)(p+2)]$ *y*

$$\begin{aligned} \delta_{n,p} =& \frac{1}{n^2} \left(2 - \frac{2\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2} + \left[\frac{\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2} \right]^2 \right) \\ &+ \frac{1}{n^3} \left[\frac{8\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2} - 6 \left(\frac{\mathbb{E}(R_i^{-2})}{\mathbb{E}(R_i^{-1})^2} \right)^2 + \frac{2\mathbb{E}(R_i^{-2})\mathbb{E}(R_i^{-3})}{\mathbb{E}(R_i^{-1})^5} - \frac{\mathbb{E}(2R_i^{-3})}{\mathbb{E}(R_i^{-1})^3} \right]. \end{aligned} \tag{2.13}$$

Las cantidades desconocidas en $\delta_{n,p}$ son $\mathbb{E}(R_i^{-2})/\mathbb{E}(R_i^{-1})^2$ y $\mathbb{E}(R_i^{-3})/\mathbb{E}(R_i^{-1})^3$, las cuales pueden estimarse de la siguiente manera. Sean

$$\widehat{R}_i = \|X_i - \widehat{\theta}_{n,p}\|, \quad \widehat{R}_{i*} = \widehat{R}_i + \widehat{\theta}'_{n,p}\widehat{U} - 2^{-1}\widehat{R}_i^{-1}\|\widehat{\theta}_{n,p}\|^2.$$

Así, sustituyendo

$$\mathbb{E}(R_i^{-k})/\mathbb{E}(R_i^{-1})^k$$

por

$$\frac{n^{n-k}\sum_{i=1}^n (\widehat{R}_{i*}^{-k})}{(\sum_{i=1}^n \widehat{R}_{i*}^{-1})^k}$$

en (2.13) obtenemos un estimador de $\delta_{n,p}$, denotado como $\hat{\delta}_{n,p}$.

Una prueba de hipótesis para (2.11) basada en este estadístico, consiste en rechazar H_0 con un nivel de significancia α si $(\tilde{Q} - p\hat{\delta}_{n,p})/\tilde{\sigma}_0 > z_\alpha$, donde z_α es un punto porcentual superior α de la distribución $N(0,1)$. Esta prueba será llamada **prueba de Zou** de nivel α.

Capítulo 3

Estudio de simulación y aplicaciones

A continuación se presenta un estudio de simulación en términos del tamaño y la potencia de las pruebas para comparar su desempeño. Para ambos problemas de hipótesis se consideró $M = 10000$ muestras aleatorias de tamaño $N = n + 1$. Los escenario contemplan varios valores de n y p, en el caso que $n > p$ se tomó $n = 500$ fijo con cinco valores de p y el caso que $p \geqslant n$ se tomó $p = 500$ fijo con cinco valores de n. Se consideró un nivel de significancia de $\alpha = 0.05$ para todas las pruebas y se calculó la proporción de veces en que se rechazaba H_0, si las pruebas son buenas estas proporciones de rechazo deben ser parecidas al nivel de significancia.

Para evaluar la potencia de las pruebas consideramos un subconjunto de matrices que se encuentran dentro de la hipótesis alternativa H_1 como en [14], que son de la siguiente forma:

$$\Sigma = I_p + hvv',$$

donde I_p es la matriz identidad de $p \times p$, h es un escalar y v es un vector unitario. El vector v se generó de forma aleatoria y h se hizo variar en un rango de valores.

3.1. Estudio de simulación para $H_0 : \Sigma = I_p$

Las pruebas correspondientes al problema de hipótesis (2.2) son: LRT_1, TW, LW, CLRT, Cai, Srivastava T_{2s}, T_2. En el caso clásico pueden aplicarse todas las pruebas y para el caso de dimensión alta pueden aplicarse todas las pruebas excepto la prueba LRT_1 y la CLRT.

3.1.1. Tamaño de las pruebas

Los resultados de los tamaños de las pruebas se exponen en los cuadros 3.1-3.2.

n, p	LRT_1	TW	CLRT	LW	Cai	T_{2s}	T_2
500,25	0.0740	0.0498	0.0555	0.0541	0.0526	0.0546	0.0552
500,50	0.2240	0.048	0.0526	0.0538	0.0519	0.0541	0.0535
500,100	0.9753	0.052	0.0526	0.0539	0.0527	0.0525	0.0531
500,200	1	0.0517	0.0513	0.0493	0.05	0.0491	0.05
500,400	1	0.0536	0.0509	0.0492	0.0487	0.0495	0.0503

Cuadro 3.1: Comparación de las pruebas con respecto al tamaño para el caso en que $p < n$.

n, p	TW	LW	Cai	T_{2s}	T_2
25,500	0.0503	0.055	0.0537	0.0541	0.0591
50,500	0.0494	0.0504	0.0509	0.0491	0.0514
100,500	0.0526	0.0527	0.0483	0.0485	0.0491
200,500	0.051	0.0549	0.0547	0.0546	0.0546
400,500	0.0528	0.0501	0.0516	0.0498	0.0508

Cuadro 3.2: Comparación de las pruebas con respecto al tamaño para el caso en que $p \geqslant n$

Las simulaciones llevadas a cabo nos muestras que las proporciones en que se rechaza la hipótesis nula de los estadísticos TW, CLRT, LW, Cai, T_{2s} y T_2 es cercano al nivel de siginificancia, tanto para el caso clásico $n \gg p$ como el de dimensión alta $p \geqslant n$ (ver cuadro 3.1 y 3.2), esto nos indica que esas pruebas son buenas en término del tamaño. Por otro lado la prueba LRT_1 no es muy buena ya que sus proporciones de rechazo están muy alejados del nivel de significancia, esta prueba podría ser buena si n es lo suficientemente grande con respecto a p. La CLRT corrige el problema que tiene la LRT_1, ya que sin importar que tan grande es n con respecto a p los resultados son buenos.

3.1.2. Potencia de las pruebas

Los resultados de las potencias al variar h, se presentan en los cuadros 3.3-3.6. La figura 3.1 ilustra el comportamiento de las potencias para el caso clásico ($p < n$) y el de dimensión alta ($p \geqslant n$).

$n = 500$, $p = 50$							
h	LRT_1	TW	CLRT	LW	Cai	T_{2s}	T_2
0.3	0.3494	0.1884	0.104	0.1266	0.1215	0.1255	0.1263
0.6	0.6777	0.9047	0.3514	0.5456	0.5373	0.5413	0.5425
0.9	0.9405	0.9997	0.7731	0.9543	0.9514	0.9524	0.9526
1.2	0.9964	1	0.9729	0.9994	0.9992	0.9993	0.9993
1.5	1	1	0.9986	1	1	1	1

Cuadro 3.3: Potencia de las pruebas para el caso en que $p < n$.

$n = 500$, $p = 200$							
h	LRT_1	TW	CLRT	LW	Cai	T_{2s}	T_2
0.5	1	0.1172	0.0751	0.095	0.0948	0.0945	0.0955
1	1	0.8746	0.1596	0.3575	0.3535	0.3538	0.3554
1.5	1	0.9998	0.3455	0.8389	0.8363	0.8361	0.8371
2	1	1	0.6014	0.9945	0.9943	0.9944	0.9943
2.5	1	1	0.8295	0.9999	0.9999	0.9999	0.9999

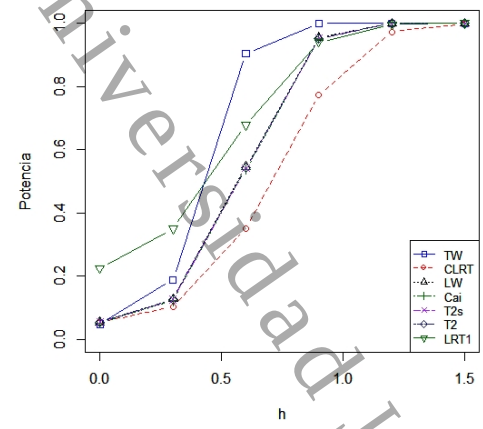
Cuadro 3.4: Potencia de las pruebas para el caso en que $p < n$.

$p = 500$, $n = 50$					
h	TW	LW	Cai	T_{2s}	T_2
1.9	0.0911	0.077	0.0739	0.0732	0.0763
3.8	0.3753	0.196	0.1881	0.1877	0.1935
5.7	0.7866	0.4707	0.4638	0.4686	0.4716
7.6	0.9577	0.7576	0.745	0.7551	.7514
9.5	0.9932	0.9139	0.9096	0.916	0.9148

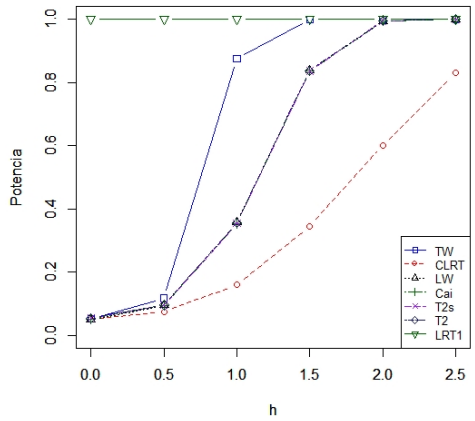
Cuadro 3.5: Potencia de las pruebas para el caso $p \geqslant n$.

$p = 500$, $n = 200$					
h	TW	LW	Cai	T_{2s}	T_2
1.2	0.1086	0.0943	0.0935	0.0945	0.0945
2.4	0.796	0.3261	0.3241	0.322	0.3211
3.6	0.9985	0.7751	0.7703	0.7735	0.7726
4.8	1	0.9788	0.9771	0.979	0.9784
6	1	0.9994	0.9993	0.9994	0.9994

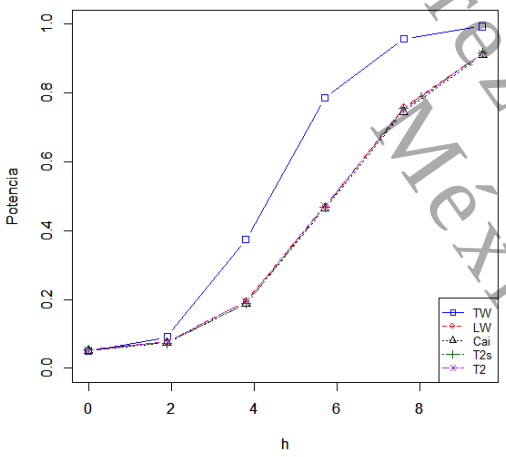
Cuadro 3.6: Potencia de las pruebas para el caso $p \geqslant n$.



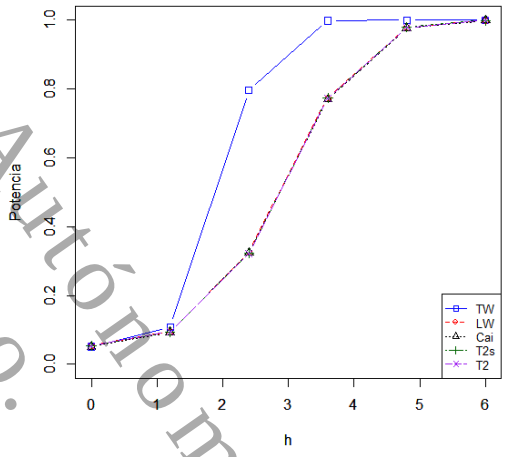
(a) $p = 50, n = 500$



(b) $p = 200, n = 500$



(c) $p = 500, n = 50$



(d) $p = 500, n = 200$

Figura 3.1: Potencia de las pruebas.

En el caso que $n > p$, en los cuadros 3.3 y 3.4 se observa que las potencias de las pruebas son buenas conforme h crece ya que éstas alcanzan o se aproximan al valor de uno, sin embargo la potencia de la prueba TW mostró superioridad con respecto a las otras pruebas (ver figura 3.1 (a) y (b)). Las potencias de las pruebas LW, Cai, T_{2s} y T_2 están tan cercanas que sus gráficas se traslapan al grado de no poder distinguirlas. A pesar de que la potencia de LRT_1 es uno para todos los valores de h en el caso en que $n = 500$ y $p = 200$ (ver cuadro 3.4 y figura 3.1 (b)) la prueba no es buena en términos del tamaño (ver cuadro 3.1), por lo tanto esta prueba no es recomendable en este caso. La potencia de CLRT aunque es la más baja es aceptable. Los cuadros 3.5 y 3.6, para $p \geqslant n$, nos indican que las potencias son aceptables en todas las pruebas, aunque la TW sigue siendo superior a las otras dejándolas abajo (ver figura 3.1 (c) y (d)).

3.2. Estudio de simulación para $H_0 : \Sigma = \lambda I_p$

Las pruebas para el problema de hipótesis (2.11) son: LRT_2, J, QLRT, T_{1s}, T_1 y Zou. Para el caso clásico se pueden utilizar todas las pruebas excepto la QLRT, mientras que para el caso de dimensión alta se pueden usar todas excepto la LRT_2.

3.2.1. Tamaño de las pruebas

Los resultados de los tamaños de las pruebas se exponen en los cuadros 3.7-3.8.

n, p	LRT_2	J	T_{1s}	T_1	Zou
500,25	0.0475	0.0493	0.0492	0.0497	0.052
500,50	0.0505	0.0515	0.052	0.0519	0.0534
500,100	0.0676	0.0526	0.0515	0.0522	0.0501
500,200	0.3673	0.0488	0.0493	0.049	0.0491
500,400	1	0.0497	0.0491	0.0496	0.0502

Cuadro 3.7: Comparación de las pruebas con respecto al tamaño para el caso en que $p < n$.

n, p	J	QLRT	T_{1s}	T_1	Zou
25,500	0.054	0.0615	0.0537	0.0586	0.0604
50,500	0.0498	0.0629	0.049	0.051	0.0519
100,500	0.0519	0.1313	0.0481	0.0489	0.0511
200,500	0.0545	0.9593	0.0546	0.0543	0.0536
400,500	0.0499	1	0.0497	0.0504	0.052

Cuadro 3.8: Comparación de las pruebas con respecto al tamaño para el caso en que $p \geqslant n$

Las proporciones de rechazo de los cuadros 3.7 y 3.8 de las pruebas basadas en los estadísticos J, T_{1s}, T_1 y Zou tienden a ser buenas para $n > p$ y $p \geqslant n$ ya que las proporciones de rechazo se parecen al nivel de significancia deseado. Por otro lado, la LRT_2 es buena para el caso en que n es grande con respecto a p, pero empieza a tener malos resultado si p es cercano a n (ver cuadro 3.7). La prueba basada en el estadístico QLRT tiene buenos resultados si p es lo suficientemente grande con respecto a n, ya que si no lo es esta prueba tiene un mal desempeño (ver cuadro 3.8).

3.2.2. Potencia de las pruebas

Los resultados de las potencias al variar h, se presentan en los cuadros 3.9-3.12. La figura 3.2 ilustra el comportamiento de las potencias para el caso clásico ($p < n$) y el de dimensión alta ($p \geqslant n$).

$n = 500$, $p = 50$					
h	LRT_2	J	T_{1s}	T_1	Zou
0.3	0.0987	0.1161	0.1159	0.1158	0.1108
0.6	0.3323	0.5133	0.5133	0.5115	0.4691
0.9	0.7453	0.9423	0.9427	0.9423	0.9139
1.2	0.9649	0.9989	0.999	0.9989	0.9978
1.5	0.998	1	1	1	1

Cuadro 3.9: Potencia de las pruebas para el caso en que $p < n$.

$n = 500$, $p = 200$					
h	LRT_2	J	T_{1s}	T_1	Zou
0.5	0.4466	0.0921	0.0927	0.0931	0.0924
1	0.6174	0.3445	0.3441	0.3443	0.3309
1.5	0.8065	0.8287	0.8286	0.8287	0.8052
2	0.9349	0.9935	0.9934	0.9936	0.9905
2.5	0.9849	0.9999	0.9999	0.9999	0.9999

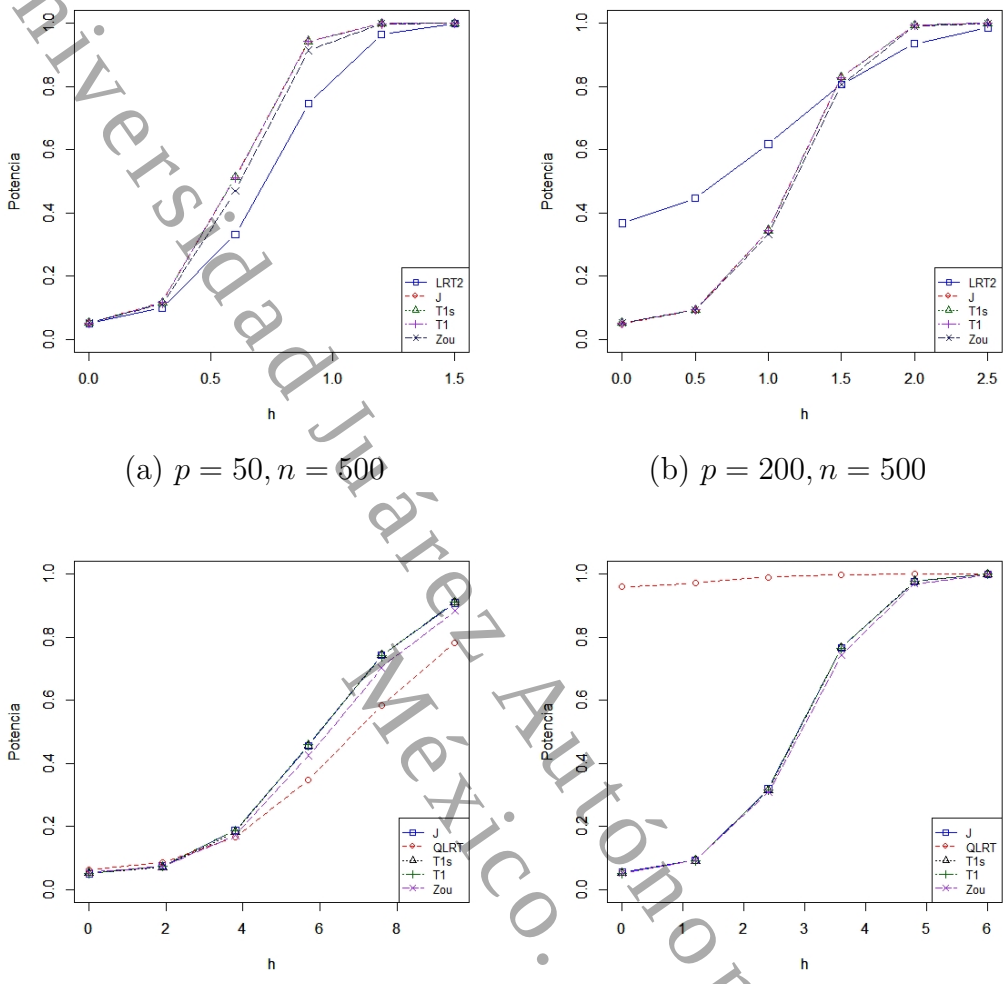
Cuadro 3.10: Potencia de las pruebas para el caso en que $p < n$.

$p = 500$, $n = 50$					
h	J	QLRT	T_{1s}	T_1	Zou
1.9	0.0745	0.0841	0.0699	0.0744	0.0722
3.8	0.1865	0.1672	0.1813	0.1863	0.1716
5.7	0.4546	0.3466	0.4573	0.4582	0.4249
7.6	0.7424	0.5834	0.7436	0.7401	0.7031
9.5	0.908	0.782	0.9106	0.91	0.8834

Cuadro 3.11: Potencia de las pruebas para el caso $p \geqslant n$.

$p = 500$, $n = 200$					
h	J	QLRT	T_{1s}	T_1	Zou
1.2	0.0929	0.9714	0.092	0.0925	0.0923
2.4	0.319	0.9902	0.3151	0.3156	0.3083
3.6	0.7653	0.9979	0.7656	0.7653	0.7434
4.8	0.9768	0.9995	.9775	0.9767	0.9695
6	0.9992	1	0.9993	0.9994	0.999

Cuadro 3.12: Potencia de las pruebas para el caso $p \geqslant n$.



(a) $p = 50, n = 500$ (b) $p = 200, n = 500$

(c) $p = 500, n = 50$ (d) $p = 500, n = 200$

Figura 3.2: Potencia de las pruebas.

Para $n > p$ las potencias de J, T_{1s}, T_1 son tan cercanas que sus gráficas se traslapan y no se pueden distinguir, además estas pruebas son superiores a la prueba LRT_2 (ver figura 3.2 (a) y (b)). En el caso en que $n = 500$ y $p = 50$ la potencia de LRT_2 es buena cuando h crece, mientras que cuando $n = 500$, $p = 200$ y h =0.5, 1, esta prueba tiene mejor potencia que J, T_{1s}, T_2 y Zou (ver cuadro 3.7 y figura 3.2 (b)), sin embargo en términos del tamaño esta prueba no es buena y por tanto no es recomendable (ver cuadro 3.7).

Cuando $p \geqslant n$ las potencias de los cuadros 3.11 y 3.12 muestran que las pruebas J, T_{1s} y T_1 son superiores a las pruebas QLRT y Zou. Cuando $p = 500$, $n = 200$ y h =1.2, 2.4, 3.6, 4.8, la potencia de la QLRT es superior a las restantes (ver cuador 3.12 y figura 3.2 (d)), pero en tamaño esta prueba es mala (ver cuadro 3.8) por lo tanto la prueba no es recomendable en ese caso.

3.3. Ejemplos de aplicación

En esta sección aplicaremos las pruebas a cinco conjuntos de datos reales con el fin de ver el comportamiento de éstas, para ello solo consideraremos el caso de mayor interés en esta tesis, es decir, de dimensión alta ($p \geqslant n$). Para ello consideraremos los problemas de hipótesis planteados en el Capítulo 2:

$$H_0 : \Sigma = I_p \quad \text{vs} \quad H_1 : \Sigma \neq I_p,$$

y

$$H_0 : \Sigma = \lambda I_p \quad \text{vs} \quad H_1 : \Sigma \neq \lambda I_p.$$

Se consideraron las pruebas de TW, J, LW, QLRT, Cai, T_{1s}, T_{2s}, T_1, T_2 y Zou ya que los datos son de dimensión alta y se calculó el p-valor para cada una de la pruebas. Se tomó un nivel de significancia de $\alpha = 0.05$. La descripción de los datos reales y los resultados se presentan a continuación.

Debido a que algunas pruebas del Capítulo 2 consideran datos normales multivariados i.i.d. con media cero y otras con media μ, en el Apéndice A se muestran las transformaciones que hay que aplicar a los datos algunas veces, antes de llevar a cabo las pruebas.

3.3.1. Datos de colon

Este conjunto de datos en un microarreglo ADN correspondiente a los niveles de expresión de 6500 genes humanos en 40 tumores y 22 tejidos normales de colon, los cuales fueron obtenidos de [8]. Se realizó una selección de 2000 genes con las intensidades más altas de la muestra. Por lo tanto, estos datos tienen un tamaño de muestra de $n = 62$ y su dimensión es de $p = 2000$. Los valores de los estadísticos de prueba para cada juego de hipótesis se muestran en las siguientes tablas.

Pruebas	Valor del estadístico
LW	39.83515
TW	60.36933
T_{2s}	211.8479
T_2	213.3861
cai	13886.73

Cuadro 3.13: Pruebas para el juego de hipótesis (2.2)

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = I_p$.

Pruebas	Valor del estadístico
J	131.0829
QLRT	1477.948
T_{1s}	2950.759
T_1	2974.033
Zou	94.7609

Cuadro 3.14: Pruebas para el juego de hipótesis (2.11)

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = \lambda I_p$.

Para ambos problemas de hipótesis al rechazar H_0 se puede concluir que la matriz de covarianza poblacional de los datos no tiene esa estructura. Esto era de esperarse ya que se sabe que existe correlación entre los genes de un mismo individuo.

3.3.2. Datos de leucemia

Este conjunto de datos contiene niveles de expresión genética de 72 pacientes que padecen leucemia linfoblástica aguda o leucemia mieloide aguda, hay 47 y 25 pacientes respectivamente, y se obtienen los datos en microarreglos oligonucleótidos Affymetric. Los datos contemplan $n = 72$ y $p = 3571$ génes, los cuales fueron obtenidos de [8]. El valor de los estadísticos de prueba para cada juego de hipótesis se muestran en las siguientes tablas.

Pruebas	Valor del estadístico
LW	60.36002
TW	171.4266
T_{2s}	352.6394
T_2	355.9055
cai	35455.42

Cuadro 3.15: Pruebas para el juego de hipótesis (2.2)

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = I_p$.

Pruebas	Valor del estadístico
J	150.5623
QLRT	1515.661
T_{1s}	3510.918
T_1	3545.076
Zou	98.88417

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = \lambda I_p$.

Al rechazar H_0 estamos concluyendo que la matriz de covarianza poblacional de los datos no tiene esa estructura. Como se mencionó en el primer ejemplo, esto era de esperarse ya que se sabe que existe correlación entre los genes de un mismo individuo.

3.3.3. Datos de Linfoma

En el siguiente conjunto de datos se examinaron muestras de biopsia de linfoma de células B grandes difuso de 240 pacientes, para determinar la expresión genética con el uso de microarreglos de ADN. Los niveles de expresión en los genes contemplan un total de 7399 genes, los datos fueron obtenidos de [4]. El valor de los estadísticos de prueba para cada juego de hipótesis se muestran en las siguientes tablas.

Pruebas	Valor del estadístico
LW	80.19308
TW	2035.514
T_{2s}	5859.268
T_2	5820.955
cai	361737.7

Cuadro 3.16: Pruebas para el juego de hipótesis (2.2)

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = I_p$.

Pruebas	Valor del estadístico
J	231.2596
QLRT	5263.1241
T_{1s}	23836.63
T_1	23679.95
Zou	181.0721

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = \lambda I_p$.

Al rechazar H_0 estamos concluyendo que la matriz de covarianza poblacional de los datos no tiene esa estructura. Como se mencionó anteriormente, esto era de esperarse ya que se sabe que existe correlación entre los genes de un mismo individuo.

3.3.4. Datos Khan

Los datos contienen muestras de tejidos correspondientes a cuatro tipos distintos de tumores. Cada muestra de tejido contiene 2308 mediciones de expresión genética para 63 sujetos. Estos datos pueden obtenerse del Software *R-project* usando la librería *ISLR*. El valor de los estadísticos de prueba se muestran en las siguientes tablas.

Pruebas	Valor del estadístico
LW	58.05825
TW	215.9323
T_{2s}	635.945
T_2	636.2333
cai	45957.16

Cuadro 3.17: Pruebas para el juego de hipótesis (2.2)

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = I_p$.

Pruebas	Valor del estadístico
J	146.9688
QLRT	1813.874
T_{1s}	3349.27
T_1	3350.813
Zou	109.1436

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = \lambda I_p$.

Al rechazar H_0 estamos concluyendo que la matriz de covarianza poblacional de los datos no tiene esa estructura. Como se mencionó anteriormente, esto era de esperarse ya que se sabe que existe correlación entre los genes de un mismo individuo.

3.3.5. Datos NCI60

Los datos contienen niveles de expresión en 6830 genes de 64 líneas celulares de cáncer. Estos datos pueden obtenerse del Software *R-project* usando la librería *ISLR*. El valor de los estadísticos de prueba se muestran en las siguientes tablas.

Pruebas	Valor del estadístico
LW	188.7283
TW	676.3819
T_{2s}	2491.531
T_2	2459.416
cai	514277.4

Cuadro 3.18: Pruebas para el juego de hipótesis (2.2)

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = I_p$.

Pruebas	Valor del estadístico
J	315.2971
QLRT	2988.593
T_{1s}	6417.741
T_1	6334.869
Zou	176.9723

Para cada una de las pruebas el p-valor fue aproximadamente 0 y por tanto hay evidencia estadística para rechazar $H_0 : \Sigma = \lambda I_p$.

Al rechazar H_0 estamos concluyendo que la matriz de covarianza poblacional de los datos no tiene esa estructura. Como se mencionó anteriormente, esto era de esperarse ya que se sabe que existe correlación entre los genes de un mismo individuo.

Conclusiones

En este trabajo se estudiaron pruebas de hipótesis para la matriz de covarianza poblacional de datos normales multivariados, siendo de mayor interés el caso cuando la dimensión de los datos es más grande que el tamaño de la muestra. Se consideraron pruebas para las hipótesis $H_0 : \Sigma = \lambda I_p$ y $H_0 : \Sigma = I_p$.

Las simulaciones llevadas a cabo en este trabajo para verificar el comportamiento de las pruebas para $H_0 : \Sigma = I_p$, nos indican que para el caso clásico ($p < n$) y el caso de simension alta ($p \geq n$) las pruebas TW, CLRT, LW, Cai, T_{2s} y T_2 son aceptables en términos del tamaño, ya que las proporciones de veces en que se rechazó la hipótesis nula fueron cercanas al nivel de significancia considerado; mientras que para la prueba LRT_1, en el caso en que $p < n$, las proporciones de rechazo resultaron estar muy alejadas del nivel de significancia deseado cuando p es cercana a n y por tanto la prueba CLRT resulta ser una mejor alternativa que la prueba LRT_1. En términos de la potencia, la prueba TW mostró superioridad con respecto a las otras pruebas. A pesar de que las potencias de las pruebas CLRT, LW, Cai, T_{2s} y T_2 se encuentran por debajo de la TW, éstas no dejan de ser buenas alternativas para realizar pruebas de hipótesis. Aunque en algunos casos la prueba LRT_1 mostró mejor potencia que la TW, ésta no es una buena prueba ya que en términos del tamaño tiene un mal comportamiento. Por lo tanto la TW resulta ser una muy buena alternativa para realizar pruebas de hipótesis, tanto para el caso $p < n$ como para el caso $p \geq n$.

Las simulaciones para las pruebas de $H_0 : \Sigma = \lambda I_p$, tanto en el caso clásico como en el de dimensión alta, nos indican que las pruebas J, T_{1s}, T_1 y Zou son aceptables en términos del tamaño, ya que las proporciones de veces en que se rechazó la hipótesis nula fueron cercanas al nivel de significancia. Sin embargo, las pruebas LRT_2 y QLRT, en algunos casos tienen proporciones de rachazo muy alejadas al nivel de significancia deseado. La prueba LRT_2 resulta ser una muy buena prueba siempre y cuando $n < p$ y n es lo suficiente grande con respecto a p, y la prueba QLRT tiene un buen resultado siempre y cuando $p \geq n$ y p es lo suficientemente grande con respecto a n. La potencia de las pruebas J, T_{1s}, T_1 y Zou tienen buen comportamiento tanto para el caso $p < n$ como para el caso $p \leq n$, por lo que estas pruebas son buenas alternativas. Las potencias de las pruebas LRT_1 y QLRT en algunos casos mostraron tener superioridad con respecto a las otras pruebas, sin embargo en términos del tamaño tuvieron malos

resultados, por lo que estas pruebas no son recomendables en esos casos.

Los resultados de las simulaciones mencionadas anteriormente, ayudan a tener una mejor idea sobre el comportamiento de diversas pruebas de hipótesis para la matriz de covarianza poblacional encontradas en la literatura, al hacer una comparación simultánea de ellas en términos del tamaño y la potencia.

Por otro lado, en este trabajo se mostró también como aplicar las pruebas a datos reales encontrados en la literatura. En los ejemplos de aplicación se consideraron cinco conjuntos de datos de microarreglos ADN y se observó que con todas las pruebas se rechazaron los dos tipos de hipótesis nula considerados, por lo que se concluye que la matriz de covarianza poblacional de los datos no es igual a la identidad ni a un múltiplo de ella. Cabe mencionar que estos resultados eran de esperarse, ya que se sabe que existe una alta correlación entre genes de un mismo individuo. Estos ejemplos muestran la utilidad del uso de estas pruebas de hipótesis para el estudio de microarreglos de ADN, lo cual es de interés debido a que muchas metodologías estadísticas para datos multivariados dependen fuertemente de la estructura de la matriz de covarianza poblacional de los datos, como por ejemplo análisis de componentes principales, clasificación y comparación de medias, por lo que es importante verificar la estructura de la matriz de covarianza poblacional mediante pruebas de hipótesis adecuadas.

Apéndice A

Detalles técnicos

En este apéndice se proporcionan algunos detalles técnicos de algunos temas abordados en la tesis.

A.1. Transformaciones de vectores aleatorio normales

En esta sección se presentan algunas transformaciones de datos normales multivariados que son útiles antes de aplicar algunas pruebas de hipótesis descritas en el Capítulo 2. Los resultados que se presentan fueron obtenidos de [1].

Lema A.1. *Si $C = (c_{\alpha\beta})$ es una matriz ortogonal, entonces*

$$\sum_{\alpha=1}^{N} x_\alpha x'_\alpha = \sum_{\alpha=1}^{N} y_\alpha y'_\alpha,$$

donde $y_\alpha = \sum_{\beta=1}^{N} c_{\alpha\beta} x_\beta$, $\alpha = 1, 2, ..., N$.

Demostración.

$$\begin{aligned}
\sum_{\alpha=1}^{N} y_\alpha y'_\alpha &= \sum_{\alpha}\sum_{\beta} c_{\alpha\beta} x_\beta \sum_{\gamma} c_{\alpha\gamma} x'_\gamma \\
&= \sum_{\beta,\gamma} \left(\sum_{\alpha} c_{\alpha\beta} c_{\alpha\gamma} \right) x_\beta x'_\gamma \\
&= \sum_{\beta,\gamma} \delta_{\beta\gamma} x_\beta x'_\gamma \\
&= \sum_{\beta} x_\beta x'_\beta,
\end{aligned}$$

donde $\delta_{\beta\gamma}$ es la delta de Kronecker. □

Lema A.2. *Sea A una matriz de $n \times p$, con $n > p$, tal que*

$$A'A = I_p,$$

entonces existe una matriz B de $n \times (n-p)$ tal que $(A\ B)$ es ortogonal.

Demostración. Sea A una matriz de rango p, entonces existe una matriz C de $n \times (n-p)$ tal que $(A\ C)$ es no singular. Tomar $D = C - AA'C$ y notar que $D'A = 0$. Usando la descomposición espectral de D tenemos

$$\begin{aligned} D'D &= O\Lambda O' \\ O'D'DO &= \Lambda \\ \Lambda^{-1/2}O'D'DO\Lambda^{-1/2} &= I. \end{aligned}$$

Tomar $E = \Lambda^{-1/2}O'$, por lo que E es una matriz de $(n-p) \times (n-p)$ y tenemos que $E'D'DE = I$, por lo tanto $B = DE$ es la matriz deseada. □

Sean $X_1, X_2, ..., X_N$ independientes, cada una con distribución $N(\mu, \Sigma)$. Entonces existe una matriz ortogonal $B = (b_{\alpha\beta})$ de $N \times N$ con la última fila de la forma

$$(1/\sqrt{N}, 1/\sqrt{N}, ..., 1/\sqrt{N}) \quad \textit{(por el Lema A.2).}$$

Sea $A = N\widehat{\Sigma}$, donde $\hat{\Sigma} = (1/N)\sum_{\alpha=1}^{N}(x_\alpha - \overline{x})(x_\alpha - \overline{x})'$ y sea $Z_\alpha = \sum_{\beta=1}^{N} b_{\alpha\beta}X_\beta$, entonces

$$\begin{aligned} Z_N &= \sum_{\beta=1}^{N} b_{N\beta}X_\beta = \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}}X_\beta \\ &= \sqrt{N}\overline{X}. \end{aligned}$$

Por el Lema A.1 tenemos

$$\begin{aligned} A &= \sum_{\alpha=1}^{N} X_\alpha X'_\alpha - N\overline{X}\overline{X}' \\ &= \sum_{\alpha=1}^{N} Z_\alpha Z'_\alpha - Z_N Z'_N \\ &= \sum_{\alpha=1}^{N-1} Z_\alpha Z'_\alpha. \end{aligned}$$

Ya que Z_N es independiente de $Z_1, Z_2, ..., Z_{N-1}$, el vector de medias $\overline{X}$ es independiente de A. Por tanto

$$\mathbb{E}Z_N = \sum_{\beta=1}^{N} b_{N\beta}\mathbb{E}X_\beta = \sum_{\beta=1}^{N} \frac{1}{\sqrt{N}}\mu = \sqrt{N}\mu,$$

por lo que Z_N tiene distribución $N(\sqrt{N}\mu, \Sigma)$ y $\overline{X} = (1/\sqrt{N})Z_N$ tiene distribución $N(\mu, (1/\sqrt{N})\Sigma)$. Notemos que

$$\begin{aligned}\mathbb{E}Z_\alpha &= \sum_{\beta=1}^{N} b_{\alpha\beta}\mathbb{E}X_\beta = \sum_{\beta=1}^{N} b_{\alpha\beta}\mu \\ &= \sum_{\beta=1}^{N} b_{\alpha\beta}b_{N\beta}\sqrt{N}\mu \\ &= 0.\end{aligned}$$

Debido a que la matriz B es ortogonal, se tiene que $Z_1, Z_2, \ldots, Z_{N-1}$ son independientes y tienen distribución $N(0, \Sigma)$.

Por lo comentado anteriormente tenemos el siguiente resultado.

Teorema A.1. *La media muestral $\overline{X}$ de una muestra de tamaño N con distribución $N(\mu, \Sigma)$ tiene distribución $N(\mu, (1/\sqrt{N})\Sigma)$ y es independiente de $\widehat{\Sigma}$, el estimador de máxima verosimilitud de Σ. $N\widehat{\Sigma}$ es de la forma $\sum_{\alpha=1}^{N-1} Z_\alpha Z'_\alpha$, donde Z_α tiene distribución $N(0, \Sigma)$, $\alpha = 1, 2, ..., N-1$, y $Z_1, Z_2, ..., Z_{N-1}$ son independientes.*

A.2. Demostración del Teorema 2.6

Demostración. Por el Teorema 2.5, haciendo un cambio de variable $x = 1+y-2\sqrt{y}\cos\theta$, donde $0 \leqslant \theta \leqslant \pi$, tenemos

$$\begin{aligned}m(g) &= \frac{g(a(y)) + g(b(y))}{4} - \frac{1}{2\pi}\int_{a(y)}^{b(y)} \frac{g(x)}{\sqrt{4y - (x-1-y)^2}}dx \\ &= \frac{(1-\sqrt{y})^2 - \log(1-\sqrt{y})^2 - 1 + (1+\sqrt{y})^2 - \log(1+\sqrt{y})^2 - 1}{4} \\ &\quad - \frac{1}{2\pi}\int_0^\pi \frac{y - 2\sqrt{y}\cos\theta - \log(1+y-2\sqrt{y})}{\sqrt{4y(1-\cos^2\theta)}} \cdot 2\sqrt{y}\text{sen}\theta d\theta \\ &= \frac{y - \log(1-\sqrt{y})}{2} - \frac{1}{2\pi}\int_0^\pi y - 2\sqrt{y}\cos\theta - log(1+y-2\sqrt{y}\cos\theta)d\theta.\end{aligned}$$

Notemos que $log(1+y-2\sqrt{y}\cos\theta) = \log|1-\sqrt{y}e^{i\theta}|^2$, donde $e^{i\theta} = \cos\theta + i\text{sen}\theta$. Además usando la propiedad de los complejos $|z|^2 = z\overline{z}$ tenemos

$$\begin{aligned}|1-\sqrt{y}e^{i\theta}|^2 &= (1-\sqrt{y}(\cos\theta + i\text{sen}\theta))(1-\sqrt{y}(\cos\theta - i\text{sen}\theta)) \\ &= 1 - 2\sqrt{y}\cos\theta + y.\end{aligned}$$

Por la paridad de la función coseno tenemos

$$\frac{y-\log(1-y)}{2} - \frac{1}{4\pi}\int_0^{2\pi}\left[y - 2\sqrt{y} - \log|1-\sqrt{y}e^{i\theta}|^2\right]d\theta,$$

Usando las propiedades de la integral y resolviendo tenemos que

$$\frac{y-\log(1-y)}{2} - \frac{1}{4\pi}\int_0^{2\pi}\left[y - 2\sqrt{y} - \log|1-\sqrt{y}e^{i\theta}|^2\right]d\theta = \frac{-\log(1-y)}{2},$$

donde $\displaystyle\int_0^{2\pi}\log|1-\sqrt{y}e^{i\theta}|^2 d\theta = 0$, (ver [3]).

Para (2.7) consideraremos F^{y_n} la cual es la ley de Marchenko-Pastur de índice y_n. Usando el cambio de variables $x = 1 + y_n - 2\sqrt{y_n}\cos\theta$, donde $0 \leqslant \theta\pi$, tenemos

$$\begin{aligned}F^{y_n}(g) &= \int_{a(y_n)}^{b(y_n)} \frac{x-\log x - 1}{2\pi x y_n}\sqrt{(b(y_n)-x)(x-a(y_n))}dx \\ &= \frac{1}{2\pi y_n}\int_0^{\pi}\frac{y_n - 2\sqrt{y_n}\cos\theta - \log(1+y_n-2\sqrt{y_n}\cos\theta)}{1+y_n-2\sqrt{y_n}\cos\theta}\cdot 4y_n\text{sen}^2\theta d\theta \\ &= \frac{1}{2\pi y_n}\int_0^{pi}\left[1 - \frac{\log(1+y_n-2\sqrt{y_n}\cos\theta)+1}{1+y_n-2\sqrt{y_n}\cos\theta}\right]4y_n\text{sen}^2\theta d\theta.\end{aligned}$$

Notemos que $log(1+y_n-2\sqrt{y_n}\cos\theta) = \log|1-\sqrt{y_n}e^{i\theta}|^2$ y por la paridad de la función dentro de la integral tenemos que

$$\frac{1}{2\pi}\int_0^{2\pi}\left[2\text{sen}^2\theta - \frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}(\log|1-\sqrt{y_n}e^{i\theta}|^2 - 1)d\theta\right].$$

Usando propiedades de la integral tenemos que esta última integral es igual a

$$\begin{aligned}\frac{1}{2\pi}\Bigg[&\int_0^{2\pi}2\text{sen}^2\theta d\theta - \int_0^{2\pi}\frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}\log|1-\sqrt{y_n}e^{i\theta}|^2 d\theta \\ &- \int_0^{2\pi}\frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}d\theta\Bigg].\end{aligned}$$

Resolviendo las tres integrales, tenemos que $\frac{1}{2\pi}\int_0^{2\pi} 2\text{sen}^2\theta d\theta = 1$,

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}\log|1-\sqrt{y_n}e^{i\theta}|^2 d\theta$$
$$= \frac{y_n-1}{y_n}\log(1-y_n)-1,$$

la cual es calculada en [3], y por último tenemos

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}d\theta = \frac{1}{\pi}\int_0^{2\pi} \frac{\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}d\theta.$$

Haciendo $a = 1+y_n$, $b = 2\sqrt{y_n}$ y usando la sustitución de Weierstrass con $t = \tan(\theta/2)$ tenemos que

$$\text{sen}\theta = \frac{2t}{1+t^2}, \quad \cos\theta = \frac{1-t^2}{1+t^2}, \quad d\theta = \frac{2}{1+t^2}dt,$$

$$\frac{1}{\pi}\int_0^{2\pi} \frac{\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}d\theta = \frac{1}{\pi}\left[\int_0^{\pi} \frac{\text{sen}^2\theta}{a-b\cos\theta}d\theta + \int_{\pi}^{2\pi} \frac{\text{sen}^2\theta}{a-b\cos\theta}d\theta\right].$$

Como $\tan(0/\pi) = 0$, $\tan(\pi/2) = \infty$ y $\tan(2\pi/2) = 0$, tenemos que

$$\frac{1}{\pi}\left[\int_0^{\infty} \frac{\left(\frac{2t}{t^2+1}\right)^2}{a-b\left(\frac{1-t^2}{1+t^2}\right)}\cdot\left(\frac{2}{t^2+1}\right)dt + \int_{-\infty}^{0} \frac{\left(\frac{2t}{t^2+1}\right)^2}{a-b\left(\frac{1-t^2}{1+t^2}\right)}\cdot\left(\frac{2}{t^2+1}\right)dt\right]$$
$$= \frac{8}{\pi(a+b)}\int_{-\infty}^{\infty} \frac{t^2}{(t^2+1)^2\left(t^2+\frac{a-b}{a+b}\right)}dt.$$

Notemos que $\frac{a-b}{a+b} = \left(\frac{1-\sqrt{y_n}}{1+\sqrt{y_n}}\right)^2 > 0$, entonces $\frac{a-b}{a+b} = c^2$, así tenemos que

$$\frac{8}{\pi(a+b)}\int_{-\infty}^{\infty} \frac{t^2}{(t^2+1)^2(t^2+c^2)}dt,$$

del método de fracciones parciales se sigue

$$\frac{t^2}{(t^2+1)^2(t^2+c^2)} = \frac{At+B}{(t^2+1)^2} + \frac{\widetilde{C}t+D}{t^2+1} + \frac{Et+F}{t^2+c^2},$$

cuya solución es $A = \widetilde{C} = E = 0$, $B = \dfrac{1}{1-c^2}$, $F = -\left(\dfrac{c}{1-c^2}\right)^2$, $D = \left(\dfrac{c}{1-c^2}\right)^2$, por la tanto la integral es

$$\begin{aligned}
&\frac{8}{\pi(a+b)}\int_{-\infty}^{\infty} \frac{t^2}{(t^2+1)^2(t^2+c^2)}dt \\
&\quad= \frac{16}{\pi(a+b)}\int_0^{\infty} \frac{1}{1-c^2}\frac{1}{(t^2+1)^2} + \left(\frac{c}{1-c^2}\right)^2\frac{1}{t^2+1} - \left(\frac{c}{1-c^2}\right)^2\frac{1}{t^2+c^2}dt \\
&\quad= \frac{16}{\pi(a+b)}\left(\frac{\pi}{4(1-c^2)} + \left(\frac{c}{1-c^2}\right)^2\frac{\pi}{2} - \frac{c\pi}{2(1-c^2)^2}\right) \\
&\quad= \frac{8}{(a+b)}\left(\frac{1}{2(1+c)^2}\right) \\
&\quad= \frac{4}{(a+b)(1+c)^2}.
\end{aligned}$$

Notemos que como $a = 1+y_n$, $b = 2\sqrt{y_n}$, $c^2 = \left(\dfrac{1-\sqrt{y_n}}{1+\sqrt{y_n}}\right)^2$ y $c = \dfrac{1-\sqrt{y_n}}{1+\sqrt{y_n}}$, entonces

$$\begin{aligned}
(a+b)(1+k)^2 &= 1+y_n+2\sqrt{y_n}\left(1+\frac{1-\sqrt{y_n}}{1+\sqrt{y_n}}\right)^2 \\
&= (1+\sqrt{y_n})^2\left(\frac{2}{1+\sqrt{y_n}}\right)^2 \\
&= (1+\sqrt{y_n})^2\frac{4}{(1+\sqrt{y_n})^2} \\
&= 4,
\end{aligned}$$

de donde

$$\frac{4}{(a+b)(1+c)^2} = 1.$$

Por lo tanto

$$\begin{aligned}
&\frac{1}{2\pi}\left[\int_0^{2\pi} 2\text{sen}^2\theta d\theta - \int_0^{2\pi} \frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta}\log|1-\sqrt{y_n}e^{i\theta}|^2 d\theta\right.\\
&\left.-\int_0^{2\pi} \frac{2\text{sen}^2\theta}{1+y_n-2\sqrt{y_n}\cos\theta} d\theta\right]\\
&= 1-\frac{y_n-1}{y_n}\log(1-y_n)+1-1\\
&= 1-\frac{y_n-1}{y_n}\log(1-y_n).
\end{aligned}$$

Debido a la complejidad para resolver (2.6), se puede consultar en [3]. □

Bibliografía

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition, 1984.

[2] Z. Bai, D. Jiang, J.-F. Yao, and S. Zheng. Corrections to lrt on large-dimensional covariance matrix by rmt. *The Annals of Statistics*, 37(6B):3822–3840, 2009.

[3] Z. D. Bai and J. W. Silverstein. Clt for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):281–333, 2008.

[4] E. Bair and R. Tibshirani. Semi-supervised methods to predict patient survival from gene expression data. *PLoS biology*, 2(4):e108, 2004.

[5] T. T. Cai and Z. Ma. Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19(5B):2359–2388, 2013.

[6] G. Casella and R. L. Berger. *Statistical inference*. Duxbury Pacific Grove, CA, 2002.

[7] D. Cortez-Elizalde. Teoría de matrices aleatorias en el estudio de la matriz de covarianza poblacional de datos de dimensión alta. Tesis de licenciatura presentada en la UJAT, 2017.

[8] T. J. Fisher, X. Sun, and C. M. Gallagher. A new test for sphericity of the covariance matrix for high dimensional data. *Journal of Multivariate Analysis*, 101(10):2554–2570, 2010.

[9] I. M. Johnstone. On the Distribution of the Largest Eigenvalue in Principal Components Analysis. *The Annals of Statistics*, 29(2):295–327, 2001.

[10] O. Ledoit and M. Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30(4):1081–1102, 2002.

[11] Z. Li and J. Yao. Testing the sphericity of a covariance matrix when the dimension is much larger than the sample size. *Electronic Journal of Statistics*, 10(2):2973–3010, 2016.

[12] Z. Ma. Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli*, 18(1):322–359, 2012.

[13] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc., Hoboken, New Jersey, second edition, 2005.

[14] A. Onatski, M. J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41(3):1204–1231, 2013.

[15] M. S. Srivastava. Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, 35(2):251–272, 2005.

[16] M. S. Srivastava, H. Yanagihara, and T. Kubokawa. Tests for covariance matrices in high dimension with less sample size. *Journal of Multivariate Analysis*, 130:289–309, 2014.

[17] C. Zou, L. Peng, L. Feng, and Z. Wang. Multivariate sign-based high-dimensional tests for sphericity. *Biometrika*, 101(1):229–236, 2013.