Evaluación de una versión modificada de la prueba Shapiro-Wilk Generalizada con estimación shrinkage de la matriz de covarianzas: caso de alta dimensión con muestras pequeñas

Elizabeth Ospina Santander

Melany Roldan Lizcano

Universidad del Valle sede Meléndez
Facultad de Ingeniería
Escuela de Estadística
Cali, Colombia
2022

Evaluación de una versión modificada de la prueba Shapiro-Wilk Generalizada con estimación shrinkage de la matriz de covarianzas: caso de alta dimensión con muestras pequeñas

Elizabeth Ospina Santander
Melany Roldan Lizcano

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico(a)

Director:
Ph.D Javier Olaya Ochoa

Codirectora:
Ph.D Luz Adriana Pereira Hoyos

Universidad del Valle sede Meléndez
Facultad de Ingeniería
Escuela de Estadística
Cali, Colombia
2022

Dedicatoria

A mis padres Elizabeth Santander, Leandro Ospina y abuelas Marlene Andrade, Romelia Rios por haberme formado como la persona que soy, por brindarme su amor y apoyo durante esta etapa.

Elizabeth Ospina Santander

A mis padres Mirian Lizcano, Jair Roldan y mis hermanos por su cariño y apoyo incondicional, durante todo este proceso.

A mi compañero Jarvi Rodríguez por su amor y valiosas discusiones acádemicas que contribuyeron al desarrollo de este trabajo.

Melany Roldan Lizcano.

Agradecimientos

Agradecemos al matemático Jarvi Rodríguez Enríquez por describir y desarrollar el algoritmo en Python para la generación de matrices de covarianzas con estructura deseada, el cual fue un insumo muy importante para conducir la simulación que permitió evaluar la prueba modificada y producir los resultados obtenidos.

Agradecemos a nuestros padres y demás familiares por brindarnos su acompañamiento y apoyo incondicional durante este proceso. A todos los profesores que han compartido con nosotras un poco de su conocimiento y además nos han formado de manera profesional, en especial a los profesores Javier Olaya Ochoa y Luz Adriana Pereira por brindarnos la oportunidad de trabajar a su lado, por su dedicación y orientación a lo largo de este proyecto. Finalmente, agradecemos a la Universidad del Valle por el todo el conocimiento adquirido durante estos últimos años.

Resumen

La hipótesis nula de la prueba Shapiro-Wilk Generalizada se define como $H_0 : \mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ es una muestra que proviene de una $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Para el cálculo del estadístico W^* de esta prueba se utilizan los estimadores de matriz de covarianzas $\mathbf{S}$ y la matriz de precisión $\mathbf{S^{-1}}$. Es bien conocido que la matriz de covarianzas muestral $\mathbf{S}$, en situaciones donde el número de variables es mayor o incluso igual al número de observaciones disponibles, genera malas estimaciones de la matriz de covarianza $\boldsymbol{\Sigma}$. Dado que $\mathbf{S^{-1}}$ hace parte del estadístico de la prueba Shapiro-Wilk Generalizada; es de esperarse que la potencia de la prueba se vea afectada.

Un estudio de Monte Carlo fue desarrollado para evidenciar el efecto sobre la potencia de la prueba. Se identificó que esta presenta problemas cuando se emplea en datos que no siguen una distribución normal multivariada. Por el contrario, dicha afectación no se evidencia cuando se emplea en datos que provienen de una distribución normal multivariada. Seguidamente, como una propuesta metodológica para enfrentar la pérdida de potencia de la prueba, se incorporó la estimación shrinkage $\mathbf{S}^*$ en la prueba de Shapiro-Wilk Generalizada y se analizó su desempeño vía simulación versus el desempeño de la prueba tradicional. La evaluación se realizó bajo distintos escenarios del tamaño de muestra, niveles de significancia y estimaciones shrinkage. Así, se concluye que la versión modificada de la prueba Shapiro-Wilk Generalizada tiene un mejor desempeño que la prueba tradicional, bajo $n \cong p$. Finalmente, se presenta una aplicación con datos reales, que consisten en mediciones de calidad del aire en la Ciudad de Cali, Colombia. La verificación del supuesto de normalidad multivariada en estos datos, toma importancia puesto que abre la posibilidad de la aplicación de novedosas propuestas metodológicas para el manejo de datos faltantes, que resultan muy frecuentes en este contexto.

Palabras clave: Shrinkage, Shapiro-Wilk Generalizada, matriz de covarianzas, mal condicionamiento, matriz de precisión, simulación, normal multivariada, calidad ambiental.

Abstract

The null hypothesis of the Generalized Shapiro-Wilk test is defined as $H_0 : \mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ is a sample that comes from a $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. For the calculation of the W^* statistic of this test, the estimators of the covariance matrix $\mathbf{S}$ and the precision matrix $\mathbf{S^{-1}}$ are used. It is well known that the sample covariance matrix $\mathbf{S}$, in situations where the number of variables is greater than or even equal to the number of available observations, generates poor estimates of the covariance matrix $\boldsymbol{\Sigma}$. Since $\mathbf{S^{-1}}$ is part of the statistic of the Generalized Shapiro-Wilk test; the power of the test is expected to be affected.

A Monte Carlo study was developed to evidence the effect on the power of the test. It was

identified that this presents problems when it is used in data that do not follow a multivariate normal distribution. On the contrary, this affectation is not evident when it is used in data that come from a multivariate normal distribution. Next, as a methodological proposal to deal with the loss of power of the test, the shrinkage estimate $\mathbf{S}^*$ was incorporated into the Generalized Shapiro-Wilk test and its performance was analyzed via simulation versus performance. of the traditional test. The evaluation was carried out under different scenarios of sample size, significance levels and shrinkage estimates. Thus, it is concluded that the modified version of the Generalized Shapiro-Wilk test has a better performance than the traditional test, under $n \cong p$. Finally, an application with real data is presented, consisting of air quality measurements in the City of Cali, Colombia. The verification of the multivariate Normality assumption in these data is important since it opens the possibility of applying new methodological proposals for handling missing data, which are very frequent in this context.

Keywords: Shrinkage, Generalized Shapiro-Wilk, covariance matrix, malconditioning, precision matrix, simulation, multivariate normal, environmental quality

Contenido

Lista de Figuras

Lista de Tablas

1 Introducción

En diversos problemas de investigación es de interés conocer o tener certeza de la distribución población de la cual proviene la muestra de datos, ya que de esta manera las inferencias realizadas a partir de dicha muestra permiten generar conclusiones más acertadas sobre los parámetros. Para ello se han desarrollado múltiples pruebas estadísticas. En particular, la generalización de la prueba Shapiro-Wilk propuesta por Villaseñor y González (2009) permite verificar si cada observación de la muestra $\mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ proviene de una población multinormalmente distribuida denotada por $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Sin embargo, se presentan situaciones donde el número de variables (dimensión de cada vector) es mayor o incluso igual al número de observaciones disponibles (cantidad de vectores), además para el cálculo del estadístico W^* de esta prueba se utilizan los estimadores de matriz de covarianzas $\mathbf{S}$ y la matriz de precisión $\mathbf{S^{-1}}$, que como lo mencionan Ledoit y Wolf (2004), la matriz de covarianzas muestral $\mathbf{S}$ bajo estas situaciones genera malas estimaciones de la matriz de covarianza $\boldsymbol{\Sigma}$ y como $\mathbf{S^{-1}}$ hace parte del estadístico de la prueba Shapiro-Wilk Generalizada; es de esperarse que la potencia de la prueba se vea afectada.

La metodología empleada se centró en desarrollar un estudio de simulación Monte Carlo comparativo para evaluar el desempeño de la prueba modificada y tradicional a través de diferentes escenarios de los parámetros, las distribuciones multivariadas, tamaños de muestra, niveles de significancia y estimaciones shrinkage óptimas para cada muestra. Los resultados obtenidos evidencian que la modificación propuesta para la prueba Shapiro-Wilk Generalizada presenta mejor rendimiento en términos de la potencia en todos los escenarios, bajo $n \cong p$ y en la medida en que $n >> p$ el desempeño es similar en ambas pruebas.

El presente trabajo tiene la siguiente estructura: en el capítulo 1 se encuentra el planteamiento del problema, la justificación, los objetivos de la investigación y los antecedentes. El capítulo 2 abarca el marco teórico; en la primera parte se encuentra la definición, propiedades de la distribución normal multivariada y estimación de los parámetros, la segunda parte hace referencia al condicionamiento de la matriz de covarianzas y en la tercera parte se encuentra las definiciones entorno a las pruebas de hipótesis. El capítulo 3 aborda la metodología propuesta para llevar a cabo la modificación y evaluación de la prueba modificada a través de los diferentes escenarios establecidos. El capítulo 4 presenta los resultados descriptivos y los obtenidos de la comparación de la prueba tradicional y modificada en términos de la potencia, además de una discusión. Finalmente, en el capítulo 5 se desarrollan las conclusiones y

recomendaciones de este trabajo.

1.1. Planteamiento del problema

Sean $\mathbf{Y_1}, ..., \mathbf{Y_n}$ vectores aleatorios independientes e idénticamente distribuidos (i.i.d) en $\mathbb{R}^p$, $p \geq 1$. Se dice que cada uno de estos vectores son independientes y siguen una distribución normal p-variada denotada por $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$. Los n vectores aleatorios corresponden a las observaciones de un fenómeno en el que se miden p variables, en muchas situaciones es importante verificar si cada vector sigue una distribución normal p-variada, para ello, existen diversas pruebas multivariadas, en particular, la prueba Shapiro Wilk Generalizada propuesta por Villaseñor y González (2009) donde la hipótesis nula se define como $H_0 : \mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ es una muestra que proviene de una población $\mathrm{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Para el cálculo del estadístico W^* de esta prueba se utiliza la matriz de covarianzas muestral $\mathbf{S}$ y la matriz de precisión muestral $\mathbf{S^{-1}}$, que corresponden a las estimaciones de la matriz de covarianzas $\boldsymbol{\Sigma}$ y la matriz de precisión $\boldsymbol{\Sigma}^{-1}$, respectivamente. Ledoit y Wolf (2004) mencionan que hay muchos problemas aplicados que requieren la estimación de ambas matrices y que están sujetas a tres posibles escenarios de la matriz de datos en los que se presentan una serie de ventajas y desventajas. El primero es cuando la dimensión de la matriz de datos p (número de variables) es finito y fijo, mientras que el tamaño de muestra n (número de observaciones) va al infinito, entonces $\mathbf{S}$ está bien condicionada (en el límite) y tiene propiedades de optimalidad deseables, como que es el estimador máximo verosímil para datos distribuidos multinormalmente. El segundo es cuando la dimensión de la matriz de datos es grande en comparación con el tamaño de la muestra disponible, entonces $\mathbf{S}$ ni siquiera es invertible. El tercero es cuando la dimensión de la matriz de datos es aproximadamente igual al tamaño de la muestra disponible, entonces $\mathbf{S}$ es invertible pero numéricamente está mal condicionada (lo que significa que invertirla implica amplificar en gran medida el error de estimación) ya que los valores propios más grandes de $\mathbf{S}$ están severamente sesgados hacia arriba y los más pequeños hacia abajo.

La matriz de covarianzas muestral $\mathbf{S}$ es una mala aproximación en muchas situaciones donde la dimensión es igual o incluso mayor que el número de observaciones disponibles. Ledoit y Wolf (2003) dan un ejemplo de la situación: "Doscientas observaciones pueden parecer muchas, pero no es suficiente si hay 100 variables: sería tan malo como usar dos observaciones para estimar la varianza de 1 variable aleatoria" (p.378). La anterior conclusión implicaría que, cuando $n \cong p$ las estimaciones $\mathbf{S}$ y $\mathbf{S^{-1}}$ no son óptimas y dado que $\mathbf{S^{-1}}$ conforma parte del estadístico de la prueba Shapiro-Wilk Generalizada; entonces, un supuesto de este estudio es que la potencia de la prueba puede verse afectada.

Ledoit y Wolf (2004) prueban que de forma general $\mathbf{S}$ tiene los valores propios más dispersos que los de la matriz de covarianzas poblacional, esto implica que su condicionamiento

es superior al de la matriz de covarianzas poblacional, por lo que propusieron un estimador shrinkage $\mathbf{U}^*$ que regula dicho condicionamiento para una mejor estimación de la matriz de covarianzas. Posteriormente, la contribución de Schäfer y Strimmer (2005) fue proponer un estimador shrinkage óptimo denotado por $\mathbf{S}^*$ que explota el lema de Ledoit y Wolf (2003).

Según Villaseñor y González (2009), la prueba Shapiro Wilk Generalizada es más potente comparada con las demás pruebas existentes. Porras (2016); concluye, al evaluar cuatro pruebas de normalidad multivariada (Mardia, Henze-Zinkler, Shapiro-Wilk Generalizada y Royston) que: "no presentan diferencias significativas en su potencia de prueba" (p.145). Aunque las diferencias no fueron significativas, se observó que la prueba Shapiro-Wilk Generalizada tuvo un comportamiento similar en todos los escenarios de simulación. Cabe resaltar que, el autor analizó diferentes escenarios que involucraron criterios como el tamaño de muestra, el número de variables y la variabilidad generalizada (definida como el determinante de la matriz de covarianzas que sirve como medida global de la independencia entre las variables). También para la generación de la matriz de covarianzas asumió independencia entre las variables y sugiere para estudios posteriores se evalúe la potencia de las pruebas considerando la correlación entre las variables. Con base en lo anterior, este estudio selecciona la prueba de Shapiro-Wilk Generalizada porque su potencia no parece ser muy sensible a cambios en las condiciones mencionadas.

Este trabajo propone incorporar la estimación shrinkage $\mathbf{S}^*$ en la prueba de Shapiro-Wilk Generalizada para analizar su desempeño vía simulación. En particular, se probará la potencia de la prueba modificada (reemplazando $\mathbf{S}^*$ por $\mathbf{S}$) en contraste con la potencia de la prueba tradicional; bajo el escenario en que $n \cong p$.

El uso de la modificación propuesta se ilustrará con datos de $PM_{2.5}$ proporcionados por el Sistema de Vigilancia de la Calidad del Aire de Santiago de Cali (SVCASC), los cuales tienen un alto número de registros faltantes que pueden suceder por fallas en el suministro de energía eléctrica, fallas en los sistemas de comunicación o situaciones relacionadas con el clima; también su relevancia político-ambiental radica en que según la Organización Mundial de la Salud - OMS (2018) las partículas suspendidas en el aire que son más dañinas para la salud son las partículas finas ($PM_{2.5}$). Dichas partículas pueden atravesar la barrera pulmonar y entrar en el sistema sanguíneo, en consecuencia, cuando se está expuesto a concentraciones altas de $PM_{2.5}$ se tiene mucho riesgo de desarrollar enfermedades cardiovasculares y respiratorias, así como cáncer de pulmón.

Por su parte, Caicedo y Jimenez (2016) con datos de $PM_{2.5}$ correspondientes al año 2015 anticipan que la distribución del $PM_{2.5}$ de la hora i en el día j es normal donde la media y varianza de la hora i en el día j se encuentran usando la media funcional y la varianza funcional del día j. Seguidamente, Otero y Presiga (2019) con datos del año 2017, validan

el mismo resultado. Así que, si se prueba que la distribución conjunta de las 24 horas de día j es normal multivariada, entonces los datos se podrían modelar con un Proceso Gaussiano (PG). Sin embargo, en las propuestas de los trabajos mencionados no se ha utilizado las correlaciones entre horas y por lo tanto, no se ha considerado la correlación entre las 24 horas, por tal motivo, Villareal y Arroyave (2020) presentaron otra propuesta de imputación con datos horarios correspondientes al año 2018 y se encontraron la dificultad de que la matriz de covarianzas muestral es un mal estimador de la matriz de covarianza poblacional debido a que $n \cong p$, puesto que en el periodo de un año, máximo se puede medir el $PM_{2.5}$ en 53 veces del día j durante las 24 horas de cada vez ($n \leq 53$ y $p = 24$); como solución proponen hacer una estimación shrinkage $\mathbf{S}^*$, la cual transforma $\mathbf{S}$ de tal manera que estime mejor a $\mathbf{\Sigma}$ cuando la muestra es pequeña. Los autores realizaron la imputación con ambas estimaciones ($\mathbf{S}$ y $\mathbf{S}^*$) de la matriz de covarianzas y concluyeron que con la matriz $\mathbf{S}^*$ se generaron imputaciones más lógicas pero se perdió precisión en el método. Se resalta el hecho de que todas las pruebas en el software R para probar normalidad multivariada se basan en la estimación tradicional de la matriz de covarianzas, por lo que, al momento de probar normalidad multivariada con la prueba de Mardia no se realizó con $\mathbf{S}^*$ si no con $\mathbf{S}$.

Si la modificación de la prueba Shapiro-Wilk Generalizada propuesta funciona, se confiaría en las estimaciones y la precisión del método de imputación. Además, permitiría continuar con la investigación de imputación de datos faltantes de $PM_{2.5}$ usando un PG.

Por último, cabe resaltar que la modificación de la prueba Shapiro Wilk Generalizada seguiría siendo útil aún cuando los datos de $PM_{2.5}$ estén completos ($n \leq 53$ y $p = 24$) y que se podría extender a otros escenarios de la matriz de covarianzas.

1.2. Justificación

El número de observaciones es muy importante al momento de realizar un estudio, el investigador se enfrenta ante tres posibles situaciones dependiendo de la naturaleza del problema y las limitaciones: La primera es que debe elegir entre métodos menos costosos (más factibles) y más costosos (requiere muchos recursos); por ejemplo, someter a calor un trozo de carbón y medir distintas características físicas, alguna podría ser: la cantidad de masa que se pierde a medida en que la temperatura va incrementando; cada medición requiere mucho tiempo y dinero para ser tomada. La segunda es cuando los datos ya fueron medidos con fallas en el instrumento de medición u omisión en los registros, y desea realizar un análisis histórico para predicciones futuras; por ejemplo, datos de $PM_{2.5}$ proporcionados por el Sistema de Vigilancia de la Calidad del Aire de Santiago de Cali (SVCASC), recordando que son partículas finas que están suspendidas en el aire y que son las más dañinas para la salud. La tercera es cuando la definición del problema hace que $n \cong p$; por ejemplo, medir diez características a las diez personas que padecen la enfermedad más rara del mundo. Así pues, es

posible e incluso común que la matriz de covarianzas muestral no sea un buen estimador ya que la cantidad de variables puede ser aproximadamente igual a la cantidad de observaciones.

Al aplicar métodos estadísticos con la expectativa de generar buenas estimaciones, el investigador debería analizar la distribución de los datos. Teóricamente, las pruebas de hipótesis formales que determinan normalidad multivariada, tales como Mardia (1970), Henze y Zirkler (1990) y Shapiro y Wilk (1965), tienen en común que parte de sus estadísticos están conformados por la matriz de covarianzas máximo verosímil, suponiendo que la cantidad de variables p es fija y la cantidad de observaciones n tiende a infinito ($n >> p$).

Este estudio parte de la presunción de que las pruebas formales no funcionan adecuadamente en el caso de que la matriz de covarianzas poblacional no esté bien representada por la matriz de covarianza muestral. En consecuencia, no subsanar esta dificultad implicaría que las estimaciones proporcionadas por el investigador no sean plenamente confiables.

Para el análisis de la distribución de los datos cuando $n \cong p$, este estudio utiliza una alternativa novedosa que consiste en incorporar la estimación shrinkage de la matriz de covarianzas en la prueba Shapiro-Wilk Generalizada. Así, se estaría considerando en el estadístico una mejor estimación de la matriz de covarianzas poblacional.

De David y MacKay (2003) se toma que, "un proceso estocástico continuo en el tiempo $\{X_t; t \in T\}$ es Gaussiano si y solo si para cualquier conjunto finito de índices $t_1, ..., t_c$ en el conjunto de índices T, entonces $X_{t_1,...,t_c} = (X_{t_1}, ..., X_{t_c})$ es una variable aleatoria normal multivariada" (p.540). De acuerdo con la definición anterior, si el investigador prueba con certeza que la distribución de los datos es normal p-variada; entonces, podría utilizar la teoría desarrollada para un Proceso Gaussiano y así generar estimaciones confiables.

En el proyecto "Imputación de datos faltantes de $PM_{2.5}$ basada en un Proceso Gaussiano", el director Olaya (2019) menciona que: "Proyectos previos del Grupo de Investigación en Estadística Aplicada - INFERIR, permiten anticipar que se dispone de distribuciones normales univariadas en mediciones de $PM_{2.5}$ asociadas con la hora i ($i = 1, ..., 24$) del día j ($j = 1, ..., 7$). Una primera mirada al problema también permite prever la existencia de distribuciones normales bivariadas (entre las horas i e i') y multivariadas (entre todas las horas del día), por lo que se estudiarían inicialmente estas distribuciones, para continuar con la exploración del uso de un PG para imputar datos faltantes" (p.1). Si la modificación de la prueba Shapiro-Wilk Generalizada funciona bien evaluada con los datos de $PM_{2.5}$; entonces, se podría proceder con la imputación de datos faltantes usando un PG y se podría replicar en el mismo u otro contexto.

La investigación planteada contribuirá a generar hipótesis del funcionamiento y caracterísi-

cas de una versión modificada de la prueba Shapiro-Wilk Generalizada bajo distintos escenarios de simulación, para futuros investigadores que contribuirían con las demostraciones formales. Asimismo, se abordarían los problemas en situaciones donde la estimación de la matriz de covarianzas reduce su calidad debido a que $n \cong p$. Además, los resultados de este estudio minimizarán la incertidumbre en la generación de mecanismos para estimaciones efectivas.

1.3. Objetivo General

Estudiar vía simulación el desempeño de una versión modificada de la prueba Shapiro-Wilk Generalizada basada en la estimación shrinkage de la matriz de covarianzas.

1.3.1. Objetivos específicos

- Proponer una modificación de la prueba Shapiro-Wilk Generalizada basada en la estimación shrinkage de la matriz de covarianzas.
- Describir las características de la prueba Shapiro-Wilk Generalizada modificada.
- Evaluar la prueba Shapiro-Wilk Generalizada modificada en los escenarios de simulación determinados.
- Realizar una implementación práctica con los datos de $PM_{2.5}$ proporcionados por el Sistema de Vigilancia de la Calidad del Aire de Santiago de Cali (SVCASC).

1.4. Pregunta de investigación

¿Cómo es el desempeño de una versión modificada de la prueba Shapiro-Wilk Generalizada basada en la estimación shrinkage de la matriz de covarianzas?

1.5. Antecedentes

A continuación se presentan los hallazgos más relevantes identificados en una revisión bibliográfica reciente sobre los diferentes estudios que han abordado estimaciones de problemas donde la cantidad de variables es aproximadamente igual a la cantidad de observaciones, especialmente cuando se tiene como premisa probar que la muestra sigue una distribución normal multivariada; estudios que soportan la modificación propuesta de la prueba Shapiro Wilk Generalizada y sobre los diferentes métodos de análisis que se han empleado para subsanar las desventajas que conlleva estar incluso en el escenario en que $n \leq p$. En consecuencia, este capítulo se divide en tres secciones: revisión de antecedentes relacionados con la problemática

de verificar la multinormalidad con alta dimensión y tamaños de muestra pequeños; revisión de antecedentes que soportan el planteamiento de la propuesta de modificación a la prueba Shapiro Wilk Generalizada; revisión de antecedentes de otros métodos. Al interior de cada subsección, los trabajos se presentan en orden cronológico, desde la publicación más antigua a la más reciente.

1.5.1. Identificación del problema

Caicedo y Jimenez (2016) con datos de $PM_{2.5}$ tomados durante el año 2015 de la estación Universidad del Valle de Santiago de Cali, prueban que la distribución del $PM_{2.5}$ de la hora i del día j es normal. Para abordar las estimaciones proponen un método de imputación de datos basado en el análisis de datos funcionales, el cual consiste en generar un número aleatorio a partir de la distribución de la hora i del día j especificada por la media y varianza que se encuentran usando la media funcional y la varianza funcional del día j, así, donde existen datos faltantes toman el valor generado como el que se debe imputar. Seguidamente Otero y Presiga (2019) con datos de $PM_{2.5}$ tomados durante el año 2017 evalúan el método propuesto por Caicedo y Jimenez (2016), con la diferencia de que generan cinco números aleatorios y el dato imputado será el promedio; concluyen que sus resultados son promisorios y recomiendan que en próximos estudios se tenga en cuenta la correlación entre las 24 horas de cada día de la semana. Villareal y Arroyave (2020) (siguiendo la recomendación de Otero y Presiga (2019)) con datos de $PM_{2.5}$ tomados durante el año 2018 proponen otro método de imputación, sin embargo, al probar que cada día de la semana se puede modelar mediante una distribución normal 24-variada, los autores reportan la dificultad de que la matriz de covarianzas muestral es un estimador impreciso (fuerte sesgo en la estimación de valores propios que brindan información sobre la dispersión de los datos) de la matriz de covarianza poblacional debido a que el tamaño de muestra no es el adecuado considerando la cantidad de variables ($n \cong p$), en consecuencia, esto podría estar afectando la potencia de la prueba de Mardia ya que, la distribución del estadístico está definida cuando $n \to \infty$, por lo que pierden precisión en el método al generar estimaciones basadas en las propiedades de la distribución normal 24-variada.

Se resalta la relevancia de probar con certeza la distribución de la que provienen los datos para proponer mecanismos con el objetivo de generar estimaciones confiables.

A continuación algunas situaciones donde no se cumple que $n \to \infty$.

McMillan y Weitzner (2003) requieren informar la experiencia de un grupo de investigadores que han tenido un año de experiencia en un ensayo clínico con pacientes de cuidados paliativos domiciliarios. El equipo de investigación mantuvo registros cuidadosos del número de parejas (paciente/cuidador) acumuladas en el estudio y las razones de la no acumulación,

así como las razones de la deserción. Respecto a la muestra comentan que: "Durante nueve meses el hospital admitió a 2.517 pacientes; el 75 % tenía cáncer y el 95 % tenía cuidadores, lo que los hacía elegibles para el estudio. Sin embargo, después de una evaluación adicional, solo el 19 % fue elegible para el contacto y solo el 5 % finalmente se incorporó al estudio. Para las 125 parejas (paciente/cuidador) que realmente participaron en el estudio, los datos de referencia se obtuvieron solo en el 50 % y los datos de seguimiento evaluables solo en el 50 %. Concluyen que incorporar pacientes a los ensayos clínicos y retenerlos cuando están críticamente enfermos y próximos a la muerte son tareas extraordinariamente difíciles. La incapacidad de reclutar y retener sujetos para ensayos clínicos tiene implicaciones para la integridad de los datos, el análisis de datos, el éxito del proyecto y el costo de llevar a cabo dichos proyectos en el futuro" (p.1).

"Los datos de 19 pacientes depresivos adquiridos al comienzo y al final de una terapia de seis semanas. Las nueve variables representan los cambios de la potencia absoluta del electroencefalograma (EEG) durante la terapia en nueve canales seleccionados (indicados por $\mathbf{X_1}$ a $\mathbf{X_9}$ respectivamente). Los datos completos se pueden encontrar en Läuter, Glimm, y Kropf (1996). Queremos probar la hipótesis $H_0 : \{\mathbf{X_1}, ..., \mathbf{X_9}\}$, es una muestra que proviene de una distribución normal $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con $p = 9$ y $n = 19$. Esto se puede considerar como un ejemplo de prueba de normalidad de alta dimensión con un tamaño de muestra pequeño." (Liang, Tang, y Chan (2009), p.3889).

De las situaciones anteriores se evidencia que no necesariamente una mala estimación de la matriz de covarianzas se da por datos faltantes, también en otros contextos sucede que $n \cong p$ y no es posible o es muy difícil tener control sobre eso.

1.5.2. Soportes del planteamiento de la prueba modificada

En la sección anterior se hace mención sobre las diversas situaciones de estudio en donde se evidencian dificultades para obtener una buena estimación de la matriz de covarianzas poblacional y cómo esto genera dificultades para aplicar los métodos estadísticos con el fin de obtener resultados confiables, es por ello que surge la necesidad de subsanar este problema.

Ledoit y Wolf (2003) sugieren estimar la matriz de covarianzas muestral mediante una transformación llamada shrinkage puesto que cuando el número de variables es considerablemente grande en relación con el número de observaciones, la matriz de covarianzas muestral $\mathbf{S}$ se estima con mucho error. La estimación shrinkage está dada por la combinación lineal $\lambda\mathbf{T} + (1 - \lambda)\mathbf{S}$, donde $\mathbf{T}$ es la matriz objetivo (estructurada), $\mathbf{S}$ es la matriz de covarianzas muestral y λ es un número entre 0 y 1 (llamado parámetro de contracción). Schäfer y Strimmer (2005) proponen un estimador de covarianzas shrinkage novedoso que explota el lema de Ledoit y Wolf (2003) para el cálculo analítico de la intensidad de contracción óptima.

Ledoit y Wolf (2003) enfatizan en que el óptimo denotado por λ^* del parámetro λ (llamado parámetro de contracción) de la estimación shrinkage debe ser consistente. Sin embargo, Schäfer y Strimmer (2005) argumentan que este es un requisito muy débil porque la consistencia es una propiedad asintótica y un requisito básico de cualquier estimador sensible. Dado que el interés es la inferencia de muestras pequeñas, para la estimación de λ^* proponen reemplazar todas las esperanzas, varianzas y covarianzas de λ^* por sus contra partes insesgadas y este nuevo estimador es llamado $\hat{\lambda}^*$. Finalmente, redefinen al estimador de contración óptimo $\hat{\lambda}^*$ por uno que además de ser óptimo también esté regulado y este se denota por $\hat{\lambda}^M$=máx$(0, min(1, \hat{\lambda}^*))$ que también estima al parámetro de contracción λ pero asegurándose de que es el óptimo y que evita la contracción excesiva o negativa. En este trabajo se denota al estimador de la intensidad de contracción a utilizar con el superíndice M para hacer referencia a que es el que se utilizará en la prueba "Modificada".

Schäfer y Strimmer (2005) garantizan que calculando la estimación shrinkage de la matriz de covarianza usando su propuesta de intensidad de contracción $\hat{\lambda}^M$ y una matriz $\mathbf{T}$ específica, esta tiene un error cuadrático medio mínimo y siempre es una matriz positiva definida incluso para tamaños de muestra pequeño. Sin embargo, por la forma en que se define a la matriz objetivo $\mathbf{T}$, la estimación shrinkage resultante solo contrae a la covarianzas y no a las varianzas, por lo que una extensión fue dada por Opgen-Rhein y Strimmer (2007) donde se propone una contracción óptima para las varianzas. Por último, se llega a la conclusión de que por un lado, contraer las correlaciones usando la misma idea de Schäfer y Strimmer (2005) tiene una serie de ventajas y por otro lado, se debe hacer uso de la contracción óptima para las varianzas para finalmente combinar ambas contracciones y reconstruir la estimación shrinkage de la matriz de covarianzas.

Los detalles del origen del planteamiento del estimador shrinkage de la matriz de covarianzas se exponen en la subsección 2.2.1 del marco teórico y los detalles de la construcción del estimador shrinkage de la matriz de covarianzas a partir de la contracción de las correlaciones y las varianzas se expone en las subsecciones 3.2.3, 3.2.4 y 3.2.5 de la metodología. La discusión anterior sugiere hacer uso de la estimación shrinkage de la matriz de covarianzas construida con las propuestas de Schäfer y Strimmer (2005) y Opgen-Rhein y Strimmer (2007)

A lo largo de los años, se ha presentado que en diversos estudios surge la necesidad de probar que la distribución de un conjunto de datos proviene de una distribución normal multivariada; esto con la finalidad de generar estimaciones basadas en sus propiedades y tener certeza en los resultados. Diversos autores como Mardia (1970); Henze y Zirkler (1990); Shapiro y Wilk (1965); y Royston (1992) han propuesto estadísticos para evaluar la normalidad multivariada. (Ver Porras (2016))

Srivastava y Hui (1987) mencionan que al evaluar la normalidad multivariada, diversos autores han propuesto procedimientos de pruebas, los cuales la mayoría de ellos están basados en generalizaciones de los estadísticos de prueba univariados. Estos autores afirman que el estadístico W de Shapiro-Wilk es la mejor prueba para detectar desviaciones de la normalidad univariada y proponen dos estadísticos de prueba basados en componentes principales, los cuales se pueden considerar como una generalización del estadístico de Shapiro-Wilk, haciendo énfasis en que los cálculos de dichos estadísticos son simples y directos. Villaseñor y González (2009) proponen una generalización de la prueba de Shapiro-Wilk para la normalidad multivariada, basada en el estadístico univariado y una estandarización empírica de las observaciones. Además, muestran a través de simulación de Monte Carlo que la generalización propuesta es más potente en comparación con las pruebas de Mardia (1970), Henze y Zirkler (1990) y Srivastava y Hui (1987), frente a una amplia gama de escenarios considerados.

Porras (2016) evaluó la potencia de cuatro pruebas de normalidad multivariada (Mardia, Henze-Zinkler, Shapiro-Wilk Generalizada y Royston) utilizando dieciséis escenarios que involucraron criterios como: tamaño de muestra, número de variables (independientes), variabilidad generalizada (baja y alta). Por un lado, el autor ha considerado en la simulación pocas variables ($p = 3$) para un nivel bajo y alto de variabilidad generalizada, de aquí concluye que no existen diferencias significativas entre las pruebas de normalidad multivariada analizadas para los distintos tamaños de muestra (30, 100, 500, 1000). Por otro lado, el autor ha considerado más variables ($p = 7$) para un nivel bajo y alto de variabilidad generalizada, de aquí concluye que a medida que el tamaño de muestra se incrementa, la potencia de la prueba de Mardia disminuye sutilmente; mientras que en la prueba de Royston incrementa de manera sutil. De acuerdo con lo anterior y analizando los resultados de Porras (2016), este estudio presume que la prueba de Shapiro-Wilk Generalizada no parece ser muy sensible a cambios en las condiciones mencionadas.

Lo anterior sugiere hacer uso de la prueba Shapiro-Wilk Generalizada, puesto que comparada con las otras pruebas que evalúan la normalidad multivariada, esta resulta se más potente según Villaseñor y González (2009) y su rendimiento es más estable viendo los resultados presentados por Porras (2016).

1.5.3. Método alternativo de análisis

Liang y cols. (2009) proponen una familia de estadísticos generalizados de Shapiro-Wilk para probar la normalidad multivariada utilizando la idea de componentes principales, tal como Srivastava y Hui (1987), la cual resulta atractiva porque es un método común para reducción de dimensión en análisis de datos de alta dimensión. Los estadísticos propuestos van dirigidos

a casos donde el tamaño de la muestra es más pequeño que la cantidad de variables. Con estudios de Monte Carlo evalúan el rendimiento en términos de potencia y tasa baja de error tipo I con algunos datos no multinormales. Se concluye que los estadísticos propuestos son superiores a los estadísticos generalizados existentes y muestran beneficios competitivos al probar la normalidad multivariada con tamaño de muestra pequeño.

Dado que Liang y cols. (2009) abordan la misma problemática de este trabajo pero desde un enfoque de reducción dimensional, este trabajo de grado usa como base algunas de sus propuestas para los escenarios de evaluación de la prueba modificada.

En la literatura encontrada no se identificaron alternativas para abordar una problemática tan común ($n \cong p$). Por lo tanto, incorporar la estimación shrinkage de la matriz de covarianzas en la prueba Shapiro-Wilk Generalizada es una alternativa novedosa y además intuitivamente sencilla de entender.

2 Marco teórico

En esta sección se presenta lo concerniente a los temas o definiciones estadísticas, las cuales conforman una base de la teoría y herramientas con las cuales se buscaría resolver el problema planteado. La presentación de los temas se hará en el orden que requiere el desarrollo del trabajo de grado.
En la notación vectorial y matricial se utilizan letras mayúsculas y en negrita, para vectores y matrices observadas se utilizan letras minúsculas con negrita, para notación de escalares se utilizan letras mayúsculas sin negrita y para valores escalares observados se utilizan letras minúsculas sin negrita; a menos que se indique lo contrario, en cuyo caso se hará claridad de la naturaleza de lo definido.

2.1. Distribución normal multivariada

En esta subsección se presenta la definición y las propiedades principales de la distribución normal multivariada, que corresponde a una generalización de la distribución normal univariada para la extensión a casos de aplicación donde se cuenta con un grupo de variables que podrían estar correlacionadas entre sí; teniendo en cuenta que la relevancia práctica de esta distribución en los métodos multivariados se deriva principalmente del teorema del límite central multivariado, puesto que, proporciona una serie de propiedades deseables. Cabe resaltar que, la distribución normal multivariada se encuentra completamente definida por los dos primeros momentos muestrales que corresponden a un vector de medias y a la matriz covarianzas.

Definición 2.1.1 *Para $p \geq 2$, sea $\mathbf{Y}$ una variable aleatoria p-dimensional con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas positiva definida $\boldsymbol{\Sigma}$ que se pueden ver como:*

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \; ; \; \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

Así, $\mathbf{Y}$ sigue una distribución normal multivariada de la forma:

$$\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} > 0$$

La función de densidad de $\mathbf{Y}$ es:

$$f(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right]; \quad \mathbf{y}\in\mathbb{R}^p \tag{2-1}$$

Además, la matriz de covarianzas puede verse en términos de las correlaciones y las desviaciones estándar, así:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \rho_{11}\sigma_{11} & \rho_{12}\sigma_{11}^{1/2}\sigma_{22}^{1/2} & \dots & \rho_{1p}\sigma_{11}^{1/2}\sigma_{pp}^{1/2} \\ \rho_{21}\sigma_{22}^{1/2}\sigma_{11}^{1/2} & \rho_{22}\sigma_{22} & \dots & \rho_{2p}\sigma_{22}^{1/2}\sigma_{pp}^{1/2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1}\sigma_{pp}^{1/2}\sigma_{11}^{1/2} & \rho_{p2}\sigma_{pp}^{1/2}\sigma_{22}^{1/2} & \dots & \rho_{pp}\sigma_{pp} \end{bmatrix}$$

Propiedades 2.1.1

1. La distribución marginal de cualquier conjunto de componentes de una variable normal multivariada es también normal multivariada.

   Si $\mathbf{Y}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, entonces para cada partición $e<p$, las distribuciones de $\mathbf{Y_1}$ y $\mathbf{Y_2}$ son $N_e(\boldsymbol{\mu_1},\boldsymbol{\Sigma}_{11})$ y $N_{p-e}(\boldsymbol{\mu_2},\boldsymbol{\Sigma}_{22})$, respectivamente.

   Para un $e<p$ fijo, las particiones de $\mathbf{Y},\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ quedan de la siguiente manera:

   $$\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{Y_1}\\ \boldsymbol{Y_2}\end{bmatrix} \;;\; \boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu_1}\\ \boldsymbol{\mu_2}\end{bmatrix} \;;\; \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$$

   donde

   $$\boldsymbol{Y_1} = (Y_1,...,Y_e)', \quad \boldsymbol{Y_2} = (Y_{e+1},...,Y_p)'$$

   $$\boldsymbol{\mu_1} = (\mu_1,...,\mu_e)', \quad \boldsymbol{\mu_2} = (\mu_{e+1},...,\mu_p)'$$

   $\boldsymbol{\Sigma_{ii}}$ es la matriz de covarianzas de $\boldsymbol{Y_i}$, $(i=1,2)$ y $\boldsymbol{\Sigma_{12}} = (\sigma_{ij})$, tal que $\sigma_{ij} = cov(Y_i,Y_j)$ para $1\le i<j\le p$.

2. Sean las variables aleatorias $\mathbf{Y_1}$ y $\mathbf{Y_2}$ definidas en la propiedad anterior, donde $\mathbf{Y}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Entonces, $\mathbf{Y_1}$ y $\mathbf{Y_2}$ son independientes si y solo si la covarianza entre $\mathbf{Y_1}$ y $\mathbf{Y_2}$ es cero, en símbolos, $\boldsymbol{\Sigma}_{12}=0$.

3. Las distribuciones de transformaciones lineales o combinaciones lineales de variables normales multivariadas son normales multivariadas.

   Si $\mathbf{Y}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ y $\mathbf{Y}=\mathbf{CY}+\mathbf{b}$, donde $\mathbf{C}$ es cualquier matriz real de orden $(\eta\times p)$ y $\mathbf{b}$ es cualquier vector real de orden $(\eta\times 1)$, entonces $\mathbf{Y}\sim N_\eta(\mathbf{C}\boldsymbol{\mu}+\mathbf{b},\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$.

   Para $\eta<p$, la transformación

   $$\mathbf{Y}^* = \begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{C}\\ \mathbf{B}\end{bmatrix}\mathbf{Y} + \begin{bmatrix}\mathbf{b}\\ \mathbf{0}_{p-\eta}\end{bmatrix}$$

donde $\mathbf{B}$ es cualquier matriz de orden $((p-\eta)\times p)$. Por lo que

$$\begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{bmatrix} \sim N_p\left[\begin{bmatrix}\mathbf{C}\boldsymbol{\mu}+\mathbf{b}\\ \mathbf{B}\boldsymbol{\mu}\end{bmatrix}, \begin{bmatrix}\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}' & \mathbf{C}\boldsymbol{\Sigma}\mathbf{B}'\\ \mathbf{B}\boldsymbol{\Sigma}\mathbf{C}' & \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}'\end{bmatrix}\right]$$

Entonces $\mathbf{Y} = \mathbf{Y_1} = \mathbf{C}\mathbf{Y}+\mathbf{b} \sim N_\eta(\mathbf{C}\boldsymbol{\mu},\ \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$

4. La distribución condicional en una distribución normal multivariada es normal multivariada. Además, el vector medio condicional es una función lineal y la matriz de covarianzas condicional depende solo de la matriz de covarianzas de la distribución conjunta.

   Sea $\mathbf{Y}$ particionada como en la propiedad 1, entonces si $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),\ \boldsymbol{\Sigma} > 0$, para cualquier $e > p$ la distribución condicional de $\mathbf{Y_1}$ dado $\mathbf{Y_2} = \mathbf{y_2}$ es $N_e(\boldsymbol{\mu}_{1|2},\ \boldsymbol{\Sigma}_{11|2})$, donde $\boldsymbol{\mu}_{1|2}$ es el vector de medias condicional y $\boldsymbol{\Sigma}_{11|2}$ es la matriz de covarianzas condicional de $\mathbf{Y_1}$ dado $\mathbf{Y_2} = \mathbf{y_2}$. Además, $\boldsymbol{\mu}_{1|2}$ es una función lineal de $\mathbf{y_2}$ y $\boldsymbol{\Sigma}_{11|2}$ no depende de $\mathbf{y_2}$.

5. La variable aleatoria $\mathbf{Y}$ tiene una distribución normal p-variada si y solo si al estandarizar $\mathbf{Y}$ su distribución es una normal p-variada estándar.

   Sea $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ con matriz de covarianzas $\boldsymbol{\Sigma}$ invertible, entonces

   $$\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}-\boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I})$$

   Para la estandarización empírica de los datos es necesario utilizar la estimación de $\boldsymbol{\Sigma}^{-1/2}$ denotada por $\mathbf{S}^{-1/2}$. El cálculo de $\mathbf{S}^{-1/2}$ se realiza mediante la descomposición de Schur, así, sea $\boldsymbol{\Delta}$ la matriz de los vectores propios de $\mathbf{S}^{-1}$ y sea $\boldsymbol{\Omega}$ una matriz diagonal con los valores propios de $\mathbf{S}^{-1}$. Entonces,

   $$\begin{aligned}\mathbf{S}^{-1/2} &= (\mathbf{S}^{-1})^{1/2}\\ &= \boldsymbol{\Delta}\boldsymbol{\Omega}^{1/2}\boldsymbol{\Delta}^{-1}\\ &= \boldsymbol{\Delta}\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Delta}'\end{aligned} \tag{2-2}$$

   que de acuerdo con la definición de raíz cuadrada, $\mathbf{S}^{1/2} = \mathbf{H}$ si $\mathbf{H}\mathbf{H} = \mathbf{S}$. Así, al multiplicar el resultado de (2-2) por sí mismo, se llega a la expresión $\boldsymbol{\Delta}\boldsymbol{\Omega}\boldsymbol{\Delta}'$, la cual coincide con la descomposición de Schur.

Las propiedades enunciadas dan cuenta de todas las implicaciones que tiene probar la normalidad multivariada. Trabajos de grado anteriores proporcionan indicios de que el vector de horas donde se miden los niveles de $PM_{2.5}$ siguen una distribución normal 24-variada. Por ello, con la modificación propuesta se espera contribuir a la verificación o refutación de dicha presunción.

Dado que los parámetros de una distribución son desconocidos, se han desarrollado algunas técnicas estadísticas para estimarlos adecuadamente dependiendo de las condiciones de los datos y las variables medidas.

2.1.1. Estimación de $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ y $\boldsymbol{\rho}$

Sean $\mathbf{y_1}, ..., \mathbf{y_n}$ vectores de dimensión $(p \times 1)$ que representan una muestra aleatoria de una población normal multivariada con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$. $\mathbf{y_1}, ..., \mathbf{y_n}$ son mutuamente independientes con distribución $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y su función de densidad conjunta es:

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{j=1}^{n} \left(\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} exp\left[-\frac{1}{2}(\mathbf{y_j} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y_j} - \boldsymbol{\mu}) \right] \right) \\ &= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} exp\left[-\frac{1}{2} \sum_{j=1}^{n} (\mathbf{y_j} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y_j} - \boldsymbol{\mu}) \right] \end{aligned} \tag{2-3}$$

Al considerar (2-3) en función de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ para un conjunto finito de observaciones, esta pasa a ser la función de verosimilitud. La técnica que mejor explica los datos observados es la estimación de máxima verosimilitud, que consiste en seleccionar los valores de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ que maximicen la densidad conjunta en (2-3) evaluada en las observaciones. Los estimadores que se encuentran de esta forma se llaman estimadores de máxima verosimilitud:

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{y_j} \quad \text{y} \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{y_j} - \bar{\mathbf{Y}})(\mathbf{y_j} - \bar{\mathbf{Y}})' \tag{2-4}$$

Sin embargo, $\hat{\boldsymbol{\Sigma}}$ en (2-4) es un estimador sesgado puesto que

$$E(\hat{\boldsymbol{\Sigma}}) = \frac{(n-1)\boldsymbol{\Sigma}}{n}$$

Por tanto, se define un estimador insesgado de $\boldsymbol{\Sigma}$ de la forma

$$\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{y_j} - \bar{\mathbf{Y}})(\mathbf{y_j} - \bar{\mathbf{Y}})' \tag{2-5}$$

En este trabajo de grado se toma como estimador tradicional de $\boldsymbol{\Sigma}$ a $\mathbf{S}$. En Díaz y Morales (2012) se encuentran los detalles de la demostración y se derivan las siguientes propiedades sobre $\bar{\mathbf{Y}}$ y $\mathbf{S}$ como los respectivos estimadores de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$:

1. El estadístico $\bar{\mathbf{Y}}$ tiene distribución $N(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$ si la muestra aleatoria de tamaño n proviene de una población $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

2. La distribución de la matriz de covarianzas muestral está ligada a una distribución Wishart, por tanto el estadístico $n\mathbf{S}$ tiene distribución $W(\boldsymbol{\Sigma}, n-1)$.

3. $\bar{\mathbf{Y}}$ y $\mathbf{S}$ son independientes.

4. $\bar{\mathbf{Y}}$ y $\mathbf{S}$ son estimadores insesgados de $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$, respectivamente.

5. $\bar{\mathbf{Y}}$ y $\mathbf{S}$ son estadísticos consistentes.

6. $\bar{\mathbf{Y}}$ y $\mathbf{S}$ son estadísticos suficientes.

Por lo que estos estadísticos cumplen con propiedades deseables para ser elegidos como candidatos óptimos. No obstante, los autores Schäfer y Strimmer (2005) sintetizan los aspectos de uso de los estimadores máximo verosímiles; como primer aspecto, se tiene el estimador máximo verosímil como medio para resumir los datos observados y generar un resumen de máxima verosimilitud. Como segundo aspecto, el procedimiento para obtener una estimación máximo verosímil. Luego, se concluye que la máxima verosimilitud es incuestionable como resumen de datos pero que tiene algunos defectos claros como procedimiento de estimación. Sin embargo, bajo problemas inferenciales de alta dimensión, el estimador de máxima verosimilitud requiere mejora (p.5).

Finalmente, la estimación de la matriz de correlaciones $\boldsymbol{\rho}$ denotada por $\mathbf{R}$ se puede expresar como

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \tag{2-6}$$

2.1.2. Estimación tradicional de la matriz de covarianzas

El estimador $\mathbf{S}$ de la matriz de covarianzas definido en (2-5), se obtiene luego de maximizar la verosimilitud de los datos provenientes de una población multinormal y modificar levemente su expresión para que sea una estimador insesgado, habitualmente se toma como una medida descriptiva cuando los datos no provienen de una población multinormal. Por tal motivo, en este trabajo se le llamará a $\mathbf{S}$ como la estimación tradicional de la matriz de covarianzas. Adicionalmente, en este trabajo se hace referencia a la prueba Shapiro Wilk Generalizada como la prueba tradicional.

2.2. Condicionamiento de las matrices de covarianzas poblacional y muestral

Se define una medida de condicionamiento que brinda indicios del comportamiento de los datos poblacionales. Este valor se define como:

$$d = \frac{\gamma_{max}}{\gamma_{min}} \geq 1, d \in \mathbb{R} \tag{2-7}$$

En (2-7) los valores propios de la matriz de covarianzas muestral que se esté utilizando (sea $\mathbf{S}$ o $\mathbf{S}^*$) son denotados por γ_i y el cociente d se llama el número de condición. El por qué este número de condición es relevante y da información acerca del comportamiento de los datos está en la interpretación de los valores y vectores propios de la matriz de covarianza. El vector propio de mayor magnitud de la matriz de covarianza apunta en la dirección de la mayor varianza de los datos y la magnitud de este vector es igual al valor propio correspondiente. El segundo vector propio más grande siempre es perpendicular al vector propio más grande y apunta en la dirección de la segunda dispersión más grande de los datos, y así sucesivamente, entonces si el menor valor propio difiere mucho en magnitud del mayor valor propio, el cambio de dispersión entre la dirección del mayor propio y el menor valor propio será grande. Se establece que la matriz de covarianzas está bien condicionada si d es pequeño.

Por su parte, Ledoit y Wolf (2004) afirman que para matrices de covarianza de gran dimensión, la matriz de covarianzas muestral, generalmente no está bien condicionada y es posible que ni siquiera sea invertible, por lo que proponen otro estimador que está bien condicionado y es más preciso que la matriz de covarianza muestral asintóticamente. Sin embargo, es de esperarse que un buen estimador se parezca a su parámetro y en el caso de que la matriz de covarianzas poblacional se encuentre mal condicionada, considerando la indicación Ledoit y Wolf (2004), ¿por qué sería de interés un estimador mejor condicionado que la estimación tradicional si en este caso se alejaría del comportamiento de la matriz de covarianzas poblacional? Para responder esta pregunta, Ledoit y Wolf (2004) se basan en el siguiente resultado del álgebra matricial.

Teorema 2.2.1 *Los valores propios son los elementos diagonales más dispersos que se pueden obtener por rotación.*

La descomposición de la matriz de covarianzas poblacional en valores propios y vectores propios: $\mathbf{\Sigma} = \mathbf{\Gamma}'\mathbf{\Lambda}\mathbf{\Gamma}$, donde $\mathbf{\Lambda}$ es una matriz diagonal y $\mathbf{\Gamma}$ es una matriz de rotación. Los elementos diagonales de $\mathbf{\Lambda}$ son los valores propios $\lambda_1, ..., \lambda_p$ y las columnas de $\mathbf{\Gamma}$ son los vectores propios $\mathit{\Pi}_\mathbf{1}, ..., \mathit{\Pi}_\mathbf{p}$. De manera similar, se descompone a la matriz de covarianza de la muestra en valores propios y vectores propios: $\mathbf{S} = \mathbf{\Psi}'\mathbf{\Upsilon}\mathbf{\Psi}$, donde $\mathbf{\Upsilon}$ es una matriz diagonal y $\mathbf{\Psi}$ es una matriz de rotación. Los elementos diagonales de $\mathbf{\Upsilon}$ son los valores propios $\varepsilon_1, ..., \varepsilon_p$, y las columnas de $\mathbf{\Psi}$ son los vectores propios $\mathit{\Delta}_\mathbf{1}, ..., \mathit{\Delta}_\mathbf{p}$. En este caso, $\mathbf{\Lambda} = \mathbf{\Gamma}\mathbf{\Sigma}\mathbf{\Gamma}'$ y $\mathbf{\Lambda} = \mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma}$ son equivalentes porque $\mathbf{\Gamma}$ es una matriz de rotación, de forma análoga ocurre con $\mathbf{\Upsilon}$.

Se plantea que $\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma}$ es un estimador insesgado de $\mathbf{\Lambda} = \mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma}$. Los elementos diagonales de $\mathbf{\Gamma}'\mathbf{S}\mathbf{\Gamma}$ están aproximadamente tan dispersos como los de $\mathbf{\Lambda} = \mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma}$. Por conveniencia, se asume que están exactamente igual de dispersos. Por el contrario, $\mathbf{\Upsilon} = \mathbf{\Psi}'\mathbf{S}\mathbf{\Psi}$ no es en absoluto un estimador imparcial de $\mathbf{\Lambda} = \mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma}$. Debido a que $\mathbf{\Upsilon}$ son exactamente los valores propios asociados a $\mathbf{S}$ y por Teorema 2.2.1, se concluye que los elementos diagonales de

$\boldsymbol{\Upsilon} = \boldsymbol{\Psi}'\mathbf{S}\boldsymbol{\Psi}$ están más dispersos que los de $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}'\mathbf{S}\boldsymbol{\Gamma}$ (y por lo tanto que los de $\boldsymbol{\Lambda} = \boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma}$). Esta es la razón por la cual los valores propios de la muestra están más dispersos que los poblacionales. Por tal motivo, sin asumir que la matriz de covarianzas poblacional tiene determinado comportamiento, siempre se deseará otro estimador que esté mejor condicionado que $\mathbf{S}$.

Dando respuesta a la pregunta planteada, en el caso de una matriz de covarianza poblacional mal condicionada, $\mathbf{S}$ estará un tanto más mal condicionada y por tanto será necesario regular su condicionamiento para que su mal condicionamiento sea similar al de su parámetro y no excesivo. Además, Ledoit y Wolf (2004) afirman que de forma general, cuando $n \cong p$, el sesgo en los valores propios aumenta, es decir que los valores propios de la muestra más grandes están severamente sesgados hacia arriba y los más pequeños hacia abajo. Por último, se resalta que no se debe confundir el mal o buen condicionamiento con un mal o buen estimador de la matriz de covarianza puesto que el condicionamiento actúa como medida que describe la dispersión de los datos y no como medida de calidad del estimador.

A continuación se define formalmente la estimación shrinkage.

2.2.1. Estimación shrinkage de la matriz de covarianzas

En esta subsección se exponen los principios generales detrás de la estimación shrinkage propuesta originalmente por Ledoit y Wolf (2003) aplicado a datos de cartera financiera y complementado por Schäfer y Strimmer (2005) aplicado al estudio genómico funcional, se resalta que en ambos escenarios está presente la gran dimensionalidad que requiere un tratamiento distinto para la calidad de sus estimadores. Este planteamiento aplica para cualquier estimador, en particular, se abordará desde el estimador de la matriz de covarianzas. Lo anterior conlleva al reto estadístico subyacente de estudiar el enfoque analítico para determinar el nivel óptimo de contracción.

Siguiendo la idea expuesta por Schäfer y Strimmer (2005), por un lado, la matriz simétrica de covarianzas sin restricciones, tiene por estimar $\frac{p^2+p}{2}$ parámetros libres que corresponden a los elementos en la diagonal principal y los que están por encima (o por debajo) de la diagonal, de aquí, sea $\mathbf{E} = (E_1, E_2, ..., E_{\frac{p^2+p}{2}})$ el vector que contiene los parámetros del modelo de interés de alta dimensión sin restricciones. Por otro lado, se considera como un ejemplo cuando todas las varianzas son iguales, esto es, $\sigma_{11} = \sigma_{22} = \cdots = \sigma_{pp}$, y se denomina como una matriz de covarianzas con restricciones, que particularmente tiene por estimar $\frac{p^2+p}{2} - (p-1)$ parámetros libres que corresponden a un elemento de la diagonal principal (el resto de la diagonal toma el mismo valor del estimado) y los que están por encima (o por debajo) de la diagonal, de aquí, sea $\boldsymbol{\Theta} = (\theta_1, \theta_2, ..., \theta_{\frac{p^2+p}{2}-(p-1)})$ los parámetros coincidentes de un submodelo restringido de menor dimensión. La explicación del submodelo restringido se aborda

con el ejemplo donde se asume homocedasticidad para ilustrar que al dar una restricción, se reduce significativamente la cantidad de parámetros a estimar, pero no es la única posible restricción que se podría plantear (covarianzas nulas, covarianzas nulas y heterocedasticidad, covarianzas nulas y varianzas unitarias, etc). Ahora bien, al ajustar cada uno de los dos modelos diferentes a los datos observados, por un lado, se obtienen las estimaciones $\hat{\mathbf{E}}$ asociadas a los $\frac{p^2+p}{2}$ elementos de la matriz de covarianzas sin restringir y se organizan en una matriz simétrica $\mathbf{S}$ de dimensión $(p \times p)$ y por otro lado se obtienen las estimaciones $\hat{\mathbf{\Theta}}$ asociadas a los $\frac{p^2+p}{2} - (p-1)$ elementos de la matriz de covarianzas restringida particularmente por la homocedasticidad y se organizan en una matriz simétrica denotada por $\mathbf{T}$ de dimensión $(p \times p)$ (la definición óptima de la matriz $\mathbf{T}$ se presenta en la sección 3.2.3). Expresamente, la estimación sin restricciones $\mathbf{S}$ tendrá una varianza comparativamente alta debido a la mayor cantidad de parámetros que deben ajustarse, mientras que su contra parte $\mathbf{T}$ de baja dimensión tendrá una varianza más baja pero potencialmente, también un sesgo considerable como estimador de la verdadera $\mathbf{\Sigma}$.

En lugar de elegir entre uno de estos dos extremos, el enfoque de contracción lineal plantea combinar ambos estimadores mediante un promedio ponderado. Así, la estimación shrinkage se define como

$$\mathbf{S}^* = \lambda\mathbf{T} + (1-\lambda)\mathbf{S} \tag{2-8}$$

Donde λ es un número entre [0, 1] que denota la intensidad de contracción, $\mathbf{T}$ es la matriz objetivo (estructurada) y $\mathbf{S}$ la matriz de covarianzas muestral.

La matriz $\mathbf{T}$ debe cumplir dos requisitos al mismo tiempo: solo implica una pequeña cantidad de parámetros libres (es decir, mucha estructura) pero también refleja importantes características de la cantidad desconocida que se está estimando.

La estimación shrinkage ofrece una forma de regular de forma sistemática la estimación de $\mathbf{S}^*$ con el objetivo de superar las estimaciones individuales $\mathbf{T}$ (varianza baja y sesgada) y $\mathbf{S}$ (varianza alta) tanto en precisión como en eficiencia estadística.

De (2-8) se observa que si la intensidad de contracción λ es 1, entonces, la estimación de contracción es igual al objetivo de contracción $\mathbf{T}$, en cambio, si λ es 0, se recupera la estimación $\mathbf{S}$ sin restricciones. Por lo anterior, es fundamental definir cómo seleccionar un valor óptimo para el parámetro de contracción y esto se propone mediante el estimador que basado en datos minimice explícitamente la siguiente función de riesgo que mide la distancia esperada entre la estimación shrinkage y la matriz de covarianzas poblacional.

$$R(\lambda) = \mathrm{E}(L(\lambda)) = \mathrm{E}(\sum_{i=1}^{p}\sum_{j=1}^{p}(s_{ij}^* - \sigma_{ij}^2))$$

Suponiendo que existen los dos primeros momentos de las distribuciones de $\mathbf{S}$ y $\mathbf{T}$, $R(\lambda)$ puede ampliarse como:

$$\begin{aligned} R(\lambda) &= \sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{Var}(s_{ij}^*) + [\mathrm{E}(s_{ij}^*) - \sigma_{ij}^2]^2 \\ &= \sum_{i=1}^{p}\sum_{j=1}^{p} \lambda^2 \mathrm{Var}(t_{ij}) + (1-\lambda)^2 \mathrm{Var}(s_{ij}) + 2\lambda(1-\lambda)\mathrm{Cov}(s_{ij}, t_{ij}) + [\lambda \mathrm{E}(t_{ij} - s_{ij}) + \mathrm{Bias}(s_{ij})]^2 \end{aligned}$$

Minimizando la expresión del riesgo con respecto a λ, Ledoit y Wolf (2003) proponen el siguiente estimador,

$$\lambda^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{Var}(s_{ij}) - \mathrm{Cov}(t_{ij}, s_{ij}) - \mathrm{Bias}(s_{ij})\mathrm{E}(t_{ij} - s_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{E}[(t_{ij} - s_{ij})^2]}$$

Si $\mathbf{S}$ es un estimador insesgado para $\mathbf{\Sigma}$, es decir, $\mathrm{E}(\mathbf{S}) = \mathbf{\Sigma}$, entonces la expresión se reduce así,

$$\lambda^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{Var}(s_{ij}) - \mathrm{Cov}(t_{ij}, s_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p} \mathrm{E}[(t_{ij} - s_{ij})^2]} \tag{2-9}$$

El cual es asintóticamente óptimo. Una inspección más cercana de (2-9) permite construir las siguiente ideas sobre la elección del λ^* óptimo.

1. El parámetro de contracción λ^* disminuye cuando la varianza de la estimación de alta dimensión de $\mathbf{S}$ también disminuye. Por lo tanto, a medida que aumenta el tamaño de la muestra, $(1 - \lambda^*)$ asigna un peso mayor a $\mathbf{S}$ y λ^* asigna un peso menor a la matriz $\mathbf{T}$ objetivo. En síntesis, cuando el tamaño de muestra incrementa, $\mathbf{S}$ mejora y requiere cada vez menos la propuesta de contracción.

2. Considerando que la covarianza se puede ver en términos de la correlación, se observa que, λ^* depende de la correlación entre el error de estimación de $\mathbf{S}$ y el de $\mathbf{T}$. Si ambos están positivamente correlacionados, el peso asignado al objetivo de contracción disminuye. Lo anterior viene de que los dos estimadores $\mathbf{S}$ y $\mathbf{T}$ se infieren del mismo conjunto de datos y que en la medida en que uno se explica por el otro, es porque requieren cada vez menos la propuesta de contracción.

3. Fijándose en el denominador, con el aumento de la diferencia cuadrática media entre $\mathbf{S}$ y $\mathbf{T}$, el cociente disminuye, por lo que λ^* también disminuye. Esto cuida la estimación de contracción $\mathbf{S}^*$ contra un objetivo $\mathbf{T}$ que esté mal especificado, en consecuencia, se asigna más peso a $\mathbf{S}$.

4. Si el estimador sin restricciones $\mathbf{S}$ está sesgado y el sesgo apunta hacia el objetivo $\mathbf{T}$, la intensidad de la contracción se reduce correspondientemente.

Con la intención de aprovechar al máximo el lema de Ledoit y Wolf (2003), los autores Schäfer y Strimmer (2005) proponen un estimador muestral shrinkage novedoso denotado como $\hat{\lambda^*}$. Dichos autores argumentan que dado que, el interés es la estimación cuando se tienen muestras pequeñas; entonces la consistencia pasa a ser un requisito muy débil ya que es una propiedad asintótica, de modo que plantean reemplazar todas las esperanzas, varianzas y covarianzas de λ^* por sus contrapartes insesgadas, así obtienen

$$\hat{\lambda^*} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \hat{Var}(s_{ij}) - \hat{Cov}(\hat{t_{ij}}, s_{ij}) - \hat{Bias}(s_{ij})(t_{ij} - s_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p}(t_{ij} - s_{ij})^2} \tag{2-10}$$

Finalmente, redefinen al estimador de contracción óptimo $\hat{\lambda}^*$ por uno que además de ser óptimo también esté regulado y este se denota por $\hat{\lambda}^M$. Así,

$$\hat{\lambda}^M = \text{máx}(0, min(1, \hat{\lambda}^*)) \tag{2-11}$$

también estima al parámetro de contracción λ pero asegurándose de que además de ser el óptimo también evita la contracción excesiva o negativa.

Schäfer y Strimmer (2005) garantizan que su propuesta de la estimación shrinkage de la matriz de covarianzas tiene un error cuadrático medio mínimo y siempre es una matriz positiva definida incluso para tamaños de muestra pequeño.

2.2.2. Estadísticos de orden

En esta subsección se presentan algunas definiciones entorno a la teoría de los estadísticos de orden debido a que la prueba Shapiro Wilk univariada se basa en estadísticos de orden y la prueba Shapiro Wilk Generalizada construye su estadístico a partir de estadísticos Shapiro Wilk univariados.

Se denota como i.i.d. al conjunto de observaciones que son independientes e idénticamente distribuidas, es decir que cada observación sigue la misma distribución de probabilidad y todas son mutuamente independientes. Ahora bien, si una característica es medida muchas veces, las observaciones pueden variar y se podría pensar a las observaciones como variables aleatorias.

Sean las variables aleatorias $X_1, ..., X_n \in \mathbb{R}$ i.i.d. con distribución de probabilidad arbitraria. Si el conjunto de variables aleatorias X_i, $i = 1, ..., n$ se ordena de manera ascendente, estos quedan denotados como

$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$,

por lo que $X_{(i)}$ denotara el i-ésimo estadístico de orden de una muestra de tamaño n. Cabe resaltar que, a diferencia de la muestra aleatoria en sí, los estadísticos de orden no son independientes, puesto que si $x_i \geq x$, entonces $x_{i+1} \geq x$, por ejemplo, el estadístico x_n (máximo)

para ser máximo depende de que todos los demás estadísticos sean menores que él.

Para ilustrar lo anterior, suponga que la variable aleatoria X fue medida en tres individuos mutuamente independientes e idénticamente distribuidos. Las observaciones fueron $x_1 = 2$, $x_2 = 3$ y $x_3 = 1$, entonces, el i-ésimo estadístico de orden para este conjunto de observaciones se define como el i-ésimo valor más pequeño de la muestra. Así, el estadístico de orden i-ésimo para este experimento es el i-ésimo valor más pequeño del conjunto $\{2, 3, 1\}$, por lo que el primer estadístico de orden es $x_{(1)} = 1$ (el mínimo); el segundo estadístico de orden es $x_{(2)} = 2$ y el tercer estadístico de orden es el tercer valor más pequeño $x_{(3)} = 3$ (el máximo). Al repetir este proceso muchas veces, el i-ésimo valor más pequeño para cada conjunto de observaciones será variable.

1 **Función de densidad conjunta**
Sea $X_1, ..., X_n \in \mathbb{R}$ variables aleatorias i.i.d. y sean $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ los estadísticos de orden correspondientes. Se denotan los valores de los n estadísticos de orden como $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$. De Balakrishnan y Clifford (1991) se toma que la densidad conjunta para los n estadísticos de orden queda expresada como

$$f_{1,...,n}(x_{(1)}, x_{(2)}, ..., x_{(n)}) = n! \prod_{i=1}^{n} f(x_{(i)}), \qquad -\infty < x_{(1)} < ... < x_{(n)} < \infty$$

Ahora, sean $X_{(i)}$ y $X_{(j)}$ $(1 \leq i \leq j \leq n)$ estadísticos de orden, la distribución conjunta está dada por

$$\begin{aligned} f_{i,j}(x_{(i)}, x_{(j)}) &= n! f(x_{(i)}) f(x_{(j)}) \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} [F(x_{(i)})]^{i-1} [F(x_{(j)}) - F(x_{(i)})]^{j-i-1} \\ &\quad [1 - F(x_{(j)})]^{n-j} f(x_{(i)}) f(x_{(j)}), \qquad -\infty < x_{(i)} < x_{(j)} < \infty \end{aligned} \tag{2-12}$$

2 **Función de densidad marginal**
La función de densidad marginal para cualquier estadístico de orden se define como

$$f_i(x_{(i)}) = \frac{n!}{(i-1)!(n-i)!} [F(x_{(i)})]^{i-1} [1 - F(x_{(i)})]^{n-i} f(x_{(i)}), \qquad -\infty < x_{(i)} < \infty \tag{2-13}$$

donde $F(x_{(i)})$ y $f(x_{(i)})$ denotan la función de distribución acumulada y la función de densidad de la variable aleatoria X, respectivamente.

Por lo que la función de densidad para el primer estadístico de orden (mínimo) es

$$f_1(x_{(1)}) = n[1 - F(x_{(1)})]^{n-1} f(x_{(1)}), \qquad -\infty < x_{(1)} < \infty$$

Del mismo modo, la función de densidad para el n-ésimo estadístico de orden (máximo) es

$$f_n(x_{(n)}) = n[F(x_{(n)})]^{n-1} f(x_{(n)}), \qquad -\infty < x_{(n)} < \infty$$

3 **Momentos muestrales**

Dado que se cuenta con la función de densidad para cualquier estadístico de orden, entonces, se pueden obtener sus respectivos momentos.

Sea $X_1, ..., X_n \in \mathbb{R}$ una muestra aleatoria que proviene de una población con función de densidad denotada por $f(x)$ y función de distribución acumulada $F(x)$, y sea $X_{(1)} \leq ... \leq X_{(n)}$ los estadísticos de orden obtenidos de la muestra anterior. Para efectos de este trabajo, se denotara al q-ésimo momento muestral como $E[X_{(i)}^q] = \tau_i^q$. De la función de densidad expuesta en (2-13), se tiene que

$$\begin{aligned}\tau_i^q =& E[X_{(i)}^q] = \int_{-\infty}^{\infty} x_q f_i(x) dx \\ =& \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x^q [F(x)]^{i-1} [1-F(x)]^{n-i} f(x) dx, \quad i = 1, 2, ..., n;\ q \geq 1\end{aligned}$$

Considerando el primer y segundo momento, la varianza se puede calcular como

$$Var(X_{(i)}) = \tau_i^2 - [\tau_i^1]^2, \quad 1 \leq i \leq n \tag{2-14}$$

Ahora, se denota al momento del producto como $\tau_{i,j}^1 = E[X_{(i)}X_{(j)}]$. De la función de densidad conjunta en (2-12) se tiene

$$\begin{aligned}\tau_{i,j}^1 = E[X_{(i)}X_{(j)}] =& \iint\limits_{-\infty<x<y<\infty} xy f_{i,j}(x,y) dy dx \\ =& \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \iint\limits_{-\infty<x<y<\infty} xy [F(x)]^{i-1} [F(y)-F(x)]^{j-i-1} \\ & [1-F(y)]^{n-j} f(x) f(y) dy dx, \quad 1 \leq i < j \leq n\end{aligned} \tag{2-15}$$

Con este último resultado, se puede calcular la covarianza entre $X_{(i)}$ y $X_{(j)}$, así

$$\text{Cov}(X_{(i)}, X_{(j)}) = \tau_{i,j}^1 - \tau_i^1 \tau_j^1, \quad 1 \leq i < j \leq n \tag{2-16}$$

- **Elementos de los estadísticos de orden cuando X sigue una distribución normal estándar**

Teniendo en cuenta las definiciones anteriores, a continuación se presentan los resultados para la distribución normal.

Sea $X_1, ..., X_n \in \mathbb{R}$ una muestra aleatoria que proviene de una población normal estándar con función de densidad

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty$$

y función de distribución acumulada

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du, \quad x \in \mathbb{R}$$

Sean los estadísticos de orden $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ obtenidos de la muestra anterior, entonces de Balakrishnan y Clifford (1991) el primer momento muestral para los estadísticos de orden queda definido como:

$$m_i^1 = \tau_i^1 = \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x \left[\frac{1}{2} - \Phi(x)\right]^{i-1} \left[\frac{1}{2} + \Phi(x)\right]^{n-i} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

donde $\Phi(x)$ denota la función de distribución acumulada de la distribución normal.

Considerando (2-14), la varianza de los estadísticos de orden para una muestra aleatoria que proviene de una distribución normal es la diferencia entre el p, así

$$Var(X_{(i)}) = m_i^2 - [m_i^1]^2, \quad 1 \leq i \leq n.$$

Finalmente, cuando X sigue una distribución normal estándar, el momento del producto $\tau_{i,j}^1 = E[X_{(i)}X_{(j)}]$ se denota por $m_{i,j}^1$ y de acuerdo con (2-16) la covarianza entre $X_{(i)}$ y $X_{(j)}$ es

$$\text{Cov}(X_{(i)}, X_{(j)}) = m_{i,j}^1 - m_i^1 m_j^1, \quad 1 \leq i < j \leq n$$

En adelante, para hacer referencia al primer momento se omitirá el superíndice 1.

2.3. Pruebas (contraste) de hipótesis

La definición de una hipótesis es bastante general, pero el punto importante es que una hipótesis hace una manifestación sobre la población. El objetivo de una prueba de hipótesis es decidir, juzgando si una propiedad que se supone en una población es compatible con lo observado en una muestra de dicha población, cuál de dos hipótesis complementarias es verdadera. En ella se considera una hipótesis nula como H_0 y una hipótesis alternativa H_1. La hipótesis H_0 nunca se considera probada, aunque puede ser rechazada por los datos.

La prueba de hipótesis como procedimiento es una regla que especifica para qué valores muestrales se toma la decisión de rechazar H_0. El conjunto del espacio muestral para el que será H_0 se denomina región de rechazo o región crítica. El complemento de la región de rechazo se llama región de aceptación.

Las pruebas de bondad de ajuste son pruebas de hipótesis para verificar si los datos observados en una muestra aleatoria se ajustan con algún nivel de significancia a determinada

distribución de probabilidad. Un interés particular de este trabajo, es cuando se desea determinar si la distribución de los datos observados proviene de una distribución normal univariada o multivariada, según sea el caso, existen diversas pruebas para verificar o refutar este supuesto, las pruebas varían en condiciones preliminares de la información. A continuación se describe de forma detallada la prueba Shapiro Wilk univariada y una generalización de dicha prueba para evaluar multinormalidad.

2.3.1. Prueba Shapiro-Wilk univariada

En esta subsección se hace un revisión de los aspectos mas relevantes de la prueba Shapiro Wilk univariada expuesta originalmente por Shapiro y Wilk (1965) y complementando con las ideas expuestas en la tesis de Hain (2010).

Esta prueba es considerada entre las pruebas más potentes para determinar si una muestra aleatoria proviene de un población normalmente distribuida.

Se plantean las siguientes hipótesis para una muestra $y_1, ..., y_n$ tal que:

$H_0 = y_1, ..., y_n \in \mathbb{R}$ provienen de una población normalmente distribuida $N(\mu, \sigma)$
$H_1 = y_1, ..., y_n \in \mathbb{R}$ no provienen de una población normalmente distribuida $N(\mu, \sigma)$

Se considera una muestra aleatoria $x_1, ..., x_n \in \mathbb{R}$, las cuales son i.i.d procedentes de una variable aleatoria X que sigue una distribución normal estándar, esto es, con media $\mu = 0$ y varianza $\sigma = 1$.

Así, la función de distribución acumulada viene dada por,

$$P(x_i \leq t) = \Phi(t) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{t} \exp\,(-\frac{(u-\mu)^2}{2\sigma})du$$
$$= \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\,(-\frac{(u)^2}{2})du$$

Los valores ordenados de la muestra aleatoria x_i son denotados por $x_{(1)} \leq ... \leq x_{(n)}$ y $\boldsymbol{X'} = (x_{(1)}, ..., x_{(n)})$ como el vector de variables aleatorias ordenadas.

De acuerdo con la subsección 2.2.2 que se refiere a los estadísticos de orden, se define $\boldsymbol{m'} = (m_1, ..., m_n)$ de tamaño $(1 \times n)$ como el vector de valores esperados de los estadísticos de orden $x_{(i)}$ y $V = (v_{ij})$ la matriz de covarianzas correspondiente de tamaño $(n \times n)$. Por tanto, sea

$E(x_{(i)}) = m_i, \quad i = 1, ..., n$

y

$$Cov(x_{(i)}, x_{(j)}) = v_{ij}, \quad i, j = 1, ..., n$$

Donde m_i se refiere al valor esperado del estadístico de orden i-ésimo de una variable aleatoria distribuida normal estándar y v_{ij} se refiere a la covarianza del i-ésimo y el j-ésimo estadístico de orden de la misma variable aleatoria.

Ahora bien, retomando el objetivo principal, se desea probar que la variable aleatoria Y está normalmente distribuida con media μ y varianza σ desconocidas. Para esto, se dispone de una muestra aleatoria $y_1, ..., y_n$ las cuales son i.i.d. Ahora, los valores ordenados en forma ascendente de la muestra aleatoria son denotados por $y_{(1)} \leq ... \leq y_{(n)}$ y $\mathbf{Y}' = (y_{(1)}, ..., y_{(n)})$ es el vector aleatorio de las observaciones ordenadas.

Para comprender la idea principal de la prueba Shapiro-Wilk univariada, inicialmente, se tendrá en cuenta que si cada y_i está normalmente distribuida, entonces, $\frac{y_i - \mu}{\sigma^{1/2}}$ sigue una distribución normal estándar. Bajo la hipótesis nula de normalidad, se puede expresar a $y_{(i)}$ como

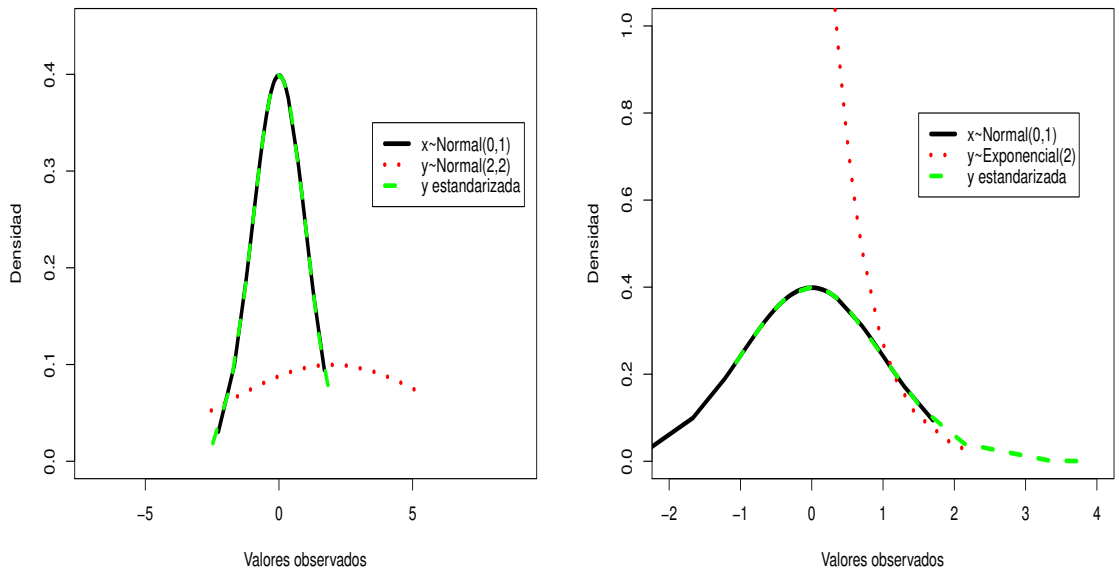
$$\begin{aligned} \frac{y_{(i)} - \mu}{\sigma^{1/2}} &= x_{(i)}, \qquad i = 1, ..., n. \\ y_{(i)} &= \mu + \sigma^{1/2} x_{(i)}, \qquad i = 1, ..., n. \end{aligned} \tag{2-17}$$

Los $y_{(i)}$ serían iguales a μ excepto por su propia variabilidad, entonces, si se interpreta a $\sigma^{1/2} x_{(i)}$ como el componente aleatorio que diferencia a $y_{(i)}$ de μ, se podría definir nuevos términos del error como $\epsilon_i = \sigma^{1/2} x_{(i)} - E(\sigma^{1/2} x_{(i)}) = \sigma^{1/2} x_{(i)} - \sigma^{1/2} m_i$, los cuales tienen valor esperado igual a cero. En consecuencia, (2-17) puede ser escrita como

$$y_{(i)} = \mu + \sigma^{1/2} m_i + \epsilon_i, \qquad i = 1, .., n. \tag{2-18}$$

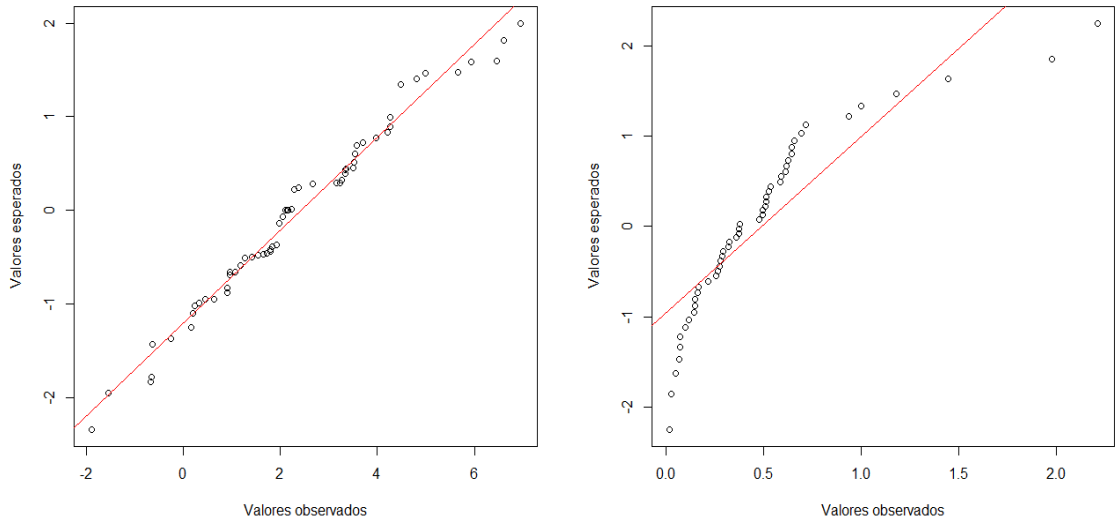
Esto indica que, bajo la condición de que las y_i provienen de una población normalmente distribuida, una gráfica de las observaciones ordenadas $y_{(1)}, ..., y_{(n)}$ y las esperanzas $m_1, ..., m_n$ de los estadísticos de orden de una distribución normal estándar debe ser aproximadamente lineal.

Como ilustración a lo expuesto en (2-17), se presenta en la Figura **2-1** una simulación para comparar los valores estandarizados de una variable aleatoria normal y los valores estandarizados de una variable aleatoria no-normal. En la Figura **2-1** (a) se observa que, cuando $Y \sim N(2, 2)$ (curva roja), entonces, $\frac{Y - \mu}{\sigma^{1/2}}$ (curva azul) es aproximadamente como $X \sim N(0, 1)$ (curva negra). En cambio, en la Figura **2-1** (b) se observa que cuando $Y \sim Exponencial(2)$ (curva roja), entonces, $\frac{Y - \mu}{\sigma^{1/2}}$ (curva azul) tiene un dominio diferente (corrido hacia la derecha) a $X \sim N(0, 1)$ (curva negra).



(a) (b)

Figura 2-1: Simulación de datos ($n = 50$) para comparar $X \sim N(0,1)$ vs Y vs $\frac{Y-\mu}{\sigma^{1/2}}$



(a) (b)

Figura 2-2: Regresión entre los valores esperados de los estadísticos de orden que son función de $X \sim N(0,1)$ vs las observaciones ordenadas de Y cuando la distribución subyacente es normal (a) y cuando es no-normal (b)

Ahora, la relación expuesta en (2-18) se ilustra en la Figura **2-2**, haciendo una comparación de la regresión cuando la distribución subyacente es normal y cuando es no-normal. En la Figura **2-2** (a) se aprecia una relación bastante lineal al graficar las observaciones ordenadas de $Y \sim N(2,2)$ y los valores esperados m_i de los estadísticos de orden que son funciones de $X \sim N(0,1)$. En contraste con la Figura **2-2** (b) donde se observa un patrón fuerte de distanciamiento de la linealidad en los extremos, al graficar las observaciones ordenadas de $Y \sim Exponencial(2)$ y los valores esperados m_i de los estadísticos de orden que son funciones de $X \sim N(0,1)$. Los valores exactos de m_i se tomaron de Parrish (1992).

Los nuevos términos del error satisfacen que,

$$E(\epsilon_i) = E(\sigma^{1/2}x_{(i)} - \sigma^{1/2}m_i) = \sigma^{1/2}m_i - \sigma^{1/2}m_i = 0, \quad i = 1, ..., n.$$

$$\begin{aligned} Cov(\epsilon_i, \epsilon_j) &= Cov(\sigma^{1/2}x_{(i)} - \sigma^{1/2}m_i, \sigma^{1/2}x_{(j)} - \sigma^{1/2}m_j) \\ &= E((\sigma^{1/2}x_{(i)} - E(\sigma^{1/2}x_{(i)}))(\sigma^{1/2}x_{(j)} - E(\sigma^{1/2}x_{(j)}))) \\ &= E((\sigma^{1/2}x_{(i)} - \sigma^{1/2}m_i)(\sigma^{1/2}x_{(j)} - \sigma^{1/2}m_j)) \\ &= E(\sigma x_{(i)}x_{(j)} - \sigma x_{(i)}m_j - \sigma m_i x_{(j)} + \sigma m_i m_j) \\ &= \sigma(E(x_{(i)}x_{(j)}) - m_j E(x_{(i)}) - m_i E(x_{(j)}) + m_i m_j) \\ &= \sigma(E(x_{(i)}x_{(j)}) - m_j m_i - m_i m_j + m_i m_j) \\ &= \sigma(E(x_{(i)}x_{(j)}) - m_j m_i) \\ &= \sigma Cov(x_{(i)}, x_{(j)}) \\ &= \sigma v_{ij}, \quad i, j = 1, ..., n. \end{aligned}$$

Para escribir a (2-18) en forma matricial, se introduce la siguiente notación vectorial: $\epsilon := (\epsilon_1, ..., \epsilon_n)'$ y $\mathbf{1} := (1, ..., 1)' \in \mathbb{R}^n$

Ahora se tiene,

$$\mathbf{y} = \mu\mathbf{1} + \sigma^{1/2}\boldsymbol{m} + \epsilon := \mathbf{PO} + \epsilon \tag{2-19}$$

donde $\mathbf{P} = (\mathbf{1} \;\; \boldsymbol{m}) \in \mathbb{R}^{n\times 2}$, $\mathbf{O'} = (\mu \;\; \sigma^{1/2}) \in \mathbb{R}^{1\times 2}$.

La matriz de covarianzas puede ser representada como,

$$\begin{aligned} Cov(\mathbf{y}) &= (\; E((y_{(i)} - E(y_{(i)}))(y_{(j)} - E(y_{(j)})))_{i,j} \\ &= (\; E((y_{(i)} - \mu - \sigma^{1/2}m_i)(y_{(j)} - \mu - \sigma^{1/2}m_j))) \;)_{i,j} \\ &= (\; E((\epsilon_i)(\epsilon_j)) \;)_{i,j} \\ &= (\; Cov(\epsilon_i, \epsilon_j) \;)_{i,j} \\ &= \sigma(\; Cov(x_{(i)}, x_{(j)}) \;)_{i,j} \\ &= \sigma\mathbf{V} \end{aligned} \tag{2-20}$$

Con $\mathbf{V} := (Cov(x_{(i)}, x_{(j)}))_{i,j=1,\dots,n} \in \mathbb{R}^{nxn}$, la matriz de covarianzas de los estadísticos de orden.

Para construir el estadístico de prueba de la prueba Shapiro-Wilk, se necesitan los mejores estimadores del modelo de regresión generalizado definido en (2-19).

Las mejores estimaciones lineales insesgadas (BLUEs) de μ y $\sigma^{1/2}$ son aquellas cantidades que minimizan la suma de cuadrados de los residuos (SCE) que corresponde a la parte de la variabilidad de la variable dependiente que no se consigue explicar con el modelo y está dada por $(\mathbf{y} - \mu\mathbf{1} - \sigma^{1/2}\boldsymbol{m})'\mathbf{V}^{-1}(\mathbf{y} - \mu\mathbf{1} - \sigma^{1/2}\boldsymbol{m})$ o $(\mathbf{y} - \mathbf{PO})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{PO})$.

Se considera la varianza generalizada dada por

$$\begin{aligned}\text{SCE} =&(\mathbf{y} - \mu\mathbf{1} - \sigma^{1/2}\boldsymbol{m})'\mathbf{V}^{-1}(\mathbf{y} - \mu\mathbf{1} - \sigma^{1/2}\boldsymbol{m})\\ =&\mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - \mu\mathbf{1}'\mathbf{V}^{-1}\mathbf{y} - \sigma^{1/2}\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{y} - \mu\mathbf{y}'\mathbf{V}^{-1}\mathbf{1} + \mu^2\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\\ &+ \mu\sigma^{1/2}\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{1} - \sigma^{1/2}\mathbf{y}'\mathbf{V}^{-1}\boldsymbol{m} + \mu\sigma^{1/2}\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m} + \sigma\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m}\\ =&\mathbf{y}'\mathbf{V}^{-1}\mathbf{y} + \mu^2\mathbf{1}'\mathbf{V}^{-1}\mathbf{1} + \sigma\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m} - 2\mu\mathbf{1}'\mathbf{V}^{-1}\mathbf{y} - 2\sigma^{1/2}\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{y} + 2\mu\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{1}.\end{aligned} \tag{2-21}$$

Para la agrupación de términos se tuvo en cuenta que $\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m}$ es un escalar y $\mathbf{V}^{-1}$ es una matriz simétrica positiva definida, en consecuencia, $\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m} = (\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m})' = \boldsymbol{m}'\mathbf{V}^{-1}\mathbf{1}$.

Ahora, minimizando la expresión de la varianza generalizada en (2-21) con respecto a μ y $\sigma^{1/2}$, se obtienen las ecuaciones:

$$\mu\mathbf{1}'\mathbf{V}^{-1}\mathbf{1} + \sigma^{1/2}\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{1} = \mathbf{1}'\mathbf{V}^{-1}\mathbf{y} \tag{2-22}$$

y

$$\mu\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{1} + \sigma^{1/2}\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m} = \boldsymbol{m}'\mathbf{V}^{-1}\mathbf{y} \tag{2-23}$$

Al resolver (2-22) y (2-23), se pueden conseguir los estimadores BLUEs de μ y $\sigma^{1/2}$, así,

$$\hat{\mu} = \frac{\boldsymbol{m}'\mathbf{V}^{-1}(\boldsymbol{m}\mathbf{1}' - \mathbf{1}\boldsymbol{m}')\mathbf{V}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m} - (\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m})^2} \tag{2-24}$$

$$\hat{\sigma}^{1/2} = \frac{\mathbf{1}'\mathbf{V}^{-1}(\mathbf{1}\boldsymbol{m}' - \boldsymbol{m}\mathbf{1}')\mathbf{V}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m} - (\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m})2} \tag{2-25}$$

Para distribuciones simétricas, en particular para la distribución normal, se tiene, $\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m} = \boldsymbol{m}'\mathbf{V}^{-1}\mathbf{1} = 0$ (Hain, 2010). Hain, mediante una demostración técnica llega a que $\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m} =$

$-\mathbf{1}'\mathbf{V}^{-1}\boldsymbol{m}$ y el único número que es igual que su opuesto es el cero. Por lo tanto se obtiene que los estimadores de μ y $\sigma^{1/2}$ se simplifican como sigue:

$$\hat{\mu} = \frac{\mathbf{1}'\mathbf{V}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{V}^{-1}\mathbf{1}}$$

$$\hat{\sigma}^{1/2} = \frac{\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{y}}{\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m}}$$

Se han definido los elementos necesarios para la conformación del estadístico de prueba denotado como W.

Definición 2.3.1 *Estadístico de prueba W*

Sea $y_{(1)}, ..., y_{(n)} \in \mathbb{R}$ una muestra de n observaciones aleatorias i.i.d ordenadas ascendentemente que provienen de la variable aleatoria Y con media desconocida $\mu \in \mathbb{R}$ y varianza desconocida $\sigma > 0$. Además, sea

$$S_n = \sum_{j=1}^{n} (y_{(j)} - \bar{y})^2$$

la estimación simétrica insesgada habitual de $(n-1)\sigma$.

El estadístico de prueba W para probar normalidad se define como:

$$W = \frac{(a'y)^2}{S_n} = \frac{(\sum_{j=1}^{n} a_j y_{(j)})^2}{\sum_{j=1}^{n}(y_{(j)} - \bar{y})^2}$$

donde

$$\mathbf{a}' = (a_1, ..., a_n) = \frac{\boldsymbol{m}'\mathbf{V}^{-1}}{(\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{V}^{-1}\boldsymbol{m})^{\frac{1}{2}}}$$

La definición del estadístico de prueba para la prueba de Shapiro-Wilk parece a primera vista un poco inesperada, especialmente las razones para el uso del vector **a** no son intuitivas.

El estadístico de prueba W puede interpretarse como una medida de linealidad y como el cociente de dos estimaciones diferentes de la desviación de una muestra.

- El estadístico W como una medida de linealidad:

  Bajo la hipótesis nula de normalidad

  $$E(y_{(i)}) = \mu + \sigma^{1/2} m_i \tag{2-26}$$

  y el estadístico W se puede expresar como

  $$W = \frac{\hat{\sigma}}{S_n} \tag{2-27}$$

Al analizar a (2-27), el numerador es la estimación de la varianza de la muestra aleatoria explicada por la regresión lineal en (2-26) y el denominador es una estimación de la varianza total. De esta manera, en la medida en que la linealidad sea más fuerte, entonces, la varianza explicada se acercará cada vez más a la varianza total, conformando así, una medida de linealidad.

- El estadístico W como la razón de dos estimaciones diferentes:

  Bajo la hipótesis nula de normalidad

  $$y_{(i)} = \mu + \sigma^{1/2} x_{(i)} \tag{2-28}$$

  el estadístico W se puede expresar como:

  $$\frac{\hat{\sigma}}{\frac{1}{n-1}S_n} = \frac{\left(\frac{\boldsymbol{m}'\mathbf{V}^{-1}\mathbf{y}}{\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m}}\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_{(i)} - \bar{y})^2} \tag{2-29}$$

  Como se observa en (2-29), el numerador representa la pendiente al cuadrado de la regresión lineal en (2-28) y el denominador representa la varianza de la muestra $y_{(1)}, ..., y_{(n)}$. Si $y_{(i)}$ se encuentra bajo la hipótesis nula, entonces la pendiente de la regresión lineal es una estimación de la desviación estándar $\sigma^{1/2}$, por tanto $\hat{\sigma}$ y $\frac{1}{n-1}S_n$ son estimaciones de la varianza y ambas deberían ser aproximadamente iguales.

Finalmente, respecto a la distribución de W bajo la hipótesis nula. Lamentablemente, según Shapiro y Wilk (1965), no hay posibilidad de dar una forma explícita de la distribución nula de W para tamaños de muestra mayores o iguales a 4. Shapiro y Wilk (1965) demostraron que existe una forma implícita para la distribución de W, para más detalles de esta prueba hay que referirse a la obra original. Por tal motivo, para obtener más información sobre la distribución y sus percentiles necesarios para probar la normalidad, se deben considerar métodos de simulación empírica.

2.3.2. Prueba Shapiro-Wilk Generalizada

En esta subsección se hace una descripción de la prueba Shapiro Wilk Generalizada expuesta originalmente por Villaseñor y González (2009), quienes proponen utilizar la teoría de la prueba Shapiro Wilk univariada con la intención de obtener una prueba de multinormalidad que herede la potencia de esta.

Se plantean las siguientes hipótesis para una muestra $\mathbf{Y_1}, ..., \mathbf{Y_n}$ en $\mathbb{R}^p$, tal que:

$H_0 = \mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ provienen de una población multinormalmente distribuida $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$H_1 = \mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ no provienen de una población multinormalmente distribuida $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Se considera a $\mathbf{Y_1}, ..., \mathbf{Y_n}$ como vectores aleatorios i.i.d de una población p-dimensional y se denota a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ como la densidad normal p-variada con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$ expresada en (2-1).

La siguiente proposición es un resultado muy importante que caracteriza a la distribución normal multivariada.

Proposición 2.3.1 $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *si y solo si* $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I})$.

A continuación se definen los elementos necesarios para la construcción del estadístico de prueba.

La media muestral $\bar{\mathbf{Y}}$ y la matriz de covarianzas muestral $\mathbf{S}$, están dadas por

$$\bar{\mathbf{Y}} = n^{-1} \sum_{j=1}^{n} \mathbf{Y_j} \qquad \mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{Y_j} - \bar{\mathbf{Y}})(\mathbf{Y_j} - \bar{\mathbf{Y}})'$$

Sea $\mathbf{S^{-1/2}}$ la raíz cuadrada definida positiva simétrica de $\mathbf{S^{-1}}$ (inversa de la matriz de $\mathbf{S}$). Ahora, considerando que las variables aleatorias $\mathbf{Y_1}, ..., \mathbf{Y_p}$ en $\mathbb{R}^n$, no necesariamente son independientes, lo cual implica que el cálculo de $\mathbf{S^{-1/2}}$ requiere un trato diferente donde se usa la descomposición de Schur. Los detalles del cálculo de $\mathbf{S^{-1/2}}$ se encuentran en la propiedad 5 de la subsección 2.1.

Cabe recordar que los elementos de la muestra aleatoria $\mathbf{Y_1}, ..., \mathbf{Y_n}$ en $\mathbb{R}^p$ son independientes e idénticamente distribuidos, en cambio, las variables aleatorias $\mathbf{Y_1}, ..., \mathbf{Y_p}$ en $\mathbb{R}^n$ no necesariamente son independientes, de hecho, en las investigaciones es de interés estudiar sus relaciones.

Explícitamente, cuando $\mathbf{Y_1}, ..., \mathbf{Y_n}$ tienen una distribución $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, los vectores aleatorios estandarizados

$$\mathbf{Z_j^*} = \mathbf{S^{-1/2}}(\mathbf{Y_j} - \bar{\mathbf{Y}})' \qquad j = 1, ..., n \tag{2-30}$$

tienen una distribución cercana a una $N_p(\mathbf{0}, \mathbf{I})$, lo que significa que las coordenadas de $\mathbf{Z_j^*}$, denotadas por $\mathbf{Z_{1j}}$,...,$\mathbf{Z_{pj}}$ son independientes con distribución normal estándar univariada, esto se deriva de la propiedad que indica que la distribución marginal de cualquier conjunto de componentes de una variable normal multivariada es también normal, en este caso, normal estándar univariada.

Definición 2.3.2 *El estadístico de prueba W^* se puede expresar como la siguiente combinación lineal*

$$W^* = \frac{1}{p}\sum_{i=1}^{p} W_{Z_i}, \tag{2-31}$$

donde W_{Z_i} es el estadístico de Shapiro-Wilk evaluado en la i-ésima coordenada de las observaciones transformadas $\mathbf{Z_{i1}}$,...,$\mathbf{Z_{in}}$, $i = 1, ..., p$. Como ilustración se considera el ejemplo para $i = 1$ donde las observaciones transformadas corresponden a $\mathbf{Z_{11}}$,...,$\mathbf{Z_{1n}}$, es decir, la primera variable estandarizada y evaluada en todos los individuos, y así para cada variable estandarizada perteneciente a la matriz original $\mathbf{Y}$. En consecuencia, bajo la hipótesis nula de multinormalidad, el estadístico W_{Z_i} calculado para cada variable, debería ser cercano a uno, luego, la suma de todos los estadísticos univariados debería ser cercana a p, y dado que W^* tiene la forma de promedio, entonces el resultado final debería ser cercano a 1.

A simple vista, no es inmediato ver la influencia de la matriz de covarianzas muestral en la construcción del estadístico de prueba W^* puesto que, se construyen p estadísticos de prueba univariados en los cuales se aproxima $\mathbf{V}$ (definida en (2-20) de la subsección 2.3.1) que depende de los estadísticos de orden asociados a la variable en cuestión, no obstante, para emplear la prueba univariada en cada variable que compone a la matriz $\mathbf{Y}$, es necesario estandarizar la matriz de datos originales $\mathbf{Y}$ y para la estandarización se utiliza la raíz de la inversa de la matriz de covarianzas muestral denotada por $\mathbf{S^{-1/2}}$, tal como se presenta en (2-30).

Finalmente, la prueba basada en W^* rechaza la hipótesis nula con un nivel de significancia α si $W^* < c_{\alpha;n,p}$, donde $c_{\alpha;n,p}$ satisface lo siguiente:

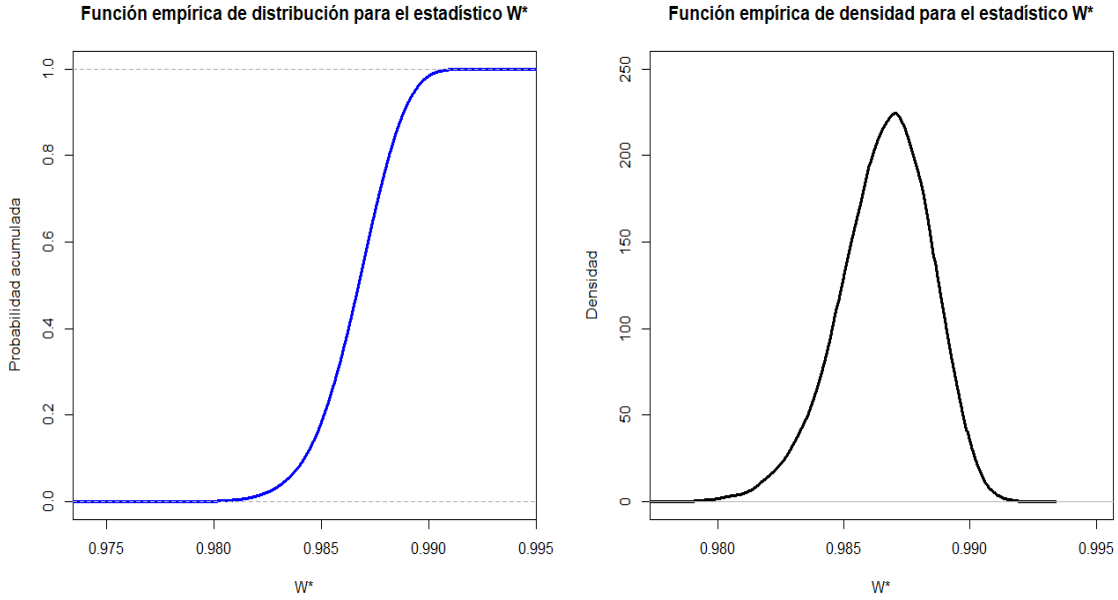
$\alpha = P\{W^* < c_{\alpha;n,p} | H_0 \textit{ es cierta}\}$

Definición 2.3.3 *Distribución nula del estadístico de prueba*

La distribución nula del estadístico W^* no se puede obtener analíticamente ya que es una función del estadístico W de la Shapiro-Wilk univariada y, hasta el momento, no se ha obtenido ninguna expresión analítica para la función de distribución de W. Por lo tanto, haciendo uso de la simulación por el algoritmo Monte Carlo, se obtienen los percentiles de $c_{\alpha;n,p}$ considerando que la función de distribución del estadístico W^* denotada por F_{W^*}, no depende de los parámetros desconocidos debido a que W^* es una función de las observaciones transformadas por el vector de medias muestral y la matriz de covarianza muestral. Así, bajo la hipótesis nula de multinormalidad, se asume la proposición 2.3.1 para n y p dados. Luego, se simulan J muestras aleatorias de tamaño n de la distribución $N_p(\mathbf{0}, \mathrm{I})$ y se calcula el respectivo valor de estadístico W^* para cada muestra. Ahora bien, usando los J valores de W^*, se estima a F_{W^*} con la función de distribución empírica denotada por $\tilde{F}_{W^*,n}$. Finalmente, $c_{\alpha;n,p}$ es tal que para un valor de $\alpha \in (0, 1)$, entonces,

$\alpha = F_{W^*}(c_{\alpha;n,p})$.

La Figura **2-3** a presenta la función de distribución empírica $\tilde{F}_{W^*,n}$ y la Tabla **2-1** presenta los percentiles empíricos asociados, ambos determinados con $n = 100$, $p = 12$ y $J = 50{,}000$.



(a) Función empírica de distribución (b) Función empírica de densidad

Figura 2-3: Función empírica de distribución y de densidad para el estadístico W^* especificada por $n = 100$, $p = 12$ y $J = 50{,}000$

5 %	10 %	15 %	20 %	25 %	30 %	35 %
0.9833769	0.9842159	0.9847410	0.9851348	0.9854630	0.9857640	0.9860240
40 %	45 %	50 %	55 %	60 %	65 %	70 %
0.9862750	0.9865120	0.9867410	0.9869660	0.9871860	0.9874130	0.9876510
75 %	80 %	85 %	90 %	95 %	100 %	
0.9879030	0.9881690	0.9884720	0.9888520	0.9893791	0.9928490	

Tabla 2-1: Percentiles empíricos con $n = 100$, $p = 12$ y $J = 50{,}000$

Analizando la forma de la función empírica de densidad del estadístico (ver Figura **2-3** (b)), se observa que no sigue una distribución normal ya que el estadístico W^* solo toma valores positivos, en consecuencia, Villaseñor y González (2009) mencionan que, una aproximación se puede dar mediante la distribución lognormal haciendo $W_1^* = log(1 - W^*)$. Los detalles de la aproximación se pueden encontrar en Villaseñor y González (2009).

2.3.3. Simulación Monte Carlo

Los métodos de Monte Carlo son algoritmos computacionales que están fundamentados en el muestreo aleatorio repetido para obtener resultados numéricos. La esencia de estos métodos consiste en usar la aleatoriedad para resolver problemas que podrían ser experimentos o fenómenos que dan lugar a cierto resultado que de antemano ya se conoce. En general, los métodos de Monte Carlo suelen ser empleados para resolver cualquier problema siempre y cuando se cuente con una interpretación probabilística.

A pesar de su simplicidad conceptual y algorítmica, el costo computacional asociado con una simulación de Monte Carlo puede ser asombrosamente alto, ya que el método requiere muchas muestras para obtener una buena aproximación, lo que puede incurrir en un tiempo de ejecución total arbitrariamente grande si el tiempo de procesamiento de una sola muestra es alto. (Mazhdrakov, Benov, y Valkanov, 2018)

Los métodos de Monte Carlo varían, pero tienden a seguir un patrón particular:

1. Definir un dominio de posibles entradas
2. Generar entradas de forma aleatoria a partir de una distribución de probabilidad sobre el dominio.
3. Realizar un cálculo determinista en las entradas.
4. Agregar los resultados

En este estudio se utilizará la simulación Monte Carlo para:

1. Obtener la distribución nula empírica del estadístico de la prueba modificada.
2. Evaluar el desempeño de la prueba Shapiro Wilk Generalizada y una versión modificada de la prueba Shapiro Wilk generalizada.

3 Metodología

En esta sección se presenta la metodología planteada para dar respuesta a la pregunta de investigación y los objetivos propuestos. Inicialmente se tomará la estimación óptima de la matriz de covarianzas shrinkage y luego se incorporará en el estadístico de la prueba Shapiro-Wilk Generalizada reemplazando a la estimación tradicional. Posteriormente se estudiarán las características de la prueba Shapiro-Wilk modificada. Ahora bien, dada la modificación propuesta se hace necesaria la evaluación de su desempeño, por lo que se generarán datos provenientes de poblaciones planeadas estratégicamente que conducirán junto a otros elementos a los escenarios de simulación. La Figura **3-1** ilustra el resumen de las actividades realizadas por fases.

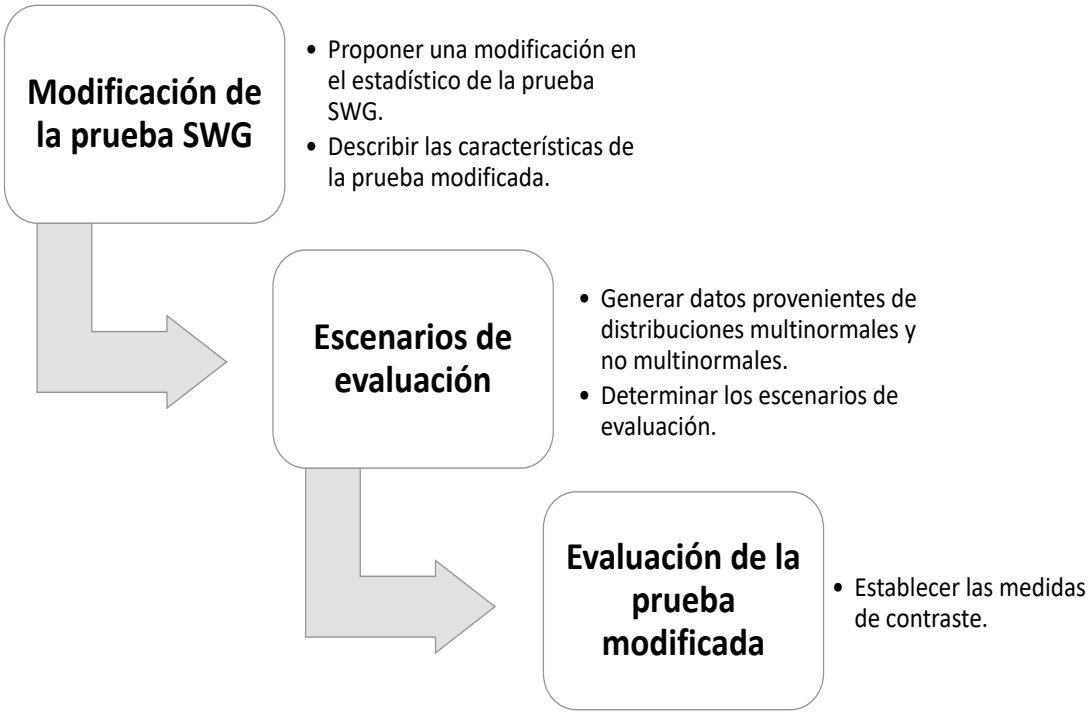


Figura 3-1: Flujo metodológico

3.1. Modificación de la prueba Shapiro-Wilk Generalizada y características

En esta sección serán descritas las características de la modificación de la prueba Shapiro-Wilk Generalizada propuesta por Villaseñor y González (2009) y descrita en la subsección 2.3.2 de este trabajo. La modificación se realiza con la finalidad de verificar el supuesto de multinormalidad para los casos en que el número de individuos n es aproximadamente igual al número de variables p y para ello se reemplaza a la estimación de la matriz de covarianzas tradicional $\mathbf{S}$ por la estimación shrinkage $\mathbf{S}^*$ propuesta por Schäfer y Strimmer (2005).

Sea $\mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ una muestra de vectores aleatorios i.i.d de una población p-dimensional. Se desea probar que los vectores aleatorios siguen una distribución normal p-variada con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$, para ello se plantean las siguientes hipótesis:

$H_0 = \mathbf{Y_1}, ..., \mathbf{Y_n}$ provienen de una población multinormalmente distribuida $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$H_1 = \mathbf{Y_1}, ..., \mathbf{Y_n}$ no provienen de una población multinormalmente distribuida $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

donde $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota la función de densidad expresada en (2-1).

Teniendo en cuenta la propiedad 5 enunciada en 2.1.1, la variable aleatoria $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ si y solo si $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{Y} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I})$.

Ahora, se denota a $\bar{\mathbf{Y}}$ como la media muestral y a la estimación shrinkage de la matriz de covarianzas muestral como $\mathbf{S}^*$, en fórmulas:

$$\bar{\mathbf{Y}} = n^{-1} \sum_{j=1}^{n} \mathbf{Y_j} \qquad \mathbf{S}^* = \hat{\lambda}^M \mathbf{T} + (1 - \hat{\lambda}^M)\mathbf{S}$$

donde $\mathbf{T}$ es la matriz objetivo, $\hat{\lambda}^M$ es un número entre [0, 1] que corresponde a una estimación óptima obtenida analíticamente luego de establecer una matriz objetivo $\mathbf{T}$ y $\mathbf{S}$ es la matriz de covarianzas muestral tradicional. La elección de $\mathbf{T}$ depende del investigador y el experto en el tema de investigación, los mencionados deben realizar una evaluación consciente generando hipótesis previas basadas en criterios sobre el comportamiento de los datos disponibles. Sin embargo, el investigador se debe asegurar que al operar con la matriz objetivo dada, la estimación shrinkage sea positiva definida.

Los autores Schäfer y Strimmer (2005) garantizan que $\mathbf{S}^*$ (calculándola con el estimador óptimo y regulado de la intensidad de contracción y especificando la matriz objetivo) es simétrica positiva definida, por lo que la raíz cuadrada de la inversa de $\mathbf{S}^*$ también es simétrica positiva definida y se denota como $\mathbf{S}^{*-1/2}$. Para calcular $\mathbf{S}^{*-1/2}$ se utiliza la descomposición de Schur expuesta en la propiedad 5 de la subsección 2.1.

Cuando los vectores aleatorios $\mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^p$ siguen una distribución $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces los vectores aleatorios estandarizados usando la estimación shrinkage de la matriz de covarianzas se definen como

$$\mathbf{Z_j^M} = \mathbf{S^{*-1/2}}(\mathbf{Y_j} - \bar{\mathbf{Y}})'$$

para cada $j = 1, ..., n$. Además, asumiendo que p es fijo; cuando n es pequeño, la estimación shrinkage de la matriz de covarianzas es un mejor estimador que la matriz de covarianzas tradicional y cuando n es grande, ambas estimaciones son similares. Por lo tanto, se presume que cada vector estandarizado sigue aproximadamente un distribución normal p-variada estándar, así

$$\mathbf{Z_j^M} \sim N_p(\mathbf{0}, \mathbf{I})$$

Por la propiedad 1 enunciada en la subsección 2.1, las coordenadas $\mathbf{Z_j^M}$ denotadas por $\mathbf{Z^*_{1j}}$,...,$\mathbf{Z^*_{pj}}$ son independientes con distribución normal estándar univariada.

La definición del estadístico de prueba modificado sigue la misma estructura del estadístico propuesto por Villaseñor y González (2009), el cual es presentado en (2-31).

Definición 3.1.1 *Estadístico de la prueba modificada*

$$W^M = \frac{1}{p}\sum_{i=1}^{p} W^*_{Z_i},$$

donde $W^*_{Z_i}$ es el estadístico Shapiro-Wilk univariado evaluado en cada coordenada de las observaciones transformadas $\mathbf{Z^*_{i1}}$,...,$\mathbf{Z^*_{in}}$, $i = 1, ..., p$. Bajo la hipótesis nula de multinormalidad, $W^*_{Z_i}$ debería ser cercano a uno para cada variable y la suma de todos ellos debería ser cercana a p. Finalmente, el estadístico W^M será el promedio de los resultados de $W^*_{Z_i}$ y dicho promedio debería ser cercano a uno.

Al momento de tomar la decisión de rechazar H_0 o no, se utiliza el percentil $c_{\alpha;n,p}$ obtenido para la prueba Shapiro Wilk Generalizada debido a lo observado en la Figura **3-2** donde las distribuciones nulas empíricas para ambos estadísticos de prueba están superpuestas. Sin embargo, un trabajo futuro podría estudiar si los percentiles de ambos estadísticos (W^* y W^M) presentan diferencias significativas que produzcan decisiones contrarias sobre la población.

Valores empíricos del estadístico de prueba

Siguiendo la idea para el cálculo de las funciones empíricas de distribución y de densidad del estadístico de prueba bajo la hipótesis nula de multinormalidad, expuesta por Villaseñor

y González (2009), descrita en el presente trabajo en la subsección 2.3.2, se presentan las funciones empíricas de densidad asociadas a los valores de W^* y W^M de muestras simuladas calculadas con la siguiente metodología:

1. Se calculan $J = 50{,}000$ muestras y se analizan para dos tamaños $n = 100$ y $n = 13$. Las muestras son simuladas a partir de una población multinormalmente distribuida especificadas por $p = 12$, $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma}$ una matriz (12×12) bien condicionada con varianzas desiguales y correlaciones asociadas positivas, negativas y nulas. Luego, se calcula el estadístico W^* y W^M a cada muestra simulada y mediante el cociente entre las frecuencias relativas y el ancho de los intervalos se construye una aproximación a la función de densidad a través de la función empírica de densidad que se puede observar en la Figura **3-2**.

2. Se calculan $J = 50{,}000$ muestras y se analizan para dos tamaños $n = 100$ y $n = 13$. Las muestras son simuladas a partir de una población no multinormalmente distribuida (simulada con el método propuesto por Qu, Liu, y Zhang (2019)) especificadas por Curtosis=180, Asimetría=3, $p = 12$, $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma}$ una matriz (12×12) bien condicionada con varianzas desiguales y correlaciones asociadas positivas, negativas y nulas. Luego, se calcula el estadístico W^* y W^M a cada muestra simulada y mediante el cociente entre las frecuencias relativas y el ancho de los intervalos se construye una aproximación a la función de densidad a través de la función empírica de densidad que se puede observar en la Figura **3-3**.

En la Figura **3-2** se observa que en las muestras de tamaño $n = 100$ los estadísticos tienen un comportamiento similar en todo su dominio, lo cual es coherente con el planteamiento de la estimación shrinkage puesto que, cuando la estimación tradicional es de calidad, entonces no es necesario aplicar una contracción fuerte, por lo tanto, ambas estimaciones deberían ser similares y producir ambos estadísticos similares. No obstante, para las muestras que tienen muy poca información ($n = 13$) para explicar tantas variables ($p = 12$), se presumía una afectación en los valores del estadístico de prueba ya que este se calcula a partir de las observaciones trasformadas por la matriz de covarianza muestral, la cual estaría mal estimada por la precaria información disponible en la muestra y que dicha dificultad se subsanaría al cambiar la estimación por una con mejor desempeño para muestra pequeñas, sin embargo, bajo normalidad no parece tener un efecto en los valores del estadístico de prueba ya que ambas curvas están superpuestas. Por último, contrastando las funciones empíricas de densidad nula con un tamaño de muestra igual a 100 versus un tamaño de muestra igual a 13, se aprecia una notable diferencia en media y dispersión de los valores de ambos estadísticos, específicamente, los valores de los estadísticos con muestras pequeñas se alejan un poco más de uno indicando quizás un riesgo mayor de rechazar la hipótesis nula de normalidad dado que es verdadera, pero la incorporación de la estimación shrinkage no afecta el riesgo de tomar la decisión incorrecta.

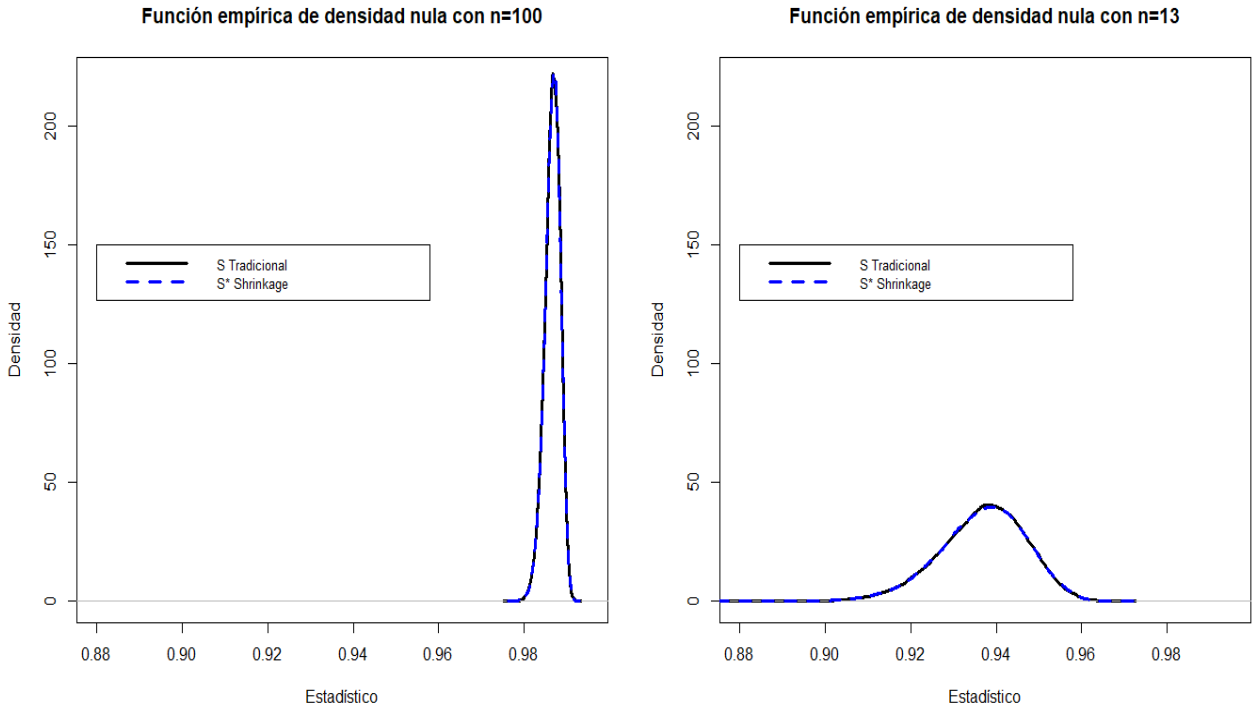


Figura 3-2: Funciones de densidad empíricas para W^* y W^M con $J = 50{,}000$ muestras de $n = 100, 13$ y $p = 12$ para datos multinormalmente distribuidos.

En la Figura **3-3**, bajo no multinormalidad, por un lado, se observa que en las muestras de tamaño $n = 100$ el estadístico modificado está tomando valores levemente menores en comparación con el estadístico tradicional, pero de forma general tienen comportamiento muy similar. Por otro lado, para las muestras que tienen muy poca información ($n = 13$) para explicar tantas variables ($p = 12$), se evidencia una diferencia en media de los valores del estadístico de prueba para la estimación shrinkage comparada con la tradicional, en este caso, la estimación shrinkage toma valores más alejado de uno que la estimación tradicional, lo que podría indicar decisiones con menos incertidumbre respecto a rechazar la hipótesis nula de normalidad dado que es falsa, conllevando a un mejor desempeño de la prueba cuando se dispone de un tamaño de muestra pequeño. Por último, contrastando las funciones empíricas de densidad nula con un tamaño de muestra igual a 100 versus un tamaño de muestra igual a 13, se aprecia una diferencia en media y dispersión de los valores de ambos estadísticos, específicamente, los valores de ambos estadísticos con muestras pequeñas se alejan un poco más de uno indicando quizás menos riesgo de no rechazar la hipótesis nula dado que es falsa ya que su máximo valor es un poco mayor a 0.96.

Para ambas muestras provenientes de dos poblaciones diferentes (multinormal y no mul-

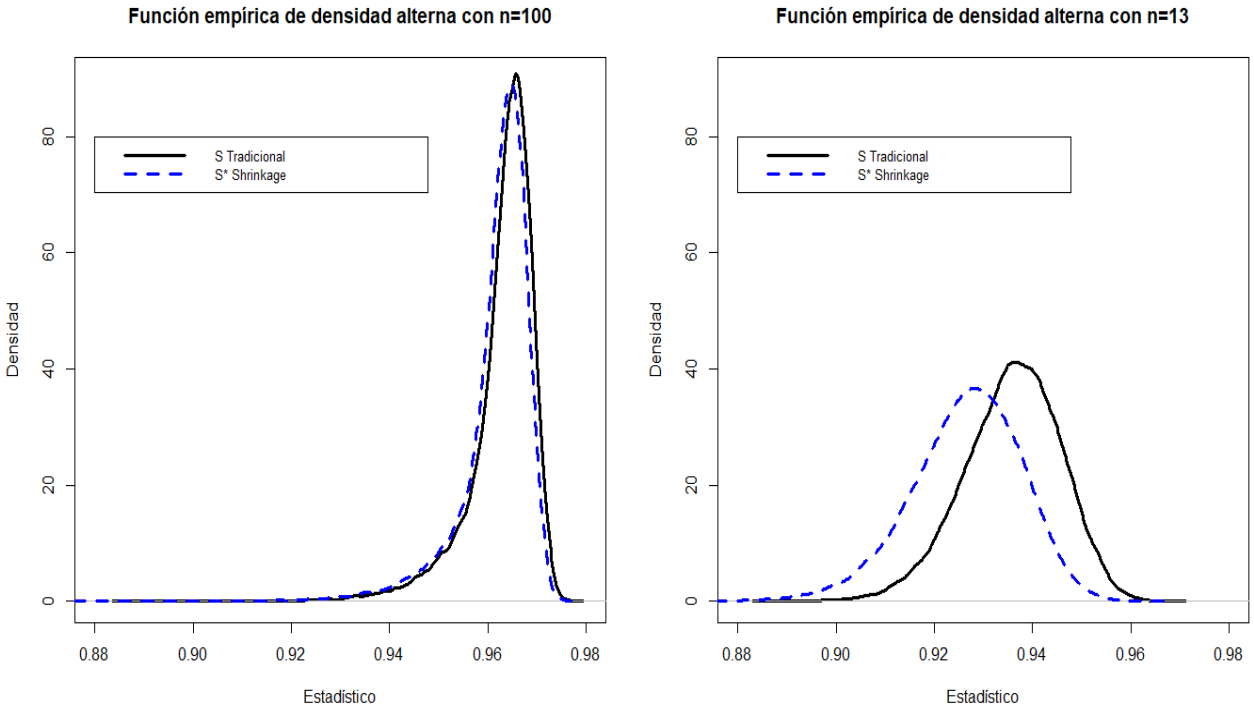


Figura 3-3: Funciones de densidad empíricas para W^* y W^M con $J = 50{,}000$ muestras de $n = 100, 13$ y $p = 12$ para datos no multinormalmente distribuidos, es decir, con Curtosis=180 y Asimetría=3

tinormal), se observa que cuando la información disponible es poca, entonces los valores de los estadísticos están menos concentrados en un punto, es decir que pueden ser más inestables respecto a la decisión.

3.2. Escenarios de evaluación

Los escenarios determinados son la base para la evaluación de la prueba tradicional y modificada. En esta subsección se expone de manera detallada cada uno de los elementos necesarios para la generación de datos tales como: tamaño de muestra, cantidad de variables, distribuciones consideradas, parámetros de las distribuciones y estimadores de los parámetros.

En la Tabla **3-1** se proponen escenarios para la generación de datos que provienen de distribuciones multinormales y no multinormales (t multivariante; $\chi^2(1)$ desplazada con marginales i.i.d.; $\chi^2(1) + nor$ y Qu-Liu-Zhang, generada con asimétria y curtosis multivariada deseada usando el método propuesto por Qu y cols.(2019)) cuyas distribuciones de probabilidad se definen en la subsección 3.2.6. $\boldsymbol{\Sigma}$ es la matriz de covarianzas poblacional y $\boldsymbol{\mu}$ es el vector de medias poblacional que conducirán la respectiva simulación. Respecto a los parámetros

para la generación de los datos que siguen distribuciones multinormales y no multinormales, se fija en todos los escenarios a $\boldsymbol{\mu} = \mathbf{0}$ y para la $\boldsymbol{\Sigma}$ hay dos propuestas, la primera es una matriz de covarianzas bien condicionada para la cual se desarrolló un algoritmo donde se define una estructura deseada, en este caso se restringe a que la matriz resultante tenga asociadas correlaciones moderadas, valores propios en un rango reducido y desviaciones estándar diferentes entre variables, por ello, en la Tabla **3-1** se formula cómo se restringe a la matriz de covarianzas de acuerdo a los criterios mencionados y en esta misma sección pero más adelante, se describe al algoritmo desarrollado. La segunda es una matriz de covarianzas mal condicionada. Luego, se procede con el cálculo de la estimación shrinkage de la matriz de covarianzas para cada pareja $(n, p = 12)$ que se construye en tres pasos; primero, contrayendo las correlaciones; segundo, contrayendo las varianzas y tercero, combinando ambas contracciones mediante la relación $s^*_{ij} = r^*_{ij} s^{*1/2}_{ii} s^{*1/2}_{jj}$, donde r^*_{ij} son las correlaciones muestrales contraídas y s^*_{ii} (s^*_{jj}) son las varianzas muestrales contraídas. La idea de contracción es la misma para cualquier estimador, es decir que se requiere un objetivo, una intensidad de contracción y un estimador muestral. La Tabla **3-2** presenta los elementos requeridos para la estimación. La matriz **S** corresponde a la estimación tradicional de la matriz de covarianzas, **T**.**var** es el vector de la varianza objetivo que permite contraer las varianzas, **T**.**cor** es la matriz objetivo que permite contraer las correlaciones, $\hat{\lambda}^M$.var es el estimador de la intensidad de contracción para las varianzas y $\hat{\lambda}^M$.cor es el estimador de la intensidad de contracción para las correlaciones.

Tabla 3-1: Generación de datos

(n,p)	Dist. de datos	Valor de parámetros	
		$\boldsymbol{\Sigma}$	$\boldsymbol{\mu}$
$(n = i, p = j)$, $i = 13, 14, ..., 60$ $j = 12$	Multinormal	Correlaciones para las variables $\begin{cases} \rho_{ij} \sim \text{U}(0.15, 0.3) & \text{si} \quad 1 \leq i, j \leq 9 \\ \rho_{ij} \sim \text{U}(0.3, 0.5) & \text{si} \quad 10 \leq i, j \leq 12 \end{cases}$ Valores propios γ_{i+1} $\gamma_{i+1} \sim \text{U}(10 - 0.8i - 0.2, 10 - 0.8i + 0.2)$ si $0 \leq i \leq 11$ Desviaciones estándar $\sigma_{ii}^{1/2}$ $\sigma_{ii}^{1/2} \sim \text{U}(\sqrt{12 - (i-1)} + 0.5, \sqrt{12 - (i-1)} + 0.5)$ si $1 \leq i \leq 12$	$\mathbf{0}$
	No Multinormal		
	Multinormal	σ_{ij}; $\sigma_{ii} = 1$ y $\sigma_{ij} = 0.9$	
	No Multinormal		

A continuación se argumenta cada elemento necesario para la generación de datos y los escenarios de evaluación.

Tabla 3-2: Escenarios de evaluación

(n,p)	Dist. de datos	Intensidad de contracción	Objetivo	Est. Shrinkage
$(n = i, p = j)$, $i = 13, 14, ..., 60$ $j = 12$ g indica la variable	Multinormal	$\hat{\lambda}^M$.var= $\text{mín}\,(1, \frac{\sum_{g=1}^{p} \hat{\text{Var}}(s_g)}{\sum_{g=1}^{p}(s_g - s_{\text{mediana}})^2})$	$\mathbf{T.var} =$ $s_{\text{mediana}}\mathbf{1'}$	$\mathbf{R^*} = [\ \hat{\lambda}^M.\text{Cor}\ (\mathbf{T.cor}) + (1 - \hat{\lambda}^M.\text{Cor})\mathbf{R}\ ]_{p\times p}$ $\mathbf{K^{*1/2}} = [\ \{diag(\hat{\lambda}^M.\text{Var}(\mathbf{T.var}) + (1 - \hat{\lambda}^M.\text{Var})\hat{\mathbf{V}})\}^{1/2}\ ]_{p\times p}$ $\mathbf{S^*} = \mathbf{R^*K^{*1/2}K^{*1/2}}$
	No Multinormal	$\hat{\lambda}^M$.cor= $\text{mín}\,(1, \frac{\sum_{i\neq j} \hat{\text{Var}}(r_{ij})}{\sum_{i\neq j} r_{ij}^2})$	$\mathbf{T.cor} =$ $\mathbf{I}$	

3.2.1. Número de observaciones y número de variables (n, p)

Los escenarios de simulación se centran principalmente en el caso de alta dimensión con tamaño de muestra pequeño. La cantidad de variables (p) permanece fija mientras que la cantidad de observaciones (n) va incrementando, esto con la finalidad de contrastar el funcionamiento de la prueba cuando la matriz de covarianzas muestral mejora su calidad como estimador ya que n rebasa cada vez más a p.

El caso más crítico es cuando $p = 12$ y $n = 13$ puesto que se acentúa más el sesgo de los valores propios asociados a la estimación tradicional de la matriz de covarianzas con respecto a la matriz de covarianzas poblacional.

Como medida preventiva para garantizar que los resultados sean válidos y reproducibles, se debe mantener un balance entre el número de observaciones y número de variables. Además, es importante especificar el valor de la semilla para conducir la simulación usando el método de Monte Carlo.

3.2.2. Matriz de covarianzas poblacional Σ y vector de medias poblacional μ

De forma general, se desea evaluar vía simulación el desempeño de la modificación de la prueba Shapiro-Wilk Generalizada, para ello se definen los parámetros poblacionales que conducen la simulación de múltiples muestras pseudoaleatorias provenientes de distribuciones multinormales y no multinormales.

Teniendo en cuenta la explicación dada en la subsección 2.2 sobre la necesidad de un estimador mejor condicionado que la estimación tradicional de la matriz de covarianzas, se plantean dos escenarios de la matriz de covarianza poblacional, una, bien condicionada y otra, mal condicionada; esto con el fin de corroborar si efectivamente la estimación shrinkage es mejor que la estimación tradicional y averiguar si esto tiene un efecto positivo en el desempeño de la prueba modificada.

En principio, se pensó que dado que la modificación de la prueba es específicamente en el cambio de estimación para obtener la matriz de covarianzas muestral bajo el escenario de muestras pequeñas, era necesario asegurarse de que el mal condicionamiento de la matriz de covarianzas muestral es consecuencia de información insuficiente en la muestra y no por un parámetro mal condicionado que produjese matrices de covarianzas muestrales mal condicionadas aún cuando el tamaño de muestra sea considerablemente suficiente para producir una buena estimación. No obstante, problemas reales sugieren relaciones fuertes positivas y negativas entre variables, varianzas marginales desiguales, que por su naturaleza, sugieren un parámetro mal condicionado. En este último escenario, aún cuando la muestra sea grande, los estimadores tendrá asociados números de condición grandes ya que en realidad se parecerá a su parámetro que originalmente está mal condicionado.

A continuación se detalla el proceso de definición de cada parámetro.

1. Matriz de covarianzas poblacional con número de condición pequeño o "bien condicionada"

   Para la definición de la matriz de covarianzas poblacional se intenta aproximar a una matriz de un problema real que posee varianzas desiguales y tiene asociadas correlaciones positivas, negativas y nulas.

   Con ideas complementarias se desarrolló un código para este trabajo de grado en el software Python (Python, 1991), para la generación de matrices de covarianzas condicionada a una estructura deseada siguiendo la idea expuesta por Arteaga y Ferrer (2013).

   En primer lugar, se debe ingresar las k desviaciones denotadas como $\{\sigma_{ii}^{1/2}\}$ donde $k \leq p$; en segundo lugar, teniendo en cuenta las variables asociadas a las desviaciones ingresadas, se establecen las correlaciones de entrada $\{\rho_{ij}\}$; en tercer lugar, se dan los h valores propios deseados $\{\gamma_i\}$ con $h \leq p$ que posteriormente se reemplazan en la matriz $\mathbf{L}$ diagonal de valores propios de una matriz simétrica arbitraria $\mathbf{F}$ dada. En el diagrama de la Figura **3-4** se expone el algoritmo propuesto.

   Al implementar el código que sigue la estructura del diagrama de la Figura **3-4**, se evidencia que la matriz converge fácilmente a las varianzas y correlaciones ingresadas, pero no a los valores propios deseados. Por ello se flexibilizaron los valores de entrada, no obstante, tantear una solución ingresando diferentes valores representaba un gasto considerable de tiempo, por lo que se implementó un código complementario (Ver Figura **3-5**) encargado de la generación aleatoria de los valores de entrada en un rango valores deseado para las desviaciones, correlaciones y valores propios de la siguiente forma:

   a) Cada valor propio se escoge siguiendo una distribución uniforme centrada en $0{,}8i$ y un margen de 0.2, es decir, $\gamma_{i+1} \sim \mathrm{U}(10 - 0.8i - 0.2, 10 - 0.8i + 0.2)$.

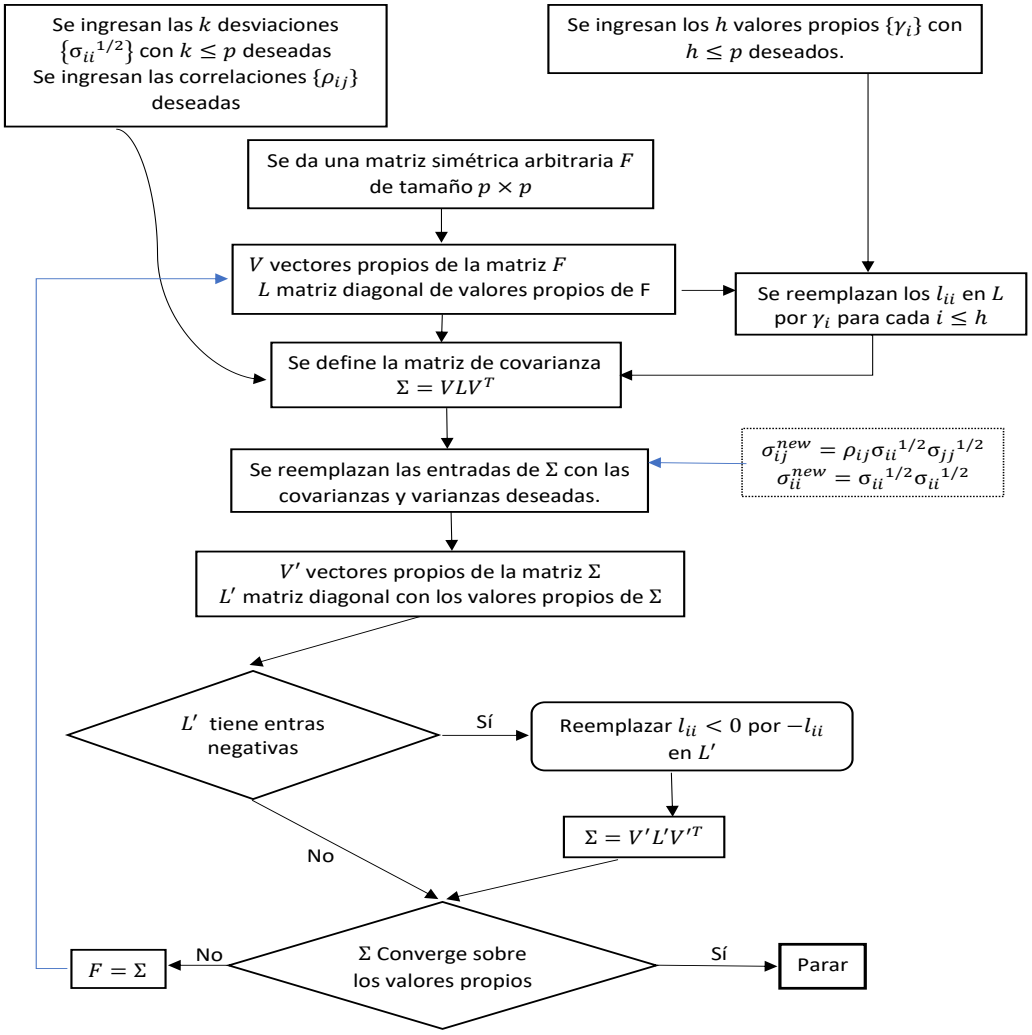


Figura 3-4: Diagrama para generación de matriz de covarianzas con estructura deseada. Fuente: Elaborado por el matemático Jarvi A. Rodriguez

b) Se establece una correlación distribuida uniformemente entre 0.15 y 0.3 para las variables 1 a 9, luego, se establece la correlación uniforme entre 0.3 y 0.5 para las variables 10 a 12 y el resto de correlaciones se dejan libres.

c) Se escogen las desviaciones considerando la distribución uniforme con centro en $\sqrt{12-(j-1)}$ y un margen de 0.5, es decir $\sigma_{ii}^{1/2} \sim \mathrm{U}(\sqrt{12-(j-1)}+0.5, \sqrt{12-(j-1)}-0.5)$.

El objetivo de los pasos mencionados es generar muchos valores de entrada que se evaluarán N veces en el código que genera la matriz de covarianzas deseada (ver Figura **3-4**). A este algoritmo se le llama algoritmo de flexibilización para la generación de la

matriz $\mathbf{\Sigma}$ poblacional (Ver Figura **3-5**). El condicional del algoritmo para la selección de la matriz $\mathbf{\Sigma}$ poblacional se establece sobre los valores propios; si la diferencia entre el mayor valor propio y el menor valor propio es pequeña se imprime por pantalla la matriz $\mathbf{\Sigma}$.

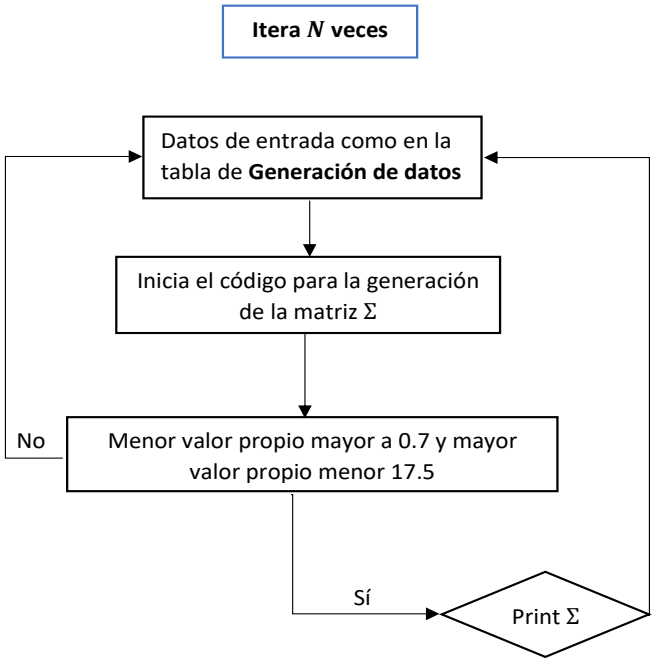


Figura 3-5: Algoritmo de flexibilización para la generación de la matriz $\mathbf{\Sigma}$ poblacional
Fuente: Elaborado por el matemático Jarvi A. Rodriguez

En la Tabla **3-3** se presenta la matriz de covarianzas hallada. En la diagonal se puede apreciar que las varianzas son desiguales y van de aproximadamente 13 a 2. Por facilidad en la interpretación se presenta en la Tabla **3-4** a la matriz de correlación poblacional, en ella, se evidencia que hay presencia de correlaciones (entre las 12 variables) moderadas que son positivas, negativas y nulas. En particular, para un contexto horario se podría pensar que esta matriz corresponde a medidas que tienen un comportamiento relacionado en horas cercanas al medio día. En la tabla **3-4** se resalta en negrita a las correlaciones más notables. Esta matriz de covarianzas tiene asociado un número de condición de 21.87206.

2. Matriz de covarianzas poblacional con número de condición grande o "mal condicionada"

   De acuerdo con la simulación de los autores Liang y cols. (2009) que comparten el objetivo de este trabajo de grado, se plantea utilizar $\mathbf{\Sigma} = (\sigma_{ij})$ con $\sigma_{ii} = 1$ y $\sigma_{ij} = 0.9$

con $1 \leq i, j \leq p$. Esta matriz de covarianzas tiene asociado un número de condición de 109.

3. Vector de medias poblacional

   Se escoge $\boldsymbol{\mu} = \mathbf{0}$ sin pérdida de generalidad.

Tabla 3-3: Matriz de covarianza poblacional con varianzas desiguales y covarianzas positivas, negativas y nulas

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
Y_1	13.0457511	1.9249115	3.1102609	-1.5973858	-2.1842978	-1.2792006	-1.3227563	1.4556690	1.3087985	1.2888275	0.6053749	-0.5959182
Y_2	1.9249115	9.2215604	-2.7945128	2.6037288	1.6839693	1.3696587	-1.5751598	0.7906141	1.1743012	0.9338424	-0.8197276	0.2555049
Y_3	3.1102609	-2.7945128	9.5521954	1.3204526	-1.7274096	1.2076476	-2.2041248	0.7860761	2.0502425	-0.8532240	-0.5461424	-0.2659541
Y_4	-1.5973858	2.6037288	1.3204526	10.5268432	1.2617634	-1.2694039	-2.3564630	0.7337672	2.0180160	1.5152041	-0.0996326	0.2119220
Y_5	-2.1842978	1.6839693	-1.7274096	1.2617634	5.7154624	1.1400372	1.3922848	0.6212800	0.5976972	-1.1100725	0.2928464	0.9106820
Y_6	-1.2792006	1.3696587	1.2076476	-1.2694039	1.1400372	7.4619515	-1.8971128	0.9303997	2.0313814	-1.4950733	0.0641156	0.4582751
Y_7	-1.3227563	-1.5751598	-2.2041248	-2.3564630	1.3922848	-1.8971128	5.9339348	0.7670789	-1.6945787	0.6041721	0.1234187	0.3887118
Y_8	1.4556690	0.7906141	0.7860761	0.7337672	0.6212800	0.9303997	0.7670789	4.6249161	0.8846730	0.7570240	-0.4779086	1.0136268
Y_9	1.3087985	1.1743012	2.0502425	2.0180160	0.5976972	2.0313814	-1.6945787	0.8846730	6.1764256	1.1577593	1.8402553	-0.3082305
Y_{10}	1.2888275	0.9338424	-0.8532240	1.5152041	-1.1100725	-1.4950733	0.6041721	0.7570240	1.1577593	4.0364865	1.1793075	-1.1727331
Y_{11}	0.6053749	-0.8197276	-0.5461424	-0.0996326	0.2928464	0.0641156	0.1234187	-0.4779086	1.8402553	1.1793075	2.7124239	-0.8767903
Y_{12}	-0.5959182	0.2555049	-0.2659541	0.2119220	0.9106820	0.4582751	0.3887118	1.0136268	-0.3082305	-1.1727331	-0.8767903	2.1021959

Tabla 3-4: Matriz de correlaciones poblacional con correlaciones nulas, moderadas positivas y negativas

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
Y_1	1.000	0.175	0.279	-0.136	-0.253	-0.130	-0.150	0.187	0.146	0.178	0.102	-0.114
Y_2	0.175	1.000	-0.298	0.264	0.232	0.165	-0.213	0.121	0.156	0.153	-0.164	0.058
Y_3	0.279	-0.298	1.000	0.132	-0.234	0.143	-0.293	0.118	0.267	-0.137	-0.107	-0.059
Y_4	-0.136	0.264	0.132	1.000	0.163	-0.143	-0.298	0.105	0.250	0.232	-0.019	0.045
Y_5	-0.253	0.232	-0.234	0.163	1.000	0.175	0.239	0.121	0.101	-0.231	0.074	0.263
Y_6	-0.130	0.165	0.143	-0.143	0.175	1.000	-0.285	0.158	0.299	-0.272	0.014	0.116
Y_7	-0.150	-0.213	-0.293	-0.298	0.239	-0.285	1.000	0.146	-0.280	0.123	0.031	0.110
Y_8	0.187	0.121	0.118	0.105	0.121	0.158	0.146	1.000	0.166	0.175	-0.135	**0.325**
Y_9	0.146	0.156	0.267	0.250	0.101	0.299	-0.280	0.166	1.000	0.232	**0.450**	-0.086
Y_{10}	0.178	0.153	-0.137	0.232	-0.231	-0.272	0.123	0.175	0.232	1.000	**0.356**	**-0.403**
Y_{11}	0.102	-0.164	-0.107	-0.019	0.074	0.014	0.031	-0.135	**0.450**	**0.356**	1.000	**-0.367**
Y_{12}	-0.114	0.058	-0.059	0.045	0.263	0.116	0.110	**0.325**	-0.086	**-0.403**	**-0.367**	1.000

Es destacable que establecer una matriz de covarianzas que tenga asociadas correlaciones entre variables, varianzas desiguales y valores propios pequeños puede ser una tarea muy difícil; sin embargo, con el apoyo en la programación y aplicando los conceptos algebraicos es posible obtenerla, aunque las correlaciones asociadas deben ser sutiles o moderadas, ya que numéricamente no fue posible obtener una matriz con correlaciones fuertes, dimensión alta y valores propios pequeños.

3.2.3. Matriz objetivo T

La matriz de covarianzas tiene por estimar los elementos de la triangular superior (o inferior por simetría) que se dividen en los elementos de la diagonal principal correspondientes a las varianzas y los elementos por fuera de la diagonal correspondientes a las covarianzas. Los autores Schäfer y Strimmer (2005) brindan una expresión del estimador óptimo de la intensidad de contracción para las covarianzas muestrales y dos años después Opgen-Rhein y Strimmer (2007) siguiendo la idea general de contracción, proponen como extensión a un estimador de intensidad de contracción óptimo para las varianzas muestrales. Dado que la contracción de las covarianzas muestrales es diferente a la contracción de las varianzas muestrales, se define una matriz objetivo para la contracción de las covarianzas muestrales y se define un vector objetivo para la contracción de las varianzas muestrales.

- **Matriz objetivo para contracción de las covarianzas muestrales:**

  La elección de la matriz objetivo se basa en el análisis de las matrices objetivo de la Tabla **3-5**.

Tabla 3-5: Seis objetivos de contracción de uso común para la matriz de covarianza y estimadores asociados de la intensidad de contracción óptima; para la discusión se puede consultar el texto principal.

A. Matriz diagonal, varianza unitaria	**B**. Matriz diagonal, varianza común
0 parámetros estimados $t_{ij} = \begin{cases} 1 & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j \end{cases}$	1 parámetro estimado: ν $t_{ij} = \begin{cases} \nu = \text{avg}(s_{ii}) & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j \end{cases}$
C. Varianza y covarianza común	**D**. Matriz diagonal, varianza desigual
2 parámetros estimados: ν, c $t_{ij} = \begin{cases} \nu = \text{avg}(s_{ii}) & \text{si} \quad i = j \\ c = \text{avg}(s_{ij}) & \text{si} \quad i \neq j \end{cases}$	p parámetros estimados: s_{ij} $t_{ij} = \begin{cases} s_{ii} & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j \end{cases}$
E. Correlación positiva perfecta	**F**. Correlación constante
p parámetros estimados: s_{ii} $t_{ij} = \begin{cases} s_{ii} & \text{si} \quad i = j \\ \sqrt{s_{ii}s_{jj}} & \text{si} \quad i \neq j \end{cases}$	$p+1$ parámetros estimados: $s_{ii}, \bar{r}$ $t_{ij} = \begin{cases} s_{ii} & \text{si} \quad i = j \\ \bar{r}\sqrt{s_{ii}s_{jj}} & \text{si} \quad i \neq j \end{cases}$

Nota: ν denota el promedio de las varianzas muestrales, c el promedio de las covarianzas muestrales, $\bar{r}$ el promedio de las correlaciones muestrales y $\bar{r}$ es el promedio de las correlaciones muestrales.

La matriz **D** tiene un modelo de "varianza desigual en diagonal" que representa un balance entre los objetivos de baja dimensión **A**, **B** y **C** y los modelos de correlación **E** y **F**, el balance se da porque de **A**, **B** y **C** se toma la reducción de las entradas fuera

de la diagonal a cero y de los modelos **E** y **F**, deja intactas las entradas diagonales, es decir, no reduce las varianzas; para ilustrar el proceso, en la siguiente operación se calcula la contracción a una matriz de covarianzas muestral arbitraria con la matriz objeto **D** y un lambda estimado arbitrario, así, se obtendría

$$0.7\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} + (1-0.7)\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0.7a + 0.3a & 0{,}3b \\ 0.3c & 0.7d + 0.3d \end{bmatrix} = \begin{bmatrix} a & 0.3b \\ 0.3c & d \end{bmatrix}$$

una estimación que solo contrae las covarianzas.

Sobre la elección de la matriz objetivo **D**, Schäfer y Strimmer (2005) afirman que, "Usar el objetivo **D** tiene otra ventaja importante: la estimación de la covarianza de contracción resultante será automáticamente definida positiva. El objetivo **D** en sí es siempre definido positivo, y la combinación convexa de una matriz definida positiva (**T**) con otra matriz que es semidefinida positiva (**S**) siempre produce una matriz definida positiva. Tenga en cuenta que esto también es cierto para los objetivos **A** y **B**, pero no para los objetivos **C**, **E** y **F** (considere como contraejemplo el objetivo **E** con todas las varianzas establecidas en uno)" (p. 13).

Al separar los parámetros de la matriz de covarianzas y aplicar un proceso de contracción distinto a cada uno, también se sugiere que para fines de reducción puede ser útil estudiar los parámetros de la matriz de covarianzas en términos de varianzas y correlaciones, en vez de varianzas y covarianzas, haciendo $s^*_{ij} = r^*_{ij} s^{*1/2}_{ii} s^{*1/2}_{jj}$, donde r^*_{ij} son las correlaciones muestrales contraídas y s^*_{ii} (s^*_{jj}) son las varianzas muestrales contraídas. En este planteamiento, la contracción se aplica a las correlaciones en lugar de a las covarianzas, lo cual trae consigo dos ventajas: Primero, los elementos fuera de la diagonal que determinan la intensidad de la contracción están todos en la misma escala. Segundo, las correlaciones (parciales) derivadas del estimador de covarianza resultante $\mathbf{S}^*$ son independientes de las transformaciones de escala y ubicación de la matriz de datos subyacente, tal como ocurre con las calculadas a partir de **S**.

La matriz objetivo **T.cor** para el caso de correlaciones es la matriz idéntica. Se resalta que el efecto de contraer con la matriz idéntica a las matriz de covarianzas es muy distinto al efecto de usar la idéntica para contraer la matriz de correlación. Usar la matriz objetivo idéntica para las correlaciones, tiene el mismo efecto de usar la matriz objetivo **D** para contraer las covarianzas.

En 3.2.4 que corresponde al cálculo del estimador de la intensidad de contracción se ilustrará la transición de las covarianzas a las correlaciones y en 3.2.5 al calcular la

estimación shrinkage de la matriz de covarianzas final se evidencia cómo se reconstruye las covarianzas contraídas a partir de las correlaciones contraídas y las varianzas contraídas.

- **Vector objetivo para contracción de las varianzas muestrales:**

  Opgen-Rhein y Strimmer (2007) proporcionan una estimación shrinkage óptima del vector de varianza. Para ello, calculan las p varianzas insesgadas muestrales de la matriz de datos denotadas por $s_{11}, s_{22}, ..., s_{pp}$. Ahora bien, para obtener el vector objetivo adecuado se calcula la mediana de las s_g y se pone en todas las entradas del vector objetivo denotado como $\mathbf{T}.\mathbf{var}$.

  Formalmente, se desea estimar es el vector de varianzas $\mathbf{V} = (\sigma_{11}, \sigma_{22}, ..., \sigma_{gg}, ..., \sigma_{pp})^{'}$, esto se hará mediante un estimador de contracción denotado por $\mathbf{V}^*$, así que

  $$\mathbf{V}^* = \hat{\lambda}^M.\text{Var}\ (\mathbf{T}.\mathbf{Var}) + (1 - \hat{\lambda}^M.\text{Var})\hat{\mathbf{V}}$$

  Donde $\hat{\mathbf{V}}$ es la estimación tradicional de cada varianza y $\mathbf{T}.\mathbf{var}$ es el vector objetivo de dimensión $(p \times 1)$ que se construye como

  $$\mathbf{T}.\mathbf{var} = s_{\text{mediana}}\mathbf{1}' = (s_{\text{mediana}}, s_{\text{mediana}}, ..., s_{\text{mediana}})^{'} \tag{3-1}$$

  Opgen-Rhein y Strimmer (2007) mencionan que en la exploración de otros posibles objetivos también se consideró la reducción contra cero y hacia la media de las varianzas muestrales. Pero, estas dos alternativas resultaron ser menos eficientes (objetivo cero) o menos sólidas (objetivo medio) que hacer la reducción hacia la mediana (p. 7).

  Cabe resaltar que si el investigador tiene un indicio bien fundamentado sobre el comportamiento de cada varianza o de un subconjunto de ellas, entonces, el vector objetivo lo debe plantear obedeciendo a su criterio. Para este trabajo se utiliza a la mediana ya que la simulación no corresponde a un problema en particular y no se asume disponibilidad de información a priori.

3.2.4. Estimador del parámetro de contracción λ^*

Recordando que la estimación shrinkage de la matriz de covarianzas está definida por $\hat{\lambda}^M\mathbf{T} + (1-\hat{\lambda}^M)\mathbf{S}$; el estimador $\hat{\lambda}^M = 0.25$ realiza un balance de información, donde $\mathbf{S}$ queda similar y se le suma poca información de $\mathbf{T}$; el estimador $\hat{\lambda}^M = 0.5$ realiza un balance equilibrado de información, donde a la mitad de la información $\mathbf{S}$ se le suma la mitad de la información de $\mathbf{T}$; el estimador $\hat{\lambda}^M = 0.75$ realiza un balance de información, donde $\mathbf{T}$ queda similar y se le suma poca información de $\mathbf{S}$. En los casos extremos en realidad no hay balance, cuando $\hat{\lambda}^M = 1$, entonces la estimación es $\mathbf{T}$ (altamente estructurada) y cuando $\hat{\lambda}^M = 0$, entonces

la estimación es $\mathbf{S}$ (mal-condicionada bajo $n \cong p$); por esto no se considerarán $\hat{\lambda}^M = 1$ y $\hat{\lambda}^M = 0$. De acuerdo con el análisis anterior se evidencia que es necesario definir una intensidad de contracción óptima que regule la intensidad dependiendo de qué tan mal estimada este la matriz de covarianza por su estimador $\mathbf{S}$.

Ahora bien, la contracción de las covarianzas tiene un proceso distinto al de las varianzas; debido a esto, se definen los respectivos $\hat{\lambda}^M$.cov y $\hat{\lambda}^M$.var haciendo uso de la matriz objetivo y vector objetivo definido en 3.2.3. Finalmente se realiza la transición de la contracción de las covarianzas a la contracción de las correlaciones y se da la expresión óptima denotada como $\hat{\lambda}^M$.cor.

- **Intensidad de contracción para las correlaciones denotada por $\hat{\lambda}^M$.cor**

  Se presenta la transición de la contracción de las covarianzas a la contracción de las correlaciones.

  En primer lugar, para obtener la intensidad de contracción óptima de las covarianzas, se reemplaza la matriz objetivo $\mathbf{D}$ en la expresión general

  $$\hat{\lambda^*} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} \hat{\mathrm{Var}}(s_{ij}) - \hat{\mathrm{Cov}}(t_{ij}, s_{ij}) - \hat{\mathrm{Bias}}(s_{ij})(t_{ij} - s_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p}(t_{ij} - s_{ij})^2} \tag{3-2}$$

  dada en (2-10). Así, se obtiene que la intensidad de contracción óptima para las covarianzas es

  $$\hat{\lambda}^M.\mathrm{cov} = \frac{\sum_{i\neq j} \hat{\mathrm{Var}}(s_{ij})}{\sum_{i\neq j} s_{ij}^2} \tag{3-3}$$

  En (3-2) se restringe para $i \neq j$ puesto que, cuando $i = j$ entonces $t_{ii} = s_{ii}$ el denominador se hace cero, el término del $\hat{\mathrm{Bias}}(s_{ii})(t_{ii} - s_{ii})$ se anula dada la diferencia de términos iguales que los acompaña y dado que s_{ii} es un estimador insesgado; por último, $\hat{\mathrm{Cov}}(t_{ii}, s_{ii})$ será la varianza de t_{ii} o de s_{ii}, la cual actúa en la expresión como restar dicha varianza que fue sumada en el término anterior $\mathrm{Var}(s_{ii})$, esto sugiere que da igual sumarla y luego restarla que solo no considerarla al hacer que la sumatoria aplique solo para $i \neq j$; De acuerdo con el análisis anterior se concluye que bajo $i = j$ la expresión general no alimenta a las sumatorias.

  Ahora bien, cuando $i \neq j$, entonces $t_{ij} = 0$; fijando la atención en cómo se transforma cada término del numerador se observa que, el término del $\hat{\mathrm{Bias}}(s_{ij})(t_{ij} - s_{ij})$ se anula porque s_{ij} son estimadores insesgados y respecto a la $\hat{\mathrm{Cov}}(t_{ij}, s_{ij})$ se tendría la $\hat{\mathrm{Cov}}(0, s_{ij})$ que es igual a 0. Finalmente, el término que sobrevive es el numerador $\sum_{i\neq j} \hat{\mathrm{Var}}(s_{ij})$. Para el denominador, como $t_{ij} = 0$, entonces la suma quede como $\sum_{i\neq j}(0 - s_{ij})^2 = \sum_{i\neq j} s_{ij}^2$. Consolidando el numerador con denominador, el resultado

luce como la expresión hallada en (3-3).

Retomando las ventajas expuestas en 3.2.3 sobre primero contraer la matriz de correlación muestral y luego reconstruir a partir de ello a la matriz de covarianzas muestral contraída, el estimador de contracción óptimo para las correlaciones se obtiene de forma análoga usando la matriz de correlación muestral $\mathbf{R}$. Así, la intensidad de contracción óptima para las correlaciones es

$$\hat{\lambda}^M.\text{cor} = \frac{\sum_{i \neq j} \hat{\text{Var}}(r_{ij})}{\sum_{i \neq j} r_{ij}^2} \tag{3-4}$$

El cálculo de $\hat{\text{Var}}(s_{ij})$ se puede consultar en el apéndice A de Schäfer y Strimmer (2005) y aplicando las fórmulas dadas por los autores a la matriz de datos estandarizada, se obtiene $\hat{\text{Var}}(r_{ij})$.

- **Intensidad de contracción para las varianzas denotada por $\hat{\lambda}^M$.var**

  Opgen-Rhein y Strimmer (2007) proporcionan la intensidad de contracción óptima para las varianzas basados en la perspectiva de James-Stein que en primer lugar, selecciona una función de pérdida y en segundo lugar, la intensidad de contracción óptima se elige de manera que se minimice el riesgo correspondiente a la estimación shrinkage, es decir, la pérdida esperada con respecto a los datos. Particularmente, se emplea cuando el error cuadrático medio (MSE) es mínimo.

  El MSE de $\mathbf{V}^* = \hat{\lambda}^M.\text{var}\ (\mathbf{T}.\mathbf{var}) + (1 - \hat{\lambda}^M.\text{var})\hat{\mathbf{V}}$ no depende de ningún término del parámetro $\mathbf{V} = (\sigma_{11}, \sigma_{22}, ..., \sigma_{gg}, ..., \sigma_{pp})'$ ya que el MSE se puede escribir como sigue:

$$\begin{aligned} MSE(V^*) &= MSE(s_g) + \lambda^2 \sum_{g=1}^{p} \{E(s_g - \text{t.var}_g)^2\} \\ &\quad - 2\lambda \sum_{g=1}^{p} \{\text{Var}(s_g) - \text{Cov}(s_g, \text{t.var}_g) \\ &\quad + \text{Bias}(s_g) E(s_g - \text{t.var}_g)\} \\ &= c + \lambda^2 b - 2\lambda a \end{aligned} \tag{3-5}$$

  Por lo tanto, la curva de riesgo MSE tiene la forma de una parábola cuyos parámetros a, b y c están completamente determinados solo por los dos primeros momentos distributivos de $\hat{\mathbf{V}}$ y $\mathbf{T}.\mathbf{Var}$. Lo anterior sugiere que la expresión es diferenciable y se puede proceder a hallar el parámetro de contracción que hace al MSE en (3-5) mínimo.

  El estimador de contracción que minimiza el MSE es

$$\hat{\lambda}^M.\text{Var} = \text{mín}(1, \frac{\sum_{g=1}^{p} \hat{\text{Var}}(s_g)}{\sum_{g=1}^{p} (s_g - s_{\text{mediana}})^2}) \tag{3-6}$$

Para esta expresión se utilizó que $\text{Cov}(s_g, s_{\text{mediana}}) \approx 0$. Intuitivamente, si las varianzas muestrales s_g se pueden determinar de manera confiable a partir de los datos y, en consecuencia, solo exhiben una pequeña varianza, habrá poca reducción, mientras que si $\hat{\text{Var}}(s_g)$ es comparativamente grande, se producirá una agrupación entre variables. Además, el denominador en (3-6) es un estimación de la especificación errónea entre el objetivo y el s_g. Por lo tanto, si el objetivo se elige incorrectamente, tampoco se producirá una contracción.

3.2.5. Cálculo de la estimación shrinkage

El cálculo de la estimación shrinkage de la matriz de covarianzas final se realizará en tres pasos usando los elementos definidos de la sección 3.2.1 a la sección 3.2.4.

1. **Contracción de las correlaciones**

   Se obtiene la matriz de correlaciones contraída de orden $(p \times p)$ así:

   $\mathbf{R}^* = [\ \hat{\lambda}^M.\text{Cor}\ (\mathbf{T}.\mathbf{Cor}) + (1 - \hat{\lambda}^M.\text{Cor})\mathbf{R}\ ]$

   donde $\mathbf{R}$ denota la matriz de correlación muestral definida en (2-6), $\mathbf{T}.\mathbf{Cor}$ la matriz objetivo que corresponde a la matriz identidad y $\hat{\lambda}^M.\text{Cor}$ la intensidad de contracción expresada en (3-4).

2. **Contracción de las varianzas**

   Se obtiene la matriz diagonal de orden $(p \times p)$ que contiene en la diagonal a las desviaciones estándar contraídas definidas en :

   $$\mathbf{K}^{*\mathbf{1/2}} = [\ \{\text{diag}(\hat{\lambda}^M.\text{Var}\ (\mathbf{T}.\mathbf{Var}) + (1 - \hat{\lambda}^M.\text{Var})\hat{\mathbf{V}})\}^{1/2}\ ] \tag{3-7}$$

   donde $\hat{\mathbf{V}}$ es el vector de varianzas muestral, $\mathbf{T}.\mathbf{Var}$ es el vector objetivo definido en (3-1) y $\hat{\lambda}^M.\text{Var}$ la intensidad de contracción óptima expuesta en (3-6).

3. **Consolidación de los pasos anteriores para construir la estimación shrinkage de la matriz de covarianzas final**

   Para construir la estimación shrinkage de la matriz de covarianzas a partir de los cálculos anteriores, se recuerda que la covarianza se puede ver como $s^*_{ij} = r^*_{ij} s^{*1/2}_{ii} s^{*1/2}_{jj}$, por tanto, la estimación shrinkage de la matriz de covarianzas final se define como

   $\mathbf{S}^* = \mathbf{R}^* \mathbf{K}^{*\mathbf{1/2}} \mathbf{K}^{*\mathbf{1/2}}$

Para ilustrar el cálculo anterior, en la siguiente operación se calcula una estimación shrinkage de la matriz de covarianzas de dimensión (2×2) con correlaciones contraídas $r^*_{12} = r^*_{21} = 0{,}7$ y con desviaciones estándar contraídas $s^{*1/2}_{12} = s^{*1/2}_{21} = 5$, así, se obtendría

$$\begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} \begin{bmatrix} 5^{1/2} & 0 \\ 0 & 5^{1/2} \end{bmatrix} \begin{bmatrix} 5^{1/2} & 0 \\ 0 & 5^{1/2} \end{bmatrix} = \begin{bmatrix} 5^{1/2} & (5)5^{1/2} \\ (5)5^{1/2} & 5^{1/2} \end{bmatrix} \begin{bmatrix} 5^{1/2} & 0 \\ 0 & 5^{1/2} \end{bmatrix} = \begin{bmatrix} 5 & (0.7)5^{1/2}5^{1/2} \\ (0.7)5^{1/2}5^{1/2} & 5 \end{bmatrix}$$

que en la expresión final coincide con la estructura de una matriz de covarianzas.

Alguien se podría preguntar, ¿por qué usar los valores que se supone están mal estimados por la matriz de covarianzas muestral bajo $n \cong p$, para presuntamente "mejorar" la estimación calculando a partir de ellos una estimación shrinkage óptima? De esto, los autores Opgen-Rhein y Strimmer (2007) mencionan que: "inevitablemente, esto conduce a un aumento del riesgo total del estimador de contracción resultante. Sin embargo, es un resultado clásico de Stein que el costo de estimar la intensidad de la contracción ya está (¡y siempre!) compensado por los ahorros en el riesgo total cuando la dimensión p es mayor que tres" (p. 4). Para ello, presentan que el error cuadrático medio de la estimación tradicional supera al de la estimación contraída en $MSE(S) - MSE(S^*) = \frac{a}{b}$, donde a y b están definidos en (3-5).

3.2.6. Distribuciones para generar los datos

Se propone evaluar el desempeño de las pruebas a muestras provenientes de cinco distribuciones diferentes que se describirán en esta subsección. Las muestras provenientes de las distribuciones a, b y e se simulan partiendo de los parámetros $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ determinados en la Tabla **3-1**. Las demás distribuciones (c y d) aunque no usan como parámetro a la matriz de covarianzas, una vez se genera la matriz de datos se pueden estudiar las relaciones entre variables, para ello es útil calcular la matriz de covarianza muestral con el estimador tradicional para incorporar en la prueba tradicional y la estimación shrinkage para incorporar en la prueba modificada. Además, se resalta que de las 5 distribuciones, las b, c y d (ítem 2) se toman del estudio de simulación conducido por Liang y cols. (2009) donde se comparte objetivo de evaluar una prueba de multinormalidad en presencia de alta dimensionalidad. La elección de las distribuciones es estratégica como se justifica a continuación.

El análisis se separa en dos ejes principales, el primero consiste en evaluar la prueba bajo multinormalidad y el segundo bajo no multinormalidad. Las muestras se generan para las siguientes distribuciones.

1. Datos provenientes de una distribución normal multivariada.
   a. La distribución normal multivariada con parámetros $\boldsymbol{\mu} = \mathbf{0}$ y $\boldsymbol{\Sigma}$.
2. Datos provenientes de distribuciones multivariadas no-normales.
   b. La distribución t multivariante con parámetros $(m = 5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma})$.

      Cuando una variable aleatoria sigue una distribución t multivariante, se cumple que las variables marginales siguen una distribución t. De forma análoga sucede con una variable aleatoria que sigue una distribución multinormal ya que implica marginales normales. Partiendo de la premisa anterior, se intenta asemejar una

distribución t multivariante a una distribución multinormal mediante las densidades marginales.

La densidad de una t es simétrica con forma acampanada tal como sucede con la densidad de una normal pero la diferencia es que la t tiene colas "más pesadas".

La definición de la distribución t multivariante puede ser encontrada en Kotz y Nadarajah (2004) y los datos pueden ser generados con el paquete **mvtnorm** (R Core Team, 2021) en el software R.

c. La $\chi^2(1)$ desplazada con marginales i.i.d. Cada marginal tiene la misma distribución que la de la variable aleatoria $Y = X - E(X)$. Donde $X \sim \chi^2(1)$ es la distribución chi-cuadrado univariante con 1 grado de libertad y $E(X) = 1$.

Al conformar cada marginal como una chi cuadrado con un grado de libertad menos la esperanza, se está desplazando a la muestra hacia 0, sin embargo, esto no quita el hecho de que las marginales continúen siendo asimétrica pero sí se establece cierta similitud con la simetría alrededor de 0 de marginales normales estándar.

d. La distribución $\chi^2(1)+nor$. Consiste en $[p/2]$ marginales i.i.d, cada marginal tiene la misma distribución que la de la variable aleatoria $Y = X-E(X)$ con $X \sim \chi^2(1)$, y $p-[p/2]$ son marginales i.i.d normales estándar, donde $[p/2]$ representa la parte entera de $p/2$.

Aquí se complementa lo propuesto en la distribución c. porque la mitad de marginales siguen distribuciones chi cuadrado desplazadas (asimétrica con media alrededor de cero) y la otra mitad siguen distribuciones normales estándar (simétricas e insesgadas).

e. La distribución Qu-Liu-Zhang. La distribución no normal multivariada con coeficiente de asimetría 3 y curtosis de 180. El nombre se le otorga en el presente trabajo debido a que se calcula con el método propuesto por Qu y cols.(2019).

Qu y cols. (2019) proporcionan un método para la generación números aleatorios multivariados con una asimetría y curtosis deseada. Lo más interesante es que la asimetría y la curtosis se plantea desde una perspectiva multivariada.

Ahora bien, la distribución normal multivariante cuenta con un coeficiente de asimetría y curtosis igual a cero. Por ello, se establecen ambas medidas alejadas de cero, provocando así una analogía de una distribución multinormal asimétrica y de colas pesadas. Los datos pueden ser generados a través del paquete **mnonr** (R Core Team, 2020) en el software R.

La idea de generar distribuciones no multinormales pero con características similares a las multinormales, es para exigirle a la prueba Shapiro-Wilk Generalizada Modificada

detectar aquellas diferencias que catalogan a las matrices de datos como no multinormales. Las características similares para b., c. y d. se proponen mediante las densidades marginales puesto que bajo multinormalidad, implica que cualquier subconjunto de las densidades marginales también son normales. Así, para el listado de distribuciones no multinormales se resume que b. es simétrica pero de colas más pesadas, c. tiene marginales asimétricas pero corregidas con un desplazamiento hacia cero, d. intentar confundir a la prueba con marginales asimétricas pero desplazadas hacia cero y marginales normales estándar. Por último, e. es la distribución más alejada de la normalidad ya que se propone con colas pesadas y asimétrica desde el punto de vista multivariado.

3.3. Evaluación de la prueba modificada

La evaluación del desempeño de la prueba se realiza mediante una comparación entre la prueba Shapiro-Wilk Generalizada modificada con la prueba tradicional Shapiro-Wilk Generalizada sometidas a un incremento gradual del tamaño de la muestra. La comparación tiene lugar en los escenarios de evaluación de la Tabla **3-2** y usando como medidas los criterios de la posterior Tabla **3-6**.

En principio, se propuso incluir en la comparación a otras pruebas de normalidad multivariada para muestras pequeñas que propusieron Liang, Li, Fang, y Fang (2000) basada en la idea de componentes principales utilizando la asimetría y curtosis multivariada de Mardia (1970) y la prueba propuesta por Liang y cols. (2009). Lamentablemente, no fue posible obtener una respuesta de los autores donde se solicitaba el código y semilla que conducía a la simulación presentada en sus artículos.

Las medidas de contraste están planteadas de acuerdo con la distribución de las muestras. Por un lado, en el caso multinormal, la hipótesis nula es verdadera, por esto, la decisión correcta sería no rechazarla. Aquí, se tiene en cuenta que α se define como la probabilidad de rechazar la hipótesis nula cuando es verdadera, por lo que, la medida de interés (decisión correcta) sería $1-\alpha$ que corresponde al complemento, es decir, la probabilidad de no rechazar la hipótesis nula cuando es verdadera. Por otro lado, en el caso no multinormal, la hipótesis nula es falsa, por esto, la decisión correcta sería rechazarla. Aquí, β se define como la probabilidad de no rechazar la hipótesis nula cuando es falsa, por lo que, la medida de interés sería $1-\beta$ que corresponde al complemento, es decir, la probabilidad de rechazar la hipótesis nula cuando es falsa, comúnmente llamada potencia de la prueba.

En la Tabla **3-6** se consolidan los criterios de evaluación de acuerdo a la procedencia de las muestras y además se mide para dos niveles de significancia distintos que buscan estudiar qué tan sensible es la decisión que se toma de acuerdo a un contraste entre el Valor p y el nivel de significancia.

Tabla 3-6: Criterios de evaluación

(n,p)	Dist. de datos	Estadístico	Criterio	Nivel de significancia
$(n = i, p = j)$, $i = 13, 14, ..., 60$ $j = 12$	Multinormal	W^* W^M	$1 - \alpha$	$\alpha = 0{,}01$ $\alpha = 0{,}05$
	No multinormal	W^* W^M	$1 - \beta$	

3.3.1. Síntesis de la evaluación de la prueba Shapiro-Wilk Generalizada modificada

Se describe de manera global el proceso de evaluación de la prueba Shapiro-Wilk Generalizada modificada.

Se calculan $J = 5{,}000$ muestras y se analizan para los tamaños $n = 13, ..., 60$. Las muestras son simuladas a partir de las distribuciones especificadas en la sección 3.2.6, con $p = 12$, parámetros $\boldsymbol{\mu} = \mathbf{0}$ y dos matrices de covarianza poblacionales (ver Tabla **3-1**): en la sección 3.2.2 se explica el procedimiento para llevar a cabo el cálculo de $\boldsymbol{\Sigma}$ que corresponde a una matriz (12×12) con número de condición pequeño, varianzas desiguales y correlaciones asociadas moderadas positivas, negativas y nulas (ver Tabla **3-3**); y $\boldsymbol{\Sigma}$ una matriz (12×12) con número de condición grande, varianzas unitarias y correlaciones asociadas fuertemente positivas. Luego, se calcula los Valores p asociados a los estadísticos W^* y W^M a cada muestra simulada.

1. H_0 verdadera

   Las muestras simuladas provienen de una población multinormal y dado que la hipótesis nula es verdadera, se hace un conteo de las 5.000 muestras que no rechazaron la hipótesis nula para un n específico.

2. H_0 falsa

   Las muestras simuladas provienen de una población no multinormal y dado que la hipótesis nula es falsa, se hace un conteo de las 5.000 muestras que rechazaron la hipótesis nula para un n específico.

La decisión de rechazar H_0 puede cambiar de acuerdo con el contraste α definido. De manera análoga se realiza lo mismo para cada siguiente n en orden secuencial, haciendo la proporción de cada conteo y plasmándolo en una gráfica donde se compare el desempeño de la prueba a medida que aumenta el tamaño de muestra.

4 Resultados

En esta sección se ilustran los efectos en el desempeño de la prueba de normalidad multivariada con la modificación propuesta. La comparación tiene lugar desde un enfoque descriptivo para la precisión y con la generación de datos para la evaluación en los escenarios propuestos y descritos en la metodología. Concretamente, la estimación shrinkage de la matriz de covarianzas se reconstruye a partir de las correlaciones contraídas y las desviaciones estándar contraídas.

Para las correlaciones contraídas: en primer lugar, se eligió la idéntica como matriz objetivo, es decir, $\mathbf{T}.\mathbf{Cor} = \mathbf{I}$ luego de discutir el porqué la estructura $\mathbf{D}$ tiene una serie de ventajas. En segundo lugar, para la matriz objetivo seleccionada se deduce la contracción óptima $\hat{\lambda}^M$.cor en (3-4) que depende de las correlaciones muestrales.

Para las desviaciones estándar contraídas: en primer lugar, se eligió a $\mathbf{T}.\mathbf{Var} = s_{\text{mediana}}\mathbf{1}'$ como vector objetivo. En segundo lugar, para el vector objetivo seleccionado se deduce la contracción óptima $\hat{\lambda}^M$.var en (3-4) que depende de las varianzas muestrales. En cuarto lugar, las varianzas contraídas se organizan en una matriz diagonal y se le aplica la raíz cuadrada para obtener la matriz diagonal de desviaciones estándar contraídas. (ver (3-7))

Particularmente, el análisis se presenta de manera detallada para las muestras generadas con la matriz de covarianzas poblacional bien condicionada, puesto que se evidenció de forma general que el desempeño de la prueba modificada es similar a cuando se evalúa para muestra generadas con la matriz de covarianzas mal condicionada. Sin embargo, al final de la sección se hacen los apuntes pertinentes.

4.1. Análisis comparativo de la precisión de la prueba tradicional y la prueba modificada

A continuación se realiza una comparación descriptiva de la precisión de ambas pruebas, al generar 5.000 repeticiones para 6 tamaños de muestra $n = 13, 20, 30, 40, 50$ y 60, evaluado en distribuciones multivariadas no normales descritas en la subsección 3.2.6.

Para muestras provenientes de una distribución t multivariante, la cual es simétrica y de

colas pesadas, la Figura **4-1** ilustra el comportamiento de ambas pruebas con respecto al valor-p. Por un lado, la prueba tradicional muestra mayor dispersión, indicando que para todos los tamaños de muestra considerados, los valores-p del estadístico están entre 0 y 1, pero mayoritariamente entre aproximadamente 0.01 y 0.5. Por otro lado, la prueba modificada presenta una dispersión mucho menor comparada con la prueba tradicional, pues el valor-p para los diferentes tamaños de muestra se localizan entre 0 y 0.4 y en la mayoría de los casos están entre aproximadamente 0 y 0.14. Aquellos valores que se encuentran por fuera de los bigotes, son valores extremos o atípicos que tienen frecuencias asociadas muy bajas. Además, se aprecia que más del 50 % de los valores-p asociados a la prueba modificada son menores a 0.05 lo que es bastante deseable ya que se H_0 es falsa y si la prueba funciona bien, para distintos tamaños de muestra, debería obtener valores-p cercanos a cero.

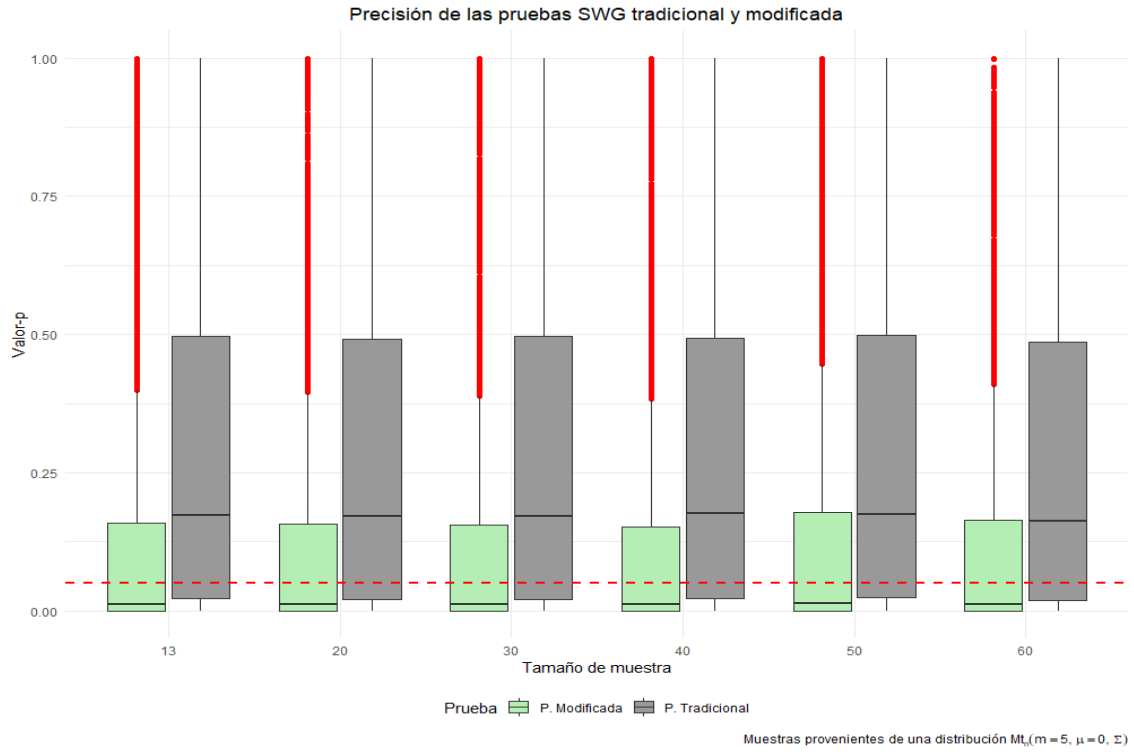
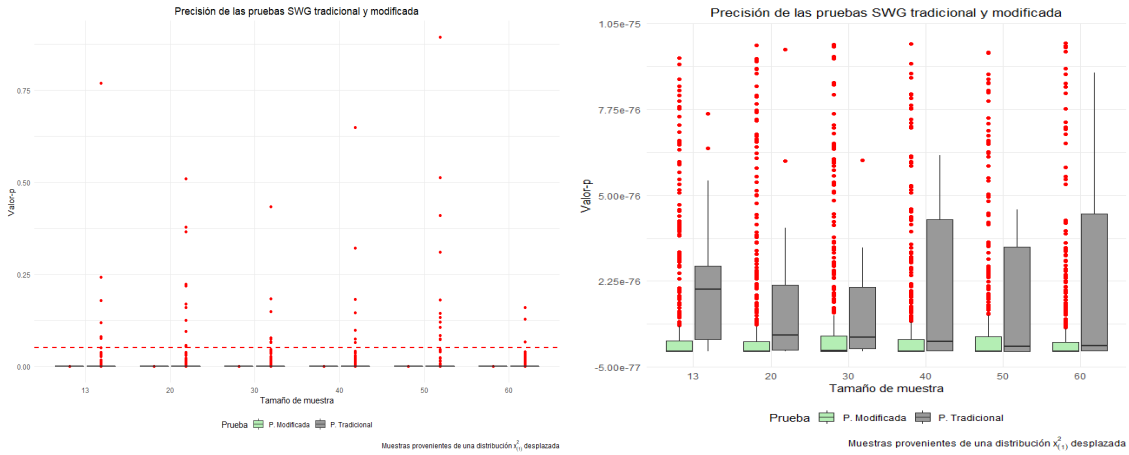


Figura 4-1: Precisión de las pruebas SWG tradicional y modificada con muestras provenientes de una distribución multi-t $(m = 5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma})$ con $\alpha = 0.05$

La Figura **4-2** y **4-3** ilustran el comportamiento de las dos pruebas para muestras provenientes de distribuciones sesgadas, las Figuras **4-2** (a) y **4-3** (a) del lado izquierdo muestra el comportamiento de ambas pruebas en todo el rango del valor-p que toma el estadístico y las Figuras **4-2** (b) y **4-3** (b) del lado derecho es un acercamiento visual al comportamiento de las cajas de ambas pruebas en el rango del valor-p. Se observa que la gran mayoría de los valores-p que toma el estadístico en ambas pruebas, se concentran en valores muy pequeños alrededor de cero. También es notorio que la prueba tradicional toma valores altos

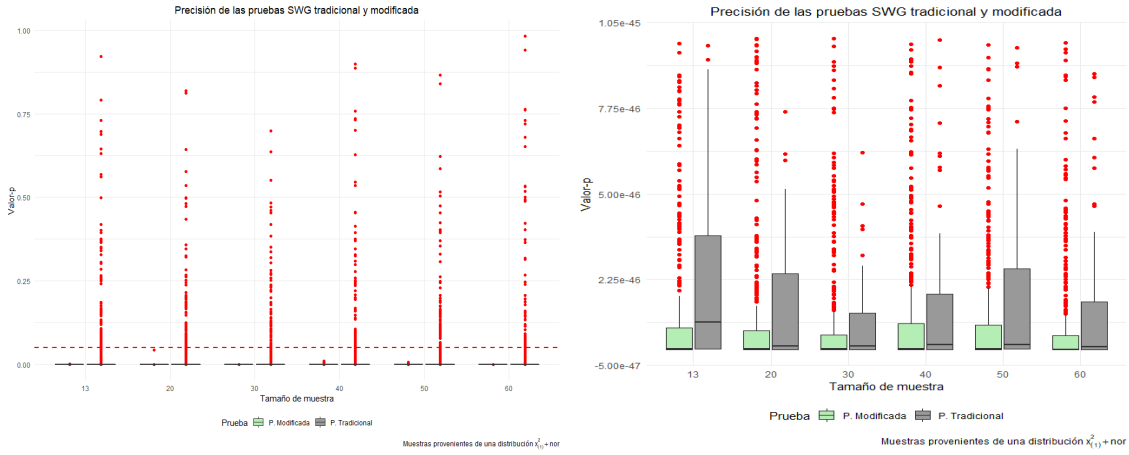
que son clasificados como atípicos, diferente a la prueba modificada que no presenta este comportamiento, lo cual indica que la prueba modificada presenta un mejor desempeño bajo este escenario puesto que los valores-p del estadístico siempre son menores o iguales a 0.05, incluso los que para su dispersión son atípicos.



(a) Comparación de valor-p en su rango completo

(b) Comparación de valor-p en un rango reducido para mejor visualización

Figura 4-2: Precisión de las pruebas SWG tradicional y modificada con muestras provenientes de una distribución $\chi^2_{(1)}$ desplazada con $\alpha = 0.05$



(a) Comparación de valor-p en su rango completo

(b) Comparación de valor-p en un rango reducido para mejor visualización

Figura 4-3: Precisión de las pruebas SWG tradicional y modificada con muestras provenientes de una distribución $\chi^2_{(1)} + nor$ con $\alpha = 0.05$

Para muestras provenientes de una distribución Qu-Lui-Zhang, la Figura **4-4** ilustra el comportamiento de la prueba tradicional y modificada para distintos tamaños de muestra. Se

observa que aunque ambas pruebas exhiben valores catalogados como extremos, es notorio que la prueba modificada presenta menor dispersión en comparación con la prueba tradicional, siendo ésta más precisa puesto que los valores-p del estadístico en su mayoría se concentran entre 0 y 0.05 aproximadamente, lo cual es deseable ya que el interés principal es obtener valores-p menores o iguales a 0.05 debido a que los datos no pertenecen a una distribución multinormal.

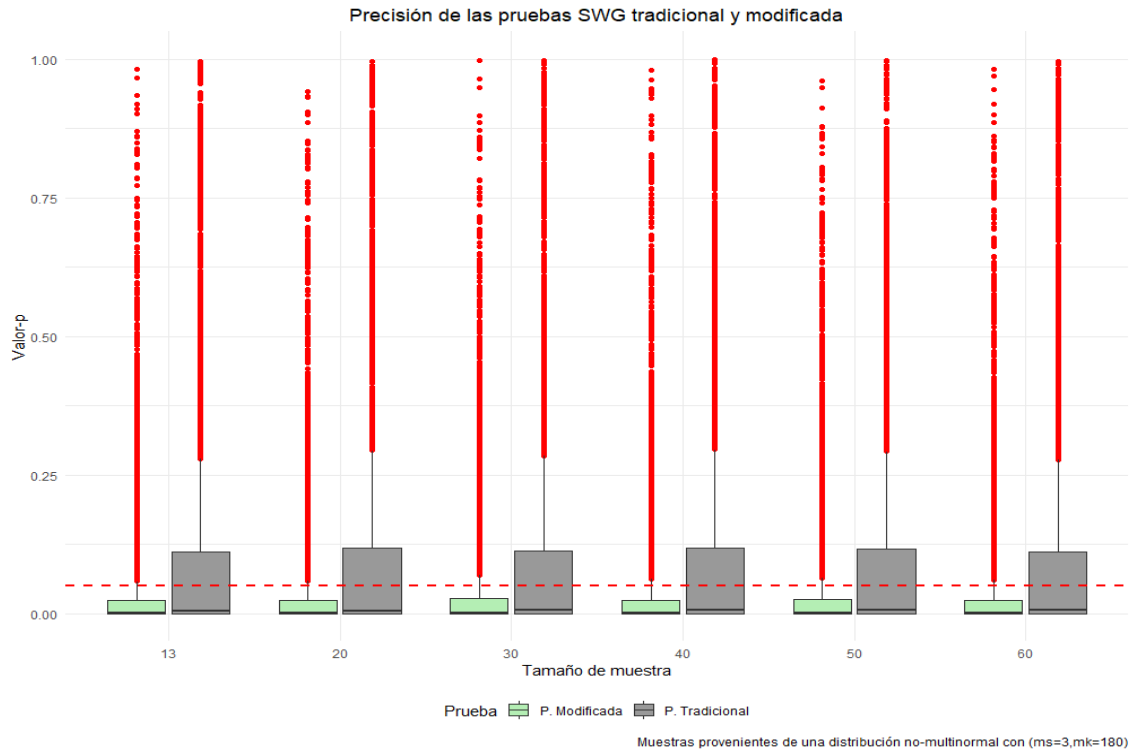


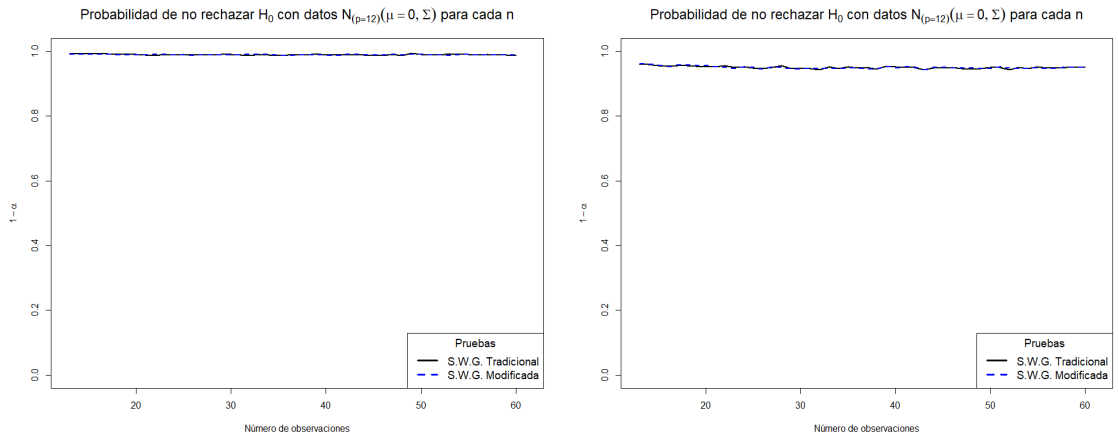
Figura 4-4: Precisión de las pruebas SWG tradicional y modificada con muestras provenientes de una distribución no-multinormal ($ms = 3, mk = 180$) con $\alpha = 0.05$

4.2. Comparación del desempeño de la prueba modificada en contraste con la prueba tradicional

Considerando los niveles de significancia de $\alpha = 0.01$ y $\alpha = 0.05$, se analiza la probabilidad de tomar la decisión correcta cuando los datos siguen una distribución multinormal; la probabilidad de error tipo II y la potencia cuando los datos provienen de una distribución no multinormal, respectivamente para ambas pruebas.

4.2.1. Población multinormal

La Figura **4-5** evalúa la probabilidad de no rechazar la hipótesis de normalidad multivariante dado que los datos provienen de una distribución multinormal. Se observa que para ambos niveles de significancia $\alpha = 0.01$ y $\alpha = 0.05$, no se presentan diferencias entre ambas pruebas y la probabilidad de tomar la decisión correcta se mantiene aproximadamente constante para cualquier tamaño de muestra.



(a) P(no rechazar $H_0|H_0$ verdadera) con $\alpha = 0.01$

(b) P(no rechazar $H_0|H_0$ verdadera) con $\alpha = 0.05$

Figura 4-5: Probabilidad de no rechazar H_0 con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) para cada n

4.2.2. Poblaciones no multinormales

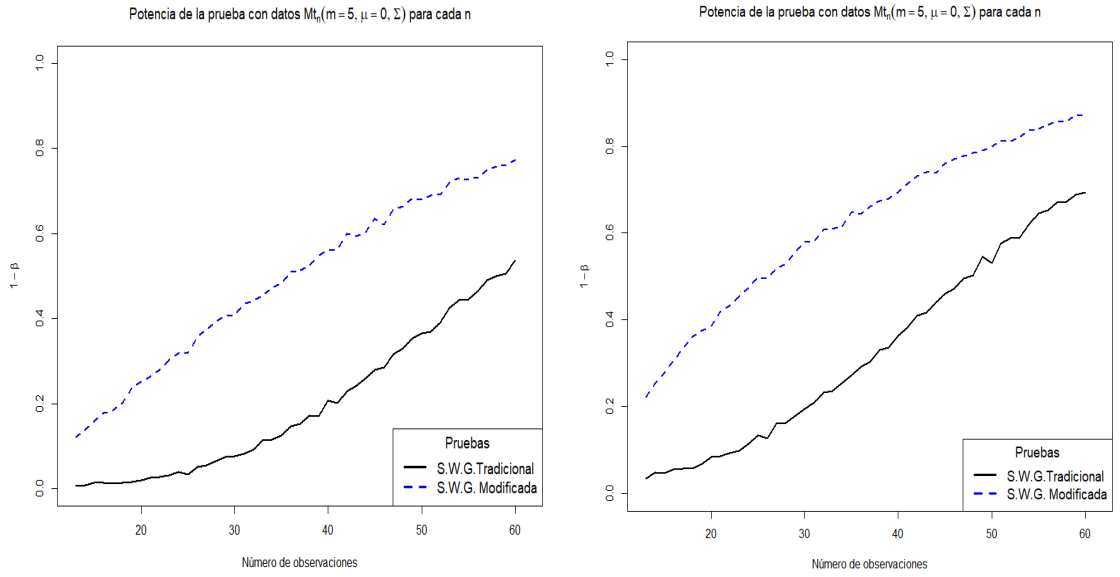
- **Datos provenientes de una distribución t multivariante**

  En la Figura **4-6** se evalúa la potencia con datos provenientes de una distribución t multivariante para cada prueba. Bajo este escenario se aprecia que para un número pequeño de observaciones, la probabilidad de rechazar la hipótesis de normalidad multivariante dado que los datos son no multinormales es considerablemente baja y a medida que el número de observaciones incrementa, la probabilidad de que ambas pruebas tomen la decisión correcta también aumenta.

  En una prueba de hipótesis, la probabilidad de error tipo I o nivel de significancia (α) hace referencia a la probabilidad de rechazar la hipótesis nula dado que es verdadera, y la probabilidad de error tipo II (β) se refiere a la probabilidad de no rechazar la hipótesis nula dado que es falsa. Al considerar distintos niveles de significancia para una misma prueba de hipótesis, la potencia de dicha prueba $(1 - \beta)$ se ve afectada, puesto que al reducir α, β incrementa y viceversa. Es por ello que en la Figura **4-6** (b) la potencia de ambas pruebas incrementa mucho más despacio, en comparación

con la potencia de la Figura **4-6** (a), ya que con $\alpha = 0.01$ se desea disminuir el error con respecto a no rechazar la hipótesis nula cuando ésta es cierta, es decir, aumentar la probabilidad de no rechazar la hipótesis nula de la prueba dado que es verdadera, lo cual aumenta consigo a β y le exige más rigurosidad a la prueba al momento de calcular la potencia. Disminuir o aumentar α implica un aumento o reducción en el número de observaciones para tomar la decisión correcta, respectivamente.

Cabe resaltar que para ambos niveles de significancia la prueba modificada presenta un desempeño notoriamente mayor al de la prueba tradicional.



(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

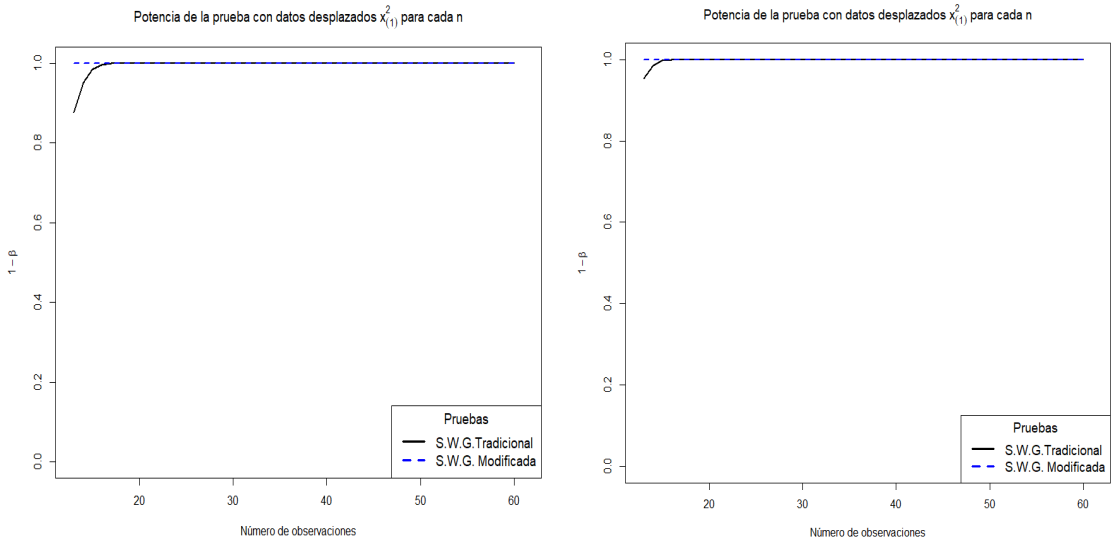
Figura 4-6: Potencia de la prueba para datos multi-t $(m = 5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma})$ con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) para cada n

- **Datos provenientes de una $\chi^2(1)$ desplazada y $\chi^2(1) + nor$**

A continuación se evalúa la sensibilidad de la prueba con datos provenientes de dos distribuciones sesgadas. Potencia de la prueba con datos chi con un alpha de 0.01 y 0.05 La Figura **4-7** estudia el desempeño de la prueba para datos provenientes de una distribución $\chi^2_{(1)}$ desplazada con marginales i.i.d, donde se establece cierta similitud con la simetría alrededor de 0 de las marginales normales estándar. Tanto para el nivel de significancia de $\alpha = 0.01$ como $\alpha = 0.05$, ambas pruebas presentan un desempeño bastante deseable; sin embargo, la prueba modificada muestra que para cualquier cantidad del número de observaciones, la probabilidad de rechazar la hipótesis de normalidad multivariante dado que dicha hipótesis es falsa, siempre es 1.

La Figura **4-8** también estudia el desempeño de la prueba pero le agrega una parte extra, puesto que la distribución que siguen los datos se compone de marginales i.i.d. $\chi^2_{(1)}$ desplazadas que son sesgadas y centradas, y $N(0,1)$ simétricas, centradas e insesgadas. Se destaca que para ambos niveles de significancia $\alpha = 0.01$ y $\alpha = 0.05$, la prueba modificada presenta una probabilidad de 1 para rechazar la hipótesis de normalidad multivariante dado que ésta es falsa para cualquier tamaño de muestra, caso contrario a la prueba tradicional que para una cantidad pequeña del número de observaciones, la probabilidad de tomar la decisión correcta es más bajo en comparación con la prueba modificada.

En síntesis, para ambas poblaciones, la prueba modificada es más potente que la prueba tradicional en muestras pequeñas y rápidamente se nivelan en términos del desempeño.

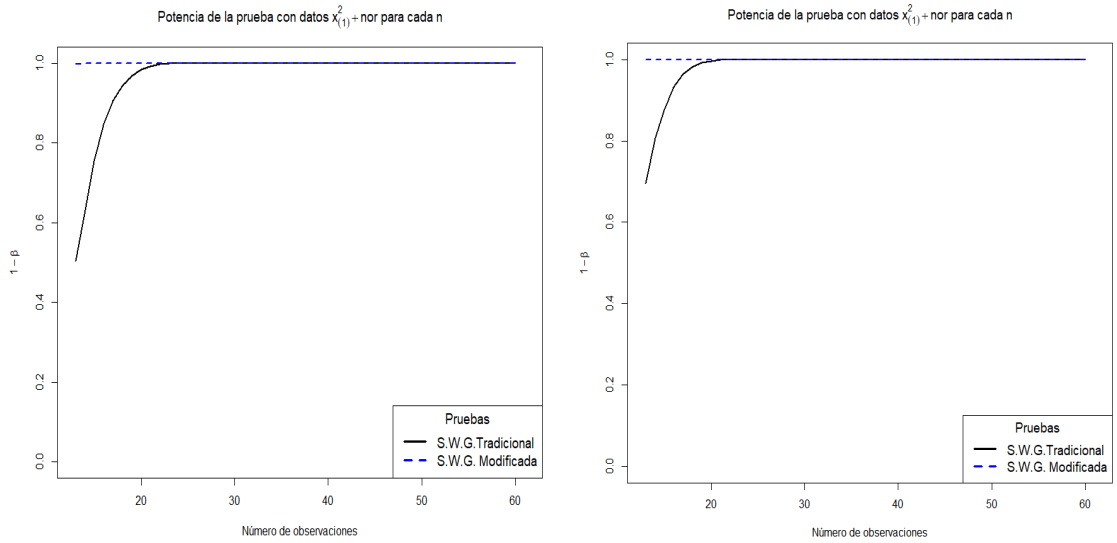


(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 4-7: Potencia de la prueba para datos desplazados $\chi^2_{(1)}$ con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) para cada n

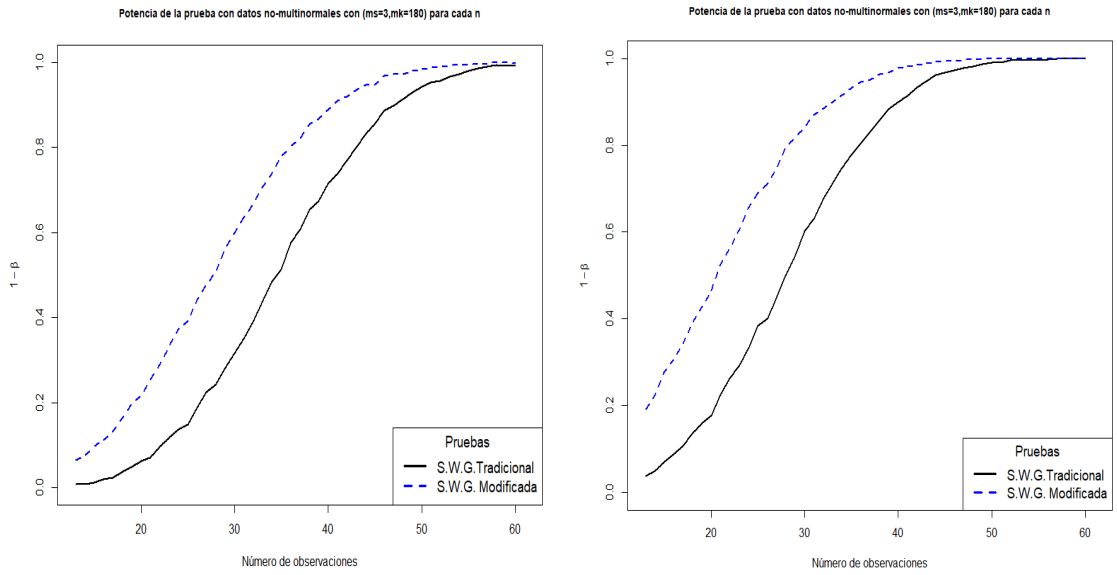
- **Distribución Qu-Lui-Zhang multivariante**

Los datos evaluados en la Figura **4-9** provienen de una Qu-Lui-Zhang asimétrica y de colas pesas, puesto que los coeficientes de asimetría y curtosis son distintos de cero. En este escenario es apreciable el aumento de la probabilidad de que ambas pruebas tomen la decisión correcta al rechazar la hipótesis de normalidad multivariante puesto que los datos evaluados son no multinormales, a medida que incrementa el tamaño de muestra. Además, para ambos niveles de significancia $\alpha = 0.01$ y $\alpha = 0.05$, la prueba modificada presenta un mejor desempeño con respecto a la prueba tradicional.



(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 4-8: Potencia de la prueba para datos $\chi^2_{(1)} + nor$ $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) para cada n



(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 4-9: Potencia de la prueba para datos distruidos Qu-Lui-Zhang $(ms = 3, mk = 180)$ con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) para cada n

Finalmente, se resalta que bajo estos escenarios, en todos los casos la prueba modificada presenta un desempeño mucho más deseable en comparación con la prueba tradicional. Además, un resultado bastante notorio es que para datos provenientes de distribuciones multivariantes asimétricas, la prueba modificada siempre toma la decisión correcta para cualquier tamaño de muestra. Para datos provenientes de distribuciones simétricas o asimétricas con colas pesadas, la prueba modificada requiere aumentar la cantidad de observaciones para acercar la probabilidad de rechazar la hipótesis de normalidad multivariante a uno (1).

En la evaluación del desempeño de la prueba con la matriz de covarianzas poblacional mal condicionada se evidenciaron resultados similares que cuando la matriz de covarianzas poblacional estaba bien condicionada, es decir, para los escenarios no multinormales que la prueba modificada tiene un mejor desempeño para muestras pequeñas y para el escenario multinormal, la prueba modificada mostró un desempeño comparativamente similar a la prueba tradicional. Los gráficos correspondientes se encuentran en el anexo 2.

4.3. Discusión

Reflexionando acerca del por qué la prueba modificada funciona mejor, se podría atribuir el mejoramiento en el desempeño a la calidad del estimador, puesto que la modificación consiste en cambiar el estimador de la matriz de covarianzas por el estimador shrinkage en la prueba de hipótesis. Bajo la premisa anterior, se evidenció lo siguiente:
La matriz de covarianzas poblacional bien condicionada tiene un número de condición de 21.87206 y la matriz de covarianzas poblacional mal condicionada tiene un número de condición de 109.

- Muestras generadas a partir de una población t-multivariada($m = 5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma}$):

  En el escenario en que las muestras fueron generadas con una matriz de covarianzas bien condicionada, se espera que a medida que incrementa el tamaño de muestra naturalmente el número de condición en ambos estimadores (tradicional y shrinkage) disminuya, ya que las estimaciones se parecen cada vez más a su parámetro que está bien condicionado. Efectivamente, en la parte izquierda de la Tabla **6-3** (Anexo 3) se evidencia que para muestras pequeñas el número de condición asociado a las estimación tradicional tiene valores mucho más elevados que los asociados a la estimación shrinkage, también a medida en que incrementa el tamaño de muestra el número de condición se va regularizando en los valores más pequeños. Lo anterior sugiere que, la prueba modificada debería funcionar bien en muestras pequeñas ya que el estimador shrinkage pudo caracterizar rápida y correctamente a su parámetro bien condicionado. Además, el número de condición de la estimación tradicional aún en el más grande de los tamaños de muestra considerados no logra acercarse al verdadero número de condición de 21.87206.

Ahora, en el escenario en que las muestras fueron generadas con una matriz de covarianzas mal condicionada, se espera que a medida que incrementa el tamaño de muestra, el número de condición en ambos estimadores (tradicional y shrinkage) se acerquen al número de condición poblacional ya que las estimaciones se parecen cada vez más a su parámetro que está mal condicionado. En la parte derecha de la Tabla **6-3** (Anexo 3), por un lado, se evidencia que para muestras pequeñas el número de condición asociado a la estimación tradicional tiene valores mucho más elevados que los asociados a la estimación shrinkage y a medida en que incrementa el tamaño de muestra se va regularizando alrededor de 300 (se considera mal condicionado); por otro lado, el estimador shrinkage de entrada tiene valores bajos y en la medida que incrementa el tamaño de muestra converje al número de condición del parámetro que es 109. Lo anterior sugiere que, para muestras pequeñas, la estimación shrinkage es más acertada que la estimación tradicional e implica que la prueba modificada debería funcionar bien en muestras pequeñas ya que el estimador pudo caracterizar rápida y correctamente a su parámetro bien condicionado.

Con los resultados obtenidos, se presume que efectivamente el desempeño mejorado de la prueba modificada se debe a que el estimador shrinkage estima mejor a la matriz de covarianza poblacional. Además, aunque el desempeño de la prueba modificada no es mejor que la prueba tradicional cuando se evalúa con datos provenientes de población multinormal, este es igual, lo cual da indicios de que en cualquier escenario la prueba modificada tiene un margen de error reducido.

- Muestras generadas a partir de una población multinormal ($p = 12, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma}$):

  En la parte izquierda de la Tabla **6-4** (Anexo 3) se evidencia que para muestras pequeñas el número de condición asociado a las estimación tradicional tiene valores mucho más elevados que los asociados a la estimación shrinkage y aunque se van regularizando, estos tiende a ser más altos que los asociados a shrinkage. Lo anterior sugiere que, la prueba modificada debería funcionar bien en muestras pequeñas ya que el estimador shrinkage pudo caracterizar rápida y correctamente a su parámetro bien condicionado. Sin embargo, aunque la estimación tradicional al parecer no estima mejor que la estimación shrinkage cuando las muestras son pequeñas, esto no se refleja en una disminución de su desempeño de la prueba. Lo mismo sucede cuando las muestras son generadas con una matriz de covarianza mal condicionada.

  Como la calidad del estimador al parecer no tiene una incidencia fuerte en la decisión que toman ambas pruebas cuando los datos provienen de una distribución multinormal, se podría tener en cuenta que para el cálculo del estadístico además del estimador de la matriz de covarianzas también se usa la matriz de datos y es posible que la forma directa de los datos prevalezca indicando la normalidad y conllevando a la decisión correcta. Una posible forma de comprobar lo anterior es, aplicar la prueba utilizando los datos multinormales pero variando un poco la estimación de la matriz de covarianzas, por

ejemplo, hacer que las entradas sean la media y variar dichos valores en un rango uniforme; así, en caso de evidenciar que la prueba sigue detectando la normalidad, permitiría afirmar con más certeza que la forma directa de los datos es más incidente para la decisión que toma la prueba de hipótesis.

Para las distribuciones $\chi^2(1)$ desplazada y $\chi^2(1) + nor$ no se puede analizar cómo se comporta el estimador de la matriz de covarianzas porque no se conoce su parámetro, pero al asumir independencia entre las variables aleatorias marginales, se espera que las covarianzas sean cercanas a cero y las varianzas no se sabe. En la simulación de la matriz de covarianzas bien condicionada se evidenció que cuando las correlaciones deseadas eran muy altas, esto elevaba el número de condición resultante además de la dimensionalidad. Por lo anterior, se podría pensar en que para estos datos generados con estas distribuciones, la estimación shrinkage también estima mejor a la matriz de covarianza poblacional ya que el estimador shrinkage en muestras pequeñas tiende a regular severamente al número de condición en comparación con la estimación tradicional y dado que en este caso las covarianzas poblacionales deberían ser cercanas a cero, entonces la estimación shrinkage da cuenta de ello para tamaños de muestra muy pequeños. Además, aunque no se pueda analizar el estimador, de los resultados se puede observar que estas distribuciones dan indicios de que la forma directa de los datos en la prueba de hipótesis parece ser muy relevante porque la prueba tradicional también detecta rápidamente la no multinormalidad. (Ver Figuras **4-7** y **4-8**)

4.4. Implementación en un caso práctico

La implementación de Shapiro-Wilk Generalizada modificada se hará con datos de partículas suspendidas en el aire que son más dañinas para la salud denominadas como $PM_{2.5}$, estos fueron proporcionados por el SVCASC y fueron medidos durante el año 2019. La base de datos presenta un alto número de registros faltantes que pueden suceder por diversos motivos técnicos. El objetivo es probar que cada día de la semana se puede modelar mediante una distribución normal 24-variada, donde la variable corresponde al vector que contiene las 24 horas del día. Sin embargo, esto no se puede abordar de manera habitual dada la presencia de registros faltantes que implican un mal condicionamiento en la matriz de covarianzas muestral debido a que $n \cong p$.

4.4.1. Estructura de los datos

Un registro corresponde al promedio de varias medidas tomadas durante una hora. Esto se realiza para cada hora durante cada día del año 2019. El día_j (j =1, 2, 3, 4, 5, 6 y 7) durante un año puede ser medido como máximo 53 veces y en el día_j se mide el $PM_{2.5}$ en las 24 horas. Con la información anterior y recordando el objetivo, se construye una matriz de datos diferente para cada día de la semana. La matriz de datos presenta algunas filas

incompletas dada la presencia de datos faltantes, así que, los tamaño de muestra para cada día podrían ser diferentes. Pero, se hace la salvedad de que en la consolidación de cada matriz de datos se tomarán las observaciones de los días que tienen registros en todas las horas. Así, la matriz de datos para el día$_j$ será dimensión ($n \leq 53 \times p = 24$).

4.4.2. Estimación shrinkage S^*

Se presenta el cálculo de la estimación shrinkage definida por Schäfer y Strimmer (2005) de la matriz de covarianzas $\mathbf{\Sigma}$ para el día lunes con los datos de $PM_{2.5}$ del año 2019.

En el año 2019, se dispone de 38 lunes que tienen registro de $PM_{2.5}$ en todas las 24 horas del respectivo día. Por tanto, la matriz de datos que se conforma es de dimensión (38×24). La estimación de la matriz de covarianzas se puede realizar ya que $p < n$, sin embargo, se presume que la estimación tradicional no es la mejor puesto que, 38 observaciones para 24 variables parecen ser poca información; de hecho, si se tuviesen todas las 53 posibles observaciones para 24 variables también parecería poca información para tantas variables.

Usando los datos del $PM_{2.5}$ se emplea la estimación shrinkage que es numéricamente óptima. La función **cov.shrink** del paquete **corpcor** (R Core Team, 2014) en el software R permite calcular la estimación shrinkage $\mathbf{S}^*$ usando el procedimiento descrito en 3.2.5.

Por un lado, en la Tabla **4-1** se presentan las varianzas muestrales insesgadas, de aquí, se observa que posiblemente las varianzas poblacionales asociadas a las horas de los lunes son desiguales, sin embargo, se recuerda que las estimaciones no son muy confiables porque no se cuenta con un buen tamaño de muestra. Por otro lado, la estimación shrinkage de las varianzas intenta regularizarlas un poco hacia la varianza mediana (ver Tabla **4-2**). Respecto a las relaciones entre variables, la estimación tradicional tiene correlaciones más fuertes que la estimación shrinkage, además, hay presencia de correlaciones moderadas positivas, negativas y nulas (ver anexo 1). Estos resultados permiten confirmar que la matriz $\mathbf{\Sigma}$ propuesta que conduce a la simulación, se ha elegido de manera adecuada con el fin de conducir la simulación a un escenario real.

Tabla 4-1: Estimación tradicional de las varianzas para las 24 horas con los datos de $PM_{2.5}$ del día lunes

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
38.12	49.31	46.43	34.25	62.17	27.49	50.82	51.86	54.73	96.24	84.37	88.23
Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	Y_{21}	Y_{22}	Y_{23}	Y_{24}
83.80	100.09	110.9	109.47	83.99	39.97	51.68	31.76	32.07	61.38	28.61	49.63

Tabla 4-2: Estimación shrinkage de las varianzas para las 24 horas con los datos de $PM_{2.5}$ del día lunes

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
43.36	50.25	48.48	40.98	58.18	36.81	51.19	51.83	53.60	79.16	71.85	74.23
Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}	Y_{21}	Y_{22}	Y_{23}	Y_{24}
71.50	81.54	87.70	87.31	71.62	44.50	51.72	39.44	39.64	57.69	37.50	50.45

Al comparar ambas estimaciones de las varianzas y las correlaciones, se observa que la estimación shrinkage conserva la estructura pero con valores regularizados. Pero, hay una gran diferencia en el comportamiento de los valores propios asociados (ver Tablas **4-3** y **4-4**). Para la estimación tradicional, el número de condición es 996.3303 y para la estimación shrinkage es de 61.81767. Como en los resultados se evidenció que la estimación shrinkage mejora la estimación de la matriz de covarianzas, entonces al parecer, para los datos de $PM_{2.5}$ la matriz de covarianza poblacional posiblemente sea medianamente bien condicionada.

Tabla 4-3: Valores propios asociados a la estimación tradicional de la matriz de covarianzas

732.0901220	168.3304179	120.8483311	74.0443543	60.1154851	56.1230494	42.8943032	33.2124002
29.0211371	24.6436167	20.8224723	19.8059149	15.8577159	13.4081251	10.6216463	9.7187850
9.0143339	7.8545555	6.6954716	3.9475392	3.0201075	2.4972704	1.2356690	0.7347865

Tabla 4-4: Valores propios asociados a la estimación shrinkage de la matriz de covarianzas

565.648230	133.944810	114.500456	69.960320	59.131783	54.942367	44.663434	37.600472
33.350555	29.952218	27.198779	26.176831	22.836118	21.656962	19.133622	17.467720
16.585758	16.217514	14.879426	12.549490	11.736455	11.342725	9.908930	9.150267

Los seis días restantes tienen asociadas sus respectivas estimaciones shrinkage de la matriz de covarianzas que se utilizarán en la prueba Shapiro Wilk modificada.

4.4.3. Shapiro-Wilk Generalizada modificada

Con el objetivo de probar que las horas de cada día de la semana proviene de una distribución normal 24-variada, se remplaza cada estimación de la matriz de covarianzas tradicional por la respectiva estimación shrinkage en la construcción estadístico de la prueba Shapiro-Wilk Generalizada.

De manera análoga, para cada día de la semana se plantea:

Sea $\mathbf{Y_1}, ..., \mathbf{Y_n} \in \mathbb{R}^{24}$ una muestra de vectores aleatorios i.i.d de una población 24-dimensional.

Se desea probar que los vectores aleatorios siguen una distribución normal 24-variada con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas $\boldsymbol{\Sigma}$, para ello se plantean las siguientes hipótesis:

$H_0 = \mathbf{Y_1}, ..., \mathbf{Y_n}$ provienen de una población multinormalmente distribuida $N_{24}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$H_1 = \mathbf{Y_1}, ..., \mathbf{Y_n}$ no provienen de una población multinormalmente distribuida $N_{24}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

donde $N_{24}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota la función de densidad expresada en 2-1.

Particularmente, los valores del estadístico de cada prueba (tradicional y modificada) para el día **lunes** son: 0.9655671 y 0.9601002 con valores p de 0.09169053 y 0.001373745, respectivamente. Haciendo el contraste con un $\alpha = 0.05$, la prueba tradicional concluye que no hay suficiente evidencia estadística para afirmar que las horas del lunes no siguen una distribución normal 24-variada; en cambio, la prueba modificada concluye que las horas del lunes no siguen una distribución normal 24-variada.

Finalmente, en la Tabla **4-5** se presenta el consolidado de resultados para los siete días de la semana. Mayoritariamente, la prueba modificada rechaza la hipótesis nula de multinormalidad con un poco más de fuerza que la prueba tradicional. Teniendo en cuenta que para tamaños de muestra pequeños la prueba modificada mostró más probabilidad de tomar la decisión correcta cuando H_0 era falsa y mostró ser igual de confiable a la prueba tradicional cuando H_0 era verdadera, se podría concluir que, no hay suficiente evidencia estadística para afirmar que el vector de horas para el día domingo no sigue una distribución normal 24-variada. Además, el vector de horas para el resto de días de la semana, no siguen un distribución normal 24-variada.

Cuando $\alpha = 0.01$, el vector de horas del viernes no rechaza H_0, sin embargo, la brecha entre 0.01 y el valor p es demasiado pequeña, por lo que, la decisión no se puede tomar con mucha certeza.

Día	(n, p)	Valor p W^*	Valor p W^{SH}	Rechaza H_0 ($\alpha = 0.01$) Tradicional / Modificada	Rechaza H_0 ($\alpha = 0.05$) Tradicional / Modificada
Lunes	$(38, 24)$	0.09169053	0.001373745	No/Sí	No/Sí
Martes	$(39, 24)$	4.392377e-08	2.089207e-10	Sí/Sí	Sí/Sí
Miércoles	$(38, 24)$	5.573273e-07	4.361044e-23	Sí/Sí	Sí/Sí
Jueves	$(37, 24)$	0.0008956209	3.036141e-16	Sí/Sí	Sí/Sí
Viernes	$(39, 24)$	0.0766598	0.01573574	No/No	No/Sí
Sábado	$(45, 24)$	0.001882453	9.563321e-05	Sí/Sí	Sí/Sí
Domingo	$(38, 24)$	0.2561594	0.2968071	No/No	No/No

Tabla 4-5: Resultados de las pruebas SWG tradicional y modificada

5 Conclusiones y recomendaciones

5.1. Conclusiones

1. De acuerdo con literatura encontrada no se identificaron alternativas para abordar una problemática tan común como lo es tener una cantidad similar de variables y observaciones, por lo tanto, incorporar la estimación shrinkage de la matriz de covarianzas en la prueba Shapiro-Wilk es una alternativa novedosa y además intuitivamente sencilla de entender.

2. Aunque el estadístico de la prueba Shapiro Wilk Generalizada es un promedio de estadísticos univariados, la incorporación de la estimación shrinkage no intervino en los p estadísticos univariados de forma directa puesto que la intervención tiene lugar antes, específicamente en la estandarización de la matriz de datos a la cual posteriormente se le calcula a cada columna el estadístico de la prueba Shapiro Wilk univariada para finalmente construir el estadístico multivariado.

3. Se esperaba que si la prueba modificada funcionaba mejor, entonces la distribución nula empírica del estadístico tradicional tuviera un dominio de valores más alejados de uno y en la prueba modificada fueran más cercanos a uno, sin embargo, las distribuciones nulas empíricas de ambos estadísticos (modificado y tradicional) están superpuestas (ver densidades en la Figura **3-2**), por lo que, al momento de tomar la decisión de rechazar H_0 o no con la prueba modificada, se utilizan los percentiles de la prueba tradicional.

4. La modificación realizada en el estadístico de prueba está orientado al problema de $n \cong p$. Si bien es cierto que para muestras provenientes de poblaciones multinormales y no multinormales, los valores p asociados a los estadísticos de ambas pruebas presentan mayor variabilidad cuando el tamaño de muestra (n) es pequeño para un número de variables fijo (p); sin embargo, comparando de forma descriptiva la dispersión de los valores p de ambas pruebas, se observó que la prueba modificada es más precisa en sus resultados, lo cual significa que presenta menor incertidumbre al momento de tomar la decisión.

5. La prueba Shapiro Wilk Generalizada modificada funciona mucho mejor para tamaños de muestra pequeños con un número de variables fijo debido a que la estimación shrinkage estima mejor a la matriz de covarianzas poblacional que la estimación tradicional.

Cabe resaltar que las distribuciones para generar datos provenientes de muestras no-multinormales se han propuesto estratégicamente para "confundir" ambas pruebas; es decir, las distribuciones establecidas comparten algunas similitudes propias de una distribución multinormal estándar, por ejemplo la simetría (distribución t multivariada), variables marginales normales estándar (distribución $\chi^2_{(1)} + nor$), variables asimétricas pero desplazadas hacia cero (distribución desplazada $\chi^2_{(1)}$) y centramiento alrededor de cero (distribución Qu-Lui-Zhang), en todos los escenarios la prueba modificada presenta un mejor desempeño.

6. Con los datos de $PM_{2.5}$: mayoritariamente, la prueba modificada rechaza la hipótesis nula de multinormalidad con valores-p más cercanos a cero que la prueba tradicional. Teniendo en cuenta que para tamaños de muestra pequeños la prueba modificada mostró más probabilidad de tomar la decisión correcta cuando H_0 era falsa y mostró ser igual de confiable a la prueba tradicional cuando H_0 era verdadera, se podría concluir que, no hay suficiente evidencia estadística para afirmar que el vector de horas para el día domingo no sigue una distribución normal 24-variada. Además, el vector de horas para el resto de días de la semana, no siguen un distribución normal 24-variada.

7. Existen problemas en que no es posible que la cantidad de observaciones sea considerablemente mayor que la cantidad de variables, así se tengan todos los recursos necesarios para tomar las observaciones, la definición del problema no lo permite. En particular, el problema de $PM_{2.5}$ aunque los datos estén completos ($n = 53$ y $p = 24$), su diferencia no es despreciable y dado que los niveles de contaminación durante un año a otro pueden variar por las distintas políticas, no se sugiere extender el problema a dos o más años. Por tal motivo, la modificación de la prueba Shapiro Wilk Generalizada seguirá siendo útil aún cuando los datos de $PM_{2.5}$ estén completos.

5.2. Recomendaciones

- Como la calidad del estimador al parecer no tiene una incidencia fuerte en la decisión que toman ambas pruebas cuando los datos provienen de una distribución multinormal, se podría tener en cuenta que para el cálculo del estadístico además del estimador de la matriz de covarianzas también se usa la matriz de datos y es posible que la forma directa de los datos prevalezca indicando la normalidad y conllevando a la decisión correcta. Una posible forma de comprobar lo anterior es, aplicar la prueba utilizando los datos multinormales pero variando un poco la estimación de la matriz de covarianzas, por ejemplo, hacer que las entradas sean la media y variar dichos valores en un rango uniforme; así, en caso de evidenciar que la prueba sigue detectando la normalidad, permitiría afirmar con más certeza que la forma directa de los datos es más incidente para la decisión que toma la prueba de hipótesis.

- Villaseñor y González (2009) muestran a través de simulación de Monte Carlo que la generalización propuesta por ellos presenta un mejor rendimiento de potencia en comparación a las pruebas de Mardia, Henze-Zirkler y Srivastava-Hui, frente a una amplia gama de escenarios considerados. ¿Se podría pensar entonces que la prueba Shapiro-Wilk Generalizada modificada es más potente que todas?, se sugiere estudiar en un futuro el desempeño de la prueba con respecto a otras pruebas de normalidad multivariante existentes definiendo una serie de escenarios estratégicos que se apliquen a todas las pruebas. Los escenarios de evaluación propuestos en este trabajo de grado podrían servir como base.

- En el caso práctico de los datos de $PM_{2.5}$ se calculó la estimación shrinkage numéricamente óptima; sin embargo, la determinación de la matriz objetivo tiene distintas opciones, en la Tabla **3-5** se definen seis objetivos de contracción. Para elegir la matriz objetivo, el investigador lo puede abordar de dos formas: la primera es usar la matriz objetivo con estructura $\mathbf{D}$ para calcular la estimación shrinkage numéricamente óptima ya deducida y utilizada en este trabajo. La segunda es basarse en resultados descriptivos similares a los presentados por Villareal y Arroyave (2020) en la sección 8, donde se muestran las estadísticas descriptivas de los datos horarios para cada día, específicamente la Figura 8-4: Gráficos de caja por hora y día de la semana y la Figura 8-5: Gráficos de caja por hora y día de la semana de la página 37. Además, un cuestionamiento podría ser si en horas pico se incrementa la variabilidad o solo se incrementa el nivel de contaminación por la dinámica de la zona. Lo anterior permitiría al investigador determinar cuál de los objetivos presentados en la Tabla **3-5** es el más adecuado para este problema. Posteriormente debe proceder con la deducción del nivel de contracción óptimo para la estructura de la matriz objetivo que eligió y finalmente, calcular la estimación shrinkage con los elementos mencionados. Cabe resaltar que el investigador debe asegurarse que la estimación resultante es positiva definida.

- De la recomendación anterior también surge la siguiente pregunta: para un caso de estudio en particular, ¿el desempeño de la prueba modificada se ve afectada si se selecciona otra matriz objetivo?

  De acuerdo con la conclusión en donde se atribuye el mejoramiento en el desempeño de la prueba modificada a que la estimación shrinkage estima mejor a la matriz de covarianzas poblacional, si se elige una matriz objetivo que sea adecuada al comportamiento de los datos del caso de estudio y de esta se puede derivar un estimador de contracción óptimo para posteriormente construir una estimación shrinkage positiva definida, entonces, no se presume una afectación negativa en el desempeño de la prueba, por el contrario, podría revelar o refutar la normalidad con más certeza.

  La estimación shrinkage numéricamente óptima con los elementos establecidos en este trabajo actúa como una estimación balanceada que se podría usar de forma general.

No obstante, en una futura investigación se podría estudiar la sensibilidad de la prueba abordando las matrices objetivo establecidas en la Tabla **3-5**.

- Un trabajo futuro podría estudiar si los percentiles de las distribuciones empíricas nulas de ambos estadísticos (tradicional W^* y modificado W^M) presentan diferencias significativas que produzcan decisiones contrarias sobre la población.

6 Anexos

6.1. Anexo 1

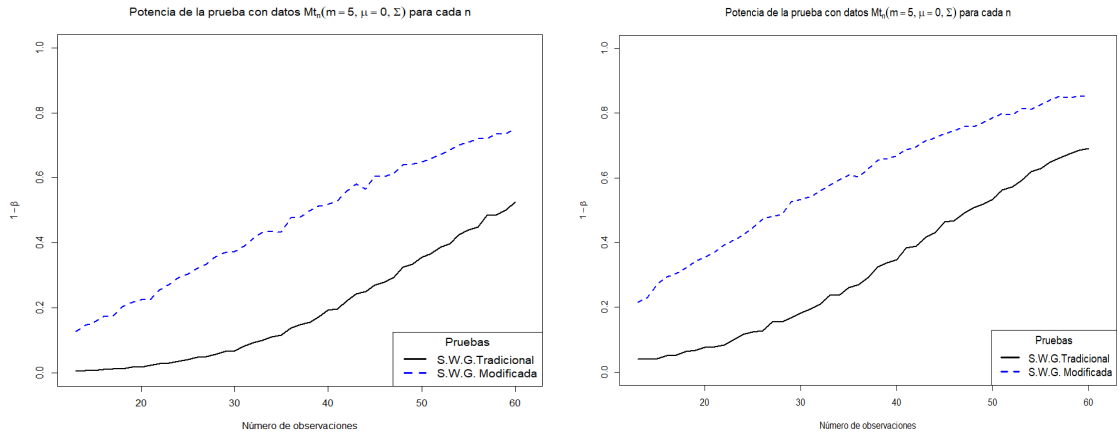
Tabla 6-1: Estimación tradicional de la matriz de correlaciones de los datos de $PM_{2.5}$ para el día lunes (reducida a 12 horas por espacio)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	1	0.655	0.720	0.684	0.619	0.692	0.604	0.611	0.546	0.596	0.704	0.561
x_2	0.655	1	0.590	0.711	0.660	0.770	0.716	0.555	0.586	0.542	0.640	0.500
x_3	0.720	0.590	1	0.752	0.630	0.714	0.580	0.575	0.466	0.567	0.657	0.541
x_4	0.684	0.711	0.752	1	0.718	0.704	0.576	0.633	0.638	0.630	0.684	0.620
x_5	0.619	0.660	0.630	0.718	1	0.632	0.724	0.627	0.724	0.626	0.688	0.672
x_6	0.692	0.770	0.714	0.704	0.632	1	0.679	0.546	0.552	0.565	0.719	0.584
x_7	0.604	0.716	0.580	0.576	0.724	0.679	1	0.738	0.716	0.730	0.804	0.658
x_8	0.611	0.555	0.575	0.633	0.627	0.546	0.738	1	0.801	0.658	0.604	0.648
x_9	0.546	0.586	0.466	0.638	0.724	0.552	0.716	0.801	1	0.715	0.731	0.647
x_{10}	0.596	0.542	0.567	0.630	0.626	0.565	0.730	0.658	0.715	1	0.878	0.721
x_{11}	0.704	0.640	0.657	0.684	0.688	0.719	0.804	0.604	0.731	0.878	1	0.763
x_{12}	0.561	0.500	0.541	0.620	0.672	0.584	0.658	0.648	0.647	0.721	0.763	1

Tabla 6-2: Estimación shrinkage de la matriz de correlaciones de los datos de $PM_{2.5}$ para el día lunes (reducida a 12 horas por espacio)

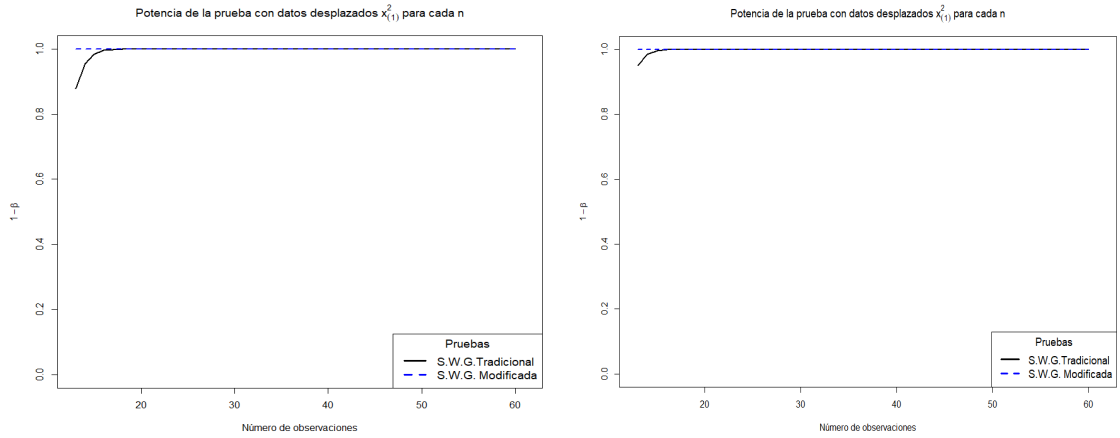
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
x_1	1	0.544	0.598	0.568	0.513	0.574	0.501	0.507	0.453	0.494	0.585	0.466
x_2	0.544	1	0.490	0.590	0.548	0.639	0.594	0.460	0.486	0.450	0.531	0.415
x_3	0.598	0.490	1	0.624	0.523	0.593	0.481	0.477	0.386	0.470	0.546	0.449
x_4	0.568	0.590	0.624	1	0.596	0.585	0.478	0.525	0.530	0.523	0.568	0.514
x_5	0.513	0.548	0.523	0.596	1	0.524	0.601	0.520	0.601	0.520	0.571	0.558
x_6	0.574	0.639	0.593	0.585	0.524	1	0.564	0.453	0.459	0.469	0.597	0.485
x_7	0.501	0.594	0.481	0.478	0.601	0.564	1	0.613	0.594	0.606	0.667	0.546
x_8	0.507	0.460	0.477	0.525	0.520	0.453	0.613	1	0.665	0.546	0.501	0.538
x_9	0.453	0.486	0.386	0.530	0.601	0.459	0.594	0.665	1	0.593	0.607	0.537
x_{10}	0.494	0.450	0.470	0.523	0.520	0.469	0.606	0.546	0.593	1	0.729	0.599
x_{12}	0.585	0.531	0.546	0.568	0.571	0.597	0.667	0.501	0.607	0.729	1	0.634
x_{13}	0.466	0.415	0.449	0.514	0.558	0.485	0.546	0.538	0.537	0.599	0.634	1

6.2. Anexo 2



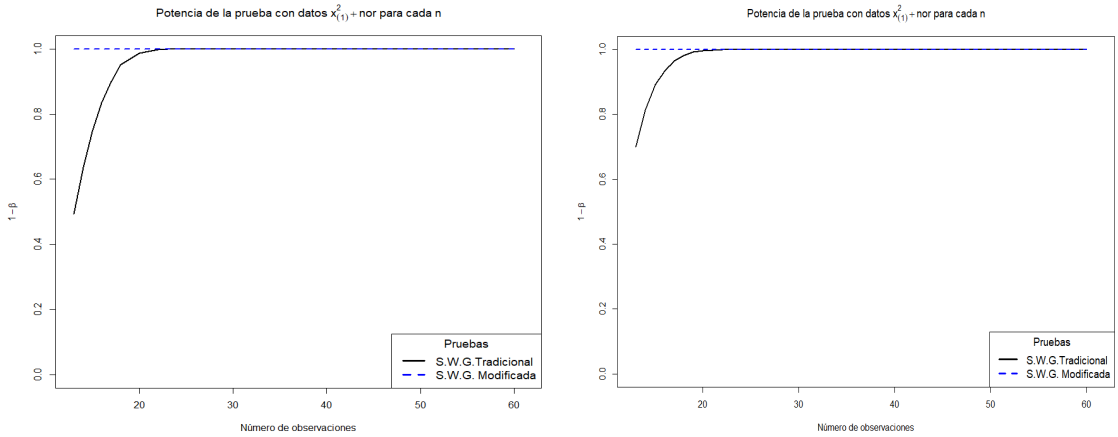
(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 6-1: Potencia de la prueba para datos multi-t $(m = 5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma})$ con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) y matriz de covarianzas mal condicionada para cada n



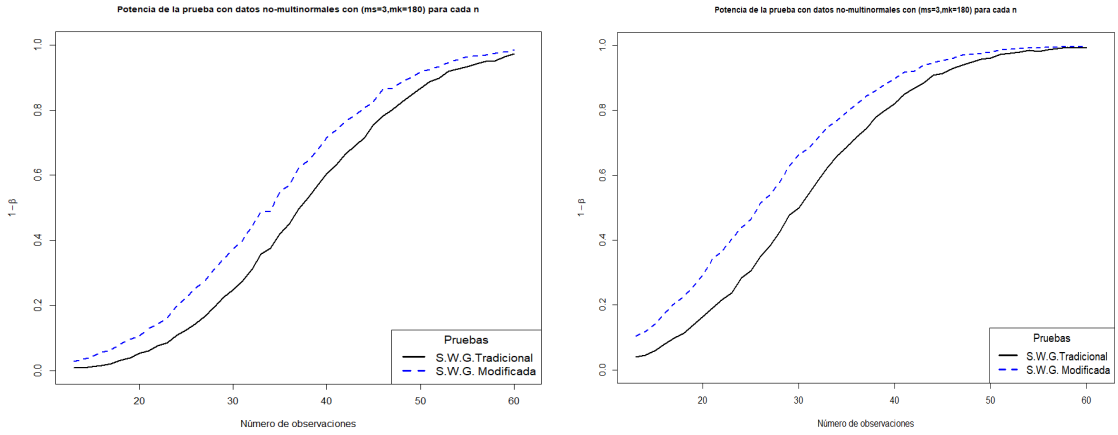
(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 6-2: Potencia de la prueba para datos desplazados $\chi^2_{(1)}$ con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) y matriz de covarianzas mal condicionada para cada n



(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 6-3: Potencia de la prueba para datos $\chi^2_{(1)} + nor$ con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) y matriz de covarianzas mal condicionada para cada n



(a) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.01$ (b) P(rechazar $H_0|H_0$ falsa) con $\alpha = 0.05$

Figura 6-4: Potencia de la prueba para datos distruidos Qu-Lui-Zhang ($ms = 3, mk = 180$) con $\alpha = 0.01$ (a) y $\alpha = 0.05$ (b) y matriz de covarianzas mal condicionada para cada n

6.3. Anexo 3

	Matriz de cov bien condicionada		Matriz de cov mal condicionada	
n	d Tradicional	d Shrinkage	d Tradicional	d Shrinkage
13	66164.32188	6.515154	8374.5496	58.58006
14	1748.39652	1.165964	4795.4803	141.88652
15	1011.80305	3.312991	4928.9387	72.13192
16	32515.07100	2.348887	3816.8180	5.89380
17	619.03306	1.950676	4823.1003	102.78245
18	224.17973	7.802973	1544.1377	52.91328
19	748.93427	3.517305	2656.5097	43.70726
20	847.38206	3.250859	3303.1946	50.58815
21	220.05102	3.170080	3148.6518	20.80482
22	161.94534	4.768991	2316.3623	36.06676
23	124.17668	2.403924	2454.5681	29.42675
24	141.07463	7.204119	1293.3237	84.52869
25	201.40796	6.713500	1899.2297	29.17516
26	290.23060	6.452254	1440.8033	88.70766
27	154.94797	7.170577	1440.3707	59.56364
28	352.86338	7.070537	755.9234	104.24574
29	402.09024	7.321256	211.8270	93.58329
30	113.52760	3.077809	1093.9490	143.55465
31	381.99627	3.432459	710.4114	135.91484
32	126.34038	3.658142	498.2254	82.02066
33	63.76306	3.857844	949.1875	80.67952
34	98.20501	6.427082	968.9002	65.56041
35	89.94543	5.790306	389.1710	113.80564
36	94.91604	8.653448	425.6834	91.79250
37	174.73547	2.231688	707.7478	103.58318
38	64.87853	3.897511	501.1211	19.40752
39	55.75157	7.786264	376.8876	102.18088
40	35.82634	6.237939	462.1078	108.31821
41	72.65130	7.480930	729.3633	139.77876
42	92.33138	4.192417	526.8367	60.46151
43	70.88752	9.644828	458.2112	155.06438
44	62.78013	6.445229	297.0623	110.86490
45	42.96648	5.555063	349.9162	159.26513
46	60.42833	5.833641	499.9348	122.67452
47	42.02948	5.426978	393.1707	141.00712
48	127.31386	4.703539	374.7781	53.20199
49	109.93819	7.177492	473.7343	41.81758
50	53.87325	7.481668	767.4709	76.33699
51	65.83064	7.392246	550.9606	163.03442
52	85.00236	7.007626	601.8495	136.47051
53	38.11491	7.285067	588.6739	168.98524
54	38.99471	6.308009	439.6507	151.60411
55	49.94932	3.006003	426.8620	106.10833
56	64.53361	3.401416	564.4587	89.27875
57	38.78373	8.024888	281.9872	100.07082
58	36.78496	6.958596	291.0929	82.05396
59	41.94763	9.109806	274.8917	153.23739
60	56.23637	8.076561	485.7679	96.78075

Tabla 6-3: Números de condición asociados al estimador tradicional y shrinkage de muestras generadas a partir de una población t-multivariada ($m = 5, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma}$) con parámetro $\boldsymbol{\Sigma}$ bien y mal condicionado

	Matriz de cov bien condicionada		Matriz de cov mal condicionada	
n	d Tradicional	d Shrinkage	d Tradicional	d Shrinkage
13	3996.59605	5.528708	34003.2680	128.16304
14	20947.85287	3.638267	59852.8509	53.61412
15	747.45632	4.438182	1851.3567	105.17858
16	704.85885	4.892451	1084.8590	63.19114
17	226.04760	3.580394	3843.9483	208.44736
18	181.56111	8.934909	2543.7145	68.26595
19	132.91613	11.786608	1104.9116	138.93473
20	122.79551	9.774689	877.8560	94.19491
21	146.17686	3.838457	727.0219	130.73703
22	152.99754	18.864189	2251.8708	86.66979
23	138.77972	3.535417	949.2444	124.80506
24	216.79188	4.252345	915.1765	67.73092
25	97.56187	4.495726	524.2296	39.56101
26	74.59386	4.871276	983.9164	102.91206
27	87.53068	10.926174	336.8892	56.31937
28	110.40503	5.179591	440.6738	115.70950
29	50.52966	7.692561	863.9977	99.01514
30	36.48369	4.213605	353.9478	105.57822
31	51.74739	5.727304	457.5018	126.12673
32	112.17187	11.594421	424.0903	73.58490
33	53.71218	10.972742	361.9715	97.59123
34	65.08654	5.447251	542.6471	114.64521
35	49.72902	5.271849	710.5436	151.24621
36	67.88281	11.671788	344.0113	121.91217
37	37.71116	6.282645	350.8183	96.90948
38	53.90353	10.213931	493.4889	181.08913
39	43.28527	5.066463	372.8448	132.72389
40	56.05800	8.514721	424.1461	103.49400
41	53.84242	9.955426	238.5569	108.74594
42	93.96780	14.192071	269.9628	94.94873
43	52.72132	7.816483	493.9956	158.73414
44	70.59326	14.011316	278.0353	133.84295
45	43.03635	7.646340	361.2317	100.57210
46	55.25553	14.699295	414.2668	197.35602
47	49.80589	10.532587	477.6333	157.02122
48	50.58262	7.487938	375.6457	195.92287
49	29.25466	8.332361	398.1458	192.46916
50	38.20361	9.321290	347.9309	155.45745
51	33.50765	8.715641	341.2729	126.34776
52	72.85652	14.888721	316.4638	149.83600
53	37.73973	9.347352	310.0154	134.33847
54	39.73613	6.943937	214.1654	136.31827
55	39.62569	8.164799	266.8625	153.19676
56	35.51260	8.018282	215.4748	125.91438
57	34.91695	6.505029	257.3528	92.65925
58	45.56848	14.482526	296.7724	110.50246
59	61.96343	13.507551	297.5469	152.01018
60	48.92941	11.998944	407.1237	183.47487

Tabla 6-4: Números de condición asociados al estimador tradicional y shrinkage de muestras generadas a partir de una población multinormal ($p = 12, \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma}$) con parámetro $\boldsymbol{\Sigma}$ bien y mal condicionado

Referencias

Arteaga, F., y Ferrer, A. (2013). Building covariance matrices with the desired structure. *Chemometrics and Intelligent Laboratory Systems*, *127*, 80–88.

Balakrishnan, N., y Clifford, A. (1991). *Order statistics and inference. estimation methods.* Statistical modeling and decision science.

Caicedo, A., y Jimenez, C. (2016). *Imputación basada en análisis de datos funcionales de observaciones faltantes de contaminación atmosférica por partículas finas suspendidas en el aire ($p.m_{2,5}$).* Trabajo de grado en Estadística. Universidad del Valle.

David, J. C., y MacKay. (2003). *Information theory, inference, and learning algorithms.* Cambridge University Press.

Díaz, L. G., y Morales, M. A. (2012). *Estadística multivariada: inferencia y métodos.* Universidad Nacional de Colombia.

Hain, J. (2010). *Comparison of common tests for normality.* Julius Maximilians Universität Würzburg.

Henze, N., y Zirkler, B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, *19*, 3595-3617.

Kotz, S., y Nadarajah, S. (2004). *Multivariate t-distributions and their applications.* Cambridge University Press.

Ledoit, O., y Wolf, M. (2003). Honey, i shrunk the sample covariance matrix. *Journal Portafolio Management*, *30*, 110–119.

Ledoit, O., y Wolf, M. (2004). A well-conditioned estimator forlarge-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*, 365–411.

Liang, J., Li, R., Fang, H., y Fang, K. T. (2000). Testing multinormality based on low-dimensional projection. *Journal of Statistical Planning and Inference*, *86*, 129—141.

Liang, J., Tang, M. L., y Chan, P. S. (2009). A generalized shapiro–wilk w statistic for testing high-dimensional normality. *Computational Statistics and Data Analysis*, *53*, 3883—3891.

Läuter, J., Glimm, E., y Kropf, S. (1996). New multivariate tests for data with an inherent structure. *Biometrical Journal*, *38*, 5–23.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika Journal*, *57*, 519–530.

Mazhdrakov, M., Benov, D., y Valkanov, N. (2018). *The monte carlo method. engineering applications.* ACMO Academic Press.

McMillan, S. C., y Weitzner, M. A. (2003). Methodologic issues in collecting data from debilitated patients with cancer near the end of life. *National Library of Medicine*, *30*, 123–129.

Olaya, J. (2019). *Imputación de datos faltantes de $p.m_{2,5}$ basada en un proceso gaussiano.* Proyecto de INFERIR. Universidad del Valle.

OMS. (2018). *9 out of 10 people worldwide breathe*

polluted air, but more countries are taking action. https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action.

Opgen-Rhein, R., y Strimmer, K. (2007). Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, *6*, 9.

Otero, A., y Presiga, M. (2019). *Evaluación de un método de imputación basado en el análisis de datos funcionales para los registros de $p.m_{2,5}$ en la ciudad de cali.* Trabajo de grado en Estadística. Universidad del Valle.

Parrish, R. (1992). Computing expected values of normal order statistics. *Communications in Statistics - Simulation and Computation*, *21*, 57–70.

Porras, J. (2016). Comparación de pruebas de normalidad multivariada. *Anales Científicos*, *77*, 141–146.

Python. (1991). Python package index [Manual de software informático]. PyPI. Descargado de https://pydata.org/

Qu, W., Liu, H., y Zhang, Z. (2019). A method of generating multivariate non-normal random numbers with desired multivariate skewness and kurtosis. *Behavior Research Methods*, *52*, 939–946.

R Core Team. (2014). R: Efficient estimation of covariance and (partial) correlation [Manual de software informático]. corpcor. Descargado de https://www.rdocumentation.org/packages/corpcor/versions/1.6.10

R Core Team. (2020). R: A generator of multivariate non-normal random numbers [Manual de software informático]. mnonr. Descargado de https://rdrr.io/cran/mnonr/src/R/unonr.R

R Core Team. (2021). R: Multivariate normal and t distributions [Manual de software informático]. mvtnorm. Descargado de https://www.rdocumentation.org/packages/mvtnorm/versions/1.1-3

Royston, J. P. (1992). Approximating the shapiro-wilk w-test for non-normality. *Statistics and Computing*, *2*, 117–119.

Schäfer, J., y Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*, 32.

Shapiro, S. S., y Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika Journal*, *52*, 591–611.

Srivastava, M. S., y Hui, T. K. (1987). On assessing multivariate normality based on shapiro-wilk w statistic. *Statistics and Probability Letter*, *5*, 15–18.

Villareal, A., y Arroyave, E. (2020). *Propuesta de un método de imputación basado en la distribución normal multivariada para los registros de $p.m_{2,5}$ en la ciudad de cali.* Trabajo de grado en Estadística. Universidad del Valle.

Villaseñor, J., y González, E. (2009). A generalization of shapiro–wilk's test for multivariate

normality. *Communications in Statistics—Theory and Methods*, *38*, 1870-–1883.