

## **ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**

**Estudio de la afectación de la Probabilidad de error tipo I para  
comparar medias, cuando se pre-testea el supuesto de normalidad bajo  
los enfoques tradicionales y de equivalencia**

**MARÍA DE LOURDES PALACIOS ROBALINO**

Trabajo de Titulación modalidad Proyectos de Investigación y Desarrollo, presentado  
ante el Instituto de Posgrado y Educación Continua de la ESPOCH, como requisito  
parcial para la obtención del grado de:

**MAGÍSTER EN MATEMÁTICA MENCIÓN MODELACIÓN Y  
DOCENCIA**

Riobamba - Ecuador

Julio 2022

**©2022, María de Lourdes Palacios Robalino**

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento, siempre y cuando se reconozca el Derecho de Autor.



## ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

EL TRIBUNAL DEL TRABAJO DE TITULACIÓN CERTIFICA QUE:

**El Trabajo de titulación modalidad proyecto de Investigación y desarrollo titulado:** Estudio de la afectación de la Probabilidad de error tipo I para comparar medias, cuando se pre-testea el supuesto de normalidad bajo los enfoques tradicionales y de equivalencia, de responsabilidad de la señorita María de Lourdes Palacios Robalino, ha sido minuciosamente revisado y se autoriza su presentación.

ING. LUIS EDUARDO HIDALGO ALMEIDA; Ph.D.  
**PRESIDENTE**

---

ING. PABLO JAVIER FLORES MUÑOZ; Mag.  
**DIRECTOR**

---

ING. WILSON JAVIER VILLAGRÁN CÁCERES; Mag.  
**MIEMBRO**

---

ING. FABIAN PATRICIO LONDO YACHAMBAY; Mag.  
**MIEMBRO**

---

Riobamba, julio 2022

## **DERECHOS INTELECTUALES**

Yo, María de Lourdes Palacios Robalino, declaro que soy responsable de las ideas, doctrinas y resultados expuesto en el Trabajo de Titulación modalidad Proyectos de Investigación y Desarrollo y que el patrimonio intelectual generado por la misma pertenece exclusivamente a la Escuela Superior Politécnica de Chimborazo.

---

MARÍA DE LOURDES PALACIOS ROBALINO

No. Cédula 060386218-6

## **DECLARACIÓN DE AUTENTICIDAD**

Yo, María de Lourdes Palacios Robalino, declaro que el presente trabajo de Titulación Modalidad de Proyectos de Investigación y Desarrollo, es de mi auditoria y que los resultados del mismo son auténticos y originales. Los textos constantes en el documento que provienen de otra fuente están debidamente citados y referenciados.

Como autor, asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Titulación de Maestría.

---

**MARÍA DE LOURDES PALACIOS ROBALINO**

No. Cédula 060386218-6

## **DEDICATORIA**

A Dios que en cada momento está conmigo, por ser la luz en mi camino y por darme la sabiduría, fortaleza para alcanzar mis objetivos.

A mis padres Mario y María de Lourdes quienes día a día me incentivaron, por su paciencia, amor y apoyo incondicional en cada etapa de mi vida, a mis hermanos por creer en mí, ya que con su apoyo me han permitido llegar a cumplir un sueño más.

A mi niña bonita Maite quien es mi inspiración y fortaleza para luchar y conseguir mis metas.

*María de Lourdes Palacios*

## **AGRADECIMIENTO**

A Dios por permitirme culminar una meta más en mi vida.

A mis padres Mario Palacios y María de Lourdes Robalino que han sido durante toda mi vida un apoyo fundamental en todos los aspectos.

Al Ing. Eduardo Villota por todo el apoyo que me brindo, motivación para seguir estudiando, por impulsarme al desarrollo de mi formación profesional.

Al Ing. Pablo Flores por compartir sus conocimientos y orientarme con paciencia durante todo el desarrollo de la tesis.

Al Ing. Javier Villagrán y al Ing. Fabian Londo quienes aportaron con sus ideas para encaminar correctamente esta investigación.

A mis amigos y conocidos que desde el primer momento me brindaron todo el apoyo, ánimo y compañía en las diferentes etapas de mi vida sin ningún interés.

*María de Lourdes Palacios*

## TABLA DE CONTENIDO

<b>RESUMEN .....</b>	<b>xiv</b>
<b>SUMMARY .....</b>	<b>xv</b>
<b>CAPÍTULO I.....</b>	<b>1</b>
<b>1 INTRODUCCIÓN.....</b>	<b>1</b>
1.1 Antecedentes .....	1
1.2 Formulación del problema.....	3
1.3 Preguntas directrices o específicas de la investigación .....	3
1.4 Justificación de la Investigación .....	4
1.5 Objetivos .....	5
1.5.1 <i>Objetivo General</i> .....	5
1.5.2 <i>Objetivos Específicos</i> .....	5
1.6 Hipótesis.....	5
<b>CAPÍTULO II .....</b>	<b>6</b>
<b>2 MARCO TEÓRICO.....</b>	<b>6</b>
2.1. Antecedentes del problema.....	6
2.2. Bases Teóricas .....	7
2.2.1. <i>Pruebas de Hipótesis para comparación de medias</i> .....	7
2.1.2 <i>Distribución muestral</i> .....	8
2.1.3 <i>Test de hipótesis para comparar medias</i> .....	15
2.1.4 <i>Test para verificar el supuesto de normalidad</i> .....	18
2.2 Método de Fleishman .....	26
2.3 Potencia de una prueba.....	28
<b>CAPÍTULO III.....</b>	<b>30</b>
<b>DESARROLLO DE LA INVESTIGACIÓN .....</b>	<b>30</b>
3.1 Metodología .....	30
3.1.1 <i>Tamaños de las muestras</i> .....	31
3.1.2 <i>Caracterización de funciones</i> .....	32
3.1.2.1 <i>Función rnonorm</i> .....	32



3.1.2.2	<i>Error de tipo I de t-Student (TIEP)</i> .....	33
3.1.2.2.1	<i>Función tiepT</i> .....	33
3.1.2.3	<i>Épsilon o Irrelevancia</i> .....	34
3.1.2.3.1	<i>Función opt_eps</i> .....	35
3.1.2.4	<i>Falta de ajuste</i> .....	36
3.1.2.4.1	<i>Función normequiv</i> .....	37
3.1.2.5	<i>Comparación de test</i> .....	37
3.1.2.5.1	<i>Función normalTest</i> .....	37
3.1.2.5.2	<i>Función rejectH0</i> .....	38
3.1.2.6	<i>Comparación de TIEP y POTENCIA</i> .....	39
3.1.2.6.1	<i>Función TIEP_POT</i> .....	39
	<b>CAPÍTULO IV</b> .....	<b>41</b>
<b>4.</b>	<b>ANÁLISIS Y RESULTADOS</b> .....	<b>41</b>
4.2.	Niveles de contaminación.....	41
4.3.	Épsilon Óptimo .....	41
4.4.	Potencia de las pruebas de normalidad .....	42
4.4.1.	<i>Resultados finales</i> .....	43
4.5.	Discusión.....	47
	<b>CONCLUSIONES</b> .....	<b>48</b>
	<b>RECOMENDACIONES</b> .....	<b>49</b>
	<b>BIBLIOGRAFÍA</b> .....	<b>50</b>
	<b>ANEXOS</b>	

## ÍNDICE DE TABLAS

Tabla 1-2.	Toma de decisiones Tipo de Error .....	14
Tabla 2-2.	Regla de decisión .....	16
Tabla 3-2.	Región de Rechazo de la prueba U de Mann Withney .....	16
Tabla 4-2.	Hipótesis alternativa de la prueba t -Student .....	17
Tabla 5-2.	Región de rechazo de la prueba t-Student a un nivel de significancia $\alpha$ . .....	18
Tabla 6-2.	Vectores mutuamente independientes .....	22
Tabla 7-2.	Datos, vectores y frecuencias .....	22
Tabla 8-2.	Diferencia entre la prueba de bondad de ajuste y falta de ajuste .....	25
Tabla 9-3.	Muestras balanceada y desbalanceada.....	32
Tabla 10-4.	Niveles de contaminación .....	41
Tabla 11-4.	Épsilon óptimo para cada par de muestra .....	41

## ÍNDICE DE GRÁFICOS

Gráfico 1-4.	Estimación de la potencia de una prueba de comparación de medias cuando se pre-testean el supuesto de normalidad usando distintos tamaños muestrales con una significancia $\alpha = 0.05$ . ....	43
Gráfico 2-4.	Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean el supuesto de normalidad, usando distintos tamaños muestrales con una significancia $\alpha = 0.05$ . ....	45

## ÍNDICE DE FIGURAS

Figura 1-2. Pretest Tradicional .....	6
Figura 2-2. p valor área bajo la curva .....	9
Figura 3-2. Región de rechazo y aceptación de $H_0$ en la distribución normal para dos colas..	11
Figura 4-2. Prueba de uno sola cola a la derecha.....	11
Figura 5-2. Prueba de uno sola cola a la izquierda .....	12
Figura 6-2. Prueba de dos colas.....	12
Figura 7-2. Distribución t-Student vs distribución normal .....	17
Figura 8-2. Función de probabilidad .....	19
Figura 9-2. Chi cuadrado es una prueba Unilateral .....	21
Figura 10-3. Proceso de hallar $\epsilon$ .....	34

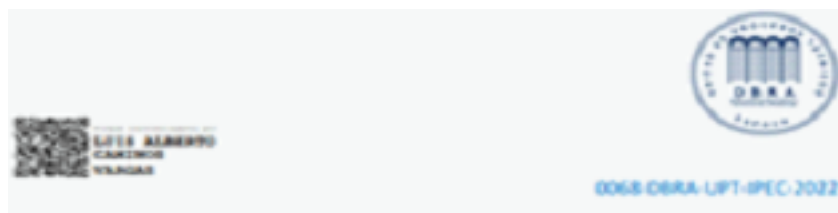
## ÍNDICE DE ANEXOS

ANEXO A.- Muestras contaminadas .....	54
ANEXO B.- Probabilidad de cometer un error de tipo I (TIEP).....	55
ANEXO C.- Épsilon o Irrelevancia .....	56
ANEXO D.- Falta de ajuste.....	58
ANEXO E.- Comparación de test .....	60
ANEXO F.- Comparación de TIEP y POTENCIA.....	62
ANEXO G.- Resultados de la simulación .....	63

## RESUMEN

La presente investigación tiene por objeto estudiar la afectación de la Probabilidad de error tipo I para comparar medias, cuando se pre-testea el supuesto de normalidad bajo los enfoques tradicionales y de equivalencia. Investigaciones realizadas previamente indican que el proceso de pre-testeo tradicional ocasiona una afectación en la probabilidad global de cometer un error de tipo I, por lo cual se desarrollara una nueva técnica basada en determinar límites de irrelevancia. En este trabajo se utilizó el lenguaje de programación R Studio, para diseñar un nuevo pretest iterativo de falta de ajuste denominado pretest de equivalencia para muestras con diferentes contaminaciones o alejamientos de la normalidad medibles, la contaminación se realiza utilizando los coeficientes de Fleishman empleando saltos iterativos de la asimetría y curtosis, considerando que las contaminaciones deben estar dentro del criterio de Cochran y de este modo se calcula el  $\epsilon$  óptimo el cual se utiliza para calcular la TIEP del test de equivalencia. La simulación consistió en aplicar 10000 repeticiones con el fin de estimar la probabilidad de error tipo I y la potencia de los pretest Chi Cuadrado y Equivalencia, así mismo de los test de comparación de medias t-Student y Wilcoxon. Los resultados obtenidos determinan que los procedimientos planteados son confiables para una prueba de comparación de medias, para los niveles de contaminación (0-6). Al realizar la comparación del test tradicional de bondad de ajuste y el pretest diseñado de falta de ajuste utilizado para probar normalidad, se demuestra que al pretestear bajo el enfoque de equivalencia se controla la probabilidad global de error tipo I y respecto a la potencia, el pretest de equivalencia es mejor que el pretest tradicional, por lo cual este enfoque se considera más confiable. Se recomienda utilizar el pretest de equivalencia desarrollado para probar el supuesto de normalidad, con el fin de reducir la afectación de la TIEP global en las pruebas de comparación de medias.

**Palabras Clave:** <CONTAMINACIÓN DE FLEISHMAN>, <BONDAD DE AJUSTE>, <PRUEBA DE EQUIVALENCIA >, <SIMULACIÓN>, <FUNCIÓN>, <ZONA DE INFERENCIA>



## SUMMARY

The present research aims to study the effect on the Probability of type I error for comparing means, when the assumption of normality is pre-tested under traditional and equivalence approaches. Previous research indicates that the traditional pre-testing process causes an effect on the overall probability of committing a type I error, so a new technique based on determining irrelevance limits will be developed. In this work the programming language R Studio was used to design a new iterative pretest of lack of fit called equivalence pretest for samples with different contaminations or measurable departures from normality, the contamination is performed using the Fleishman coefficients using iterative jumps of the skewness and kurtosis, considering that the contaminations must be within Cochran's criterion and thus calculating the optimal epsilon which is used to calculate the TIEP of the equivalence test. The simulation consisted of applying 10000 repetitions in order to estimate the probability of type I error and the power of the Chi-Square and Equivalence pretests, as well as of the t-Student and Wilcoxon mean comparison tests. The results obtained from the study conclude that the procedures proposed are reliable for a mean comparison test for contamination levels (0-6). When comparing the traditional goodness-of-fit test and the pretest designed for lack of fit used to test normality, it is shown that pretesting under the equivalence approach controls the overall probability of type I error, and with respect to power, the equivalence pretest is better than the traditional pretest, so this approach is considered more reliable. It is recommended to use the equivalence pretest developed to test the assumption of normality, in order to reduce the effect of the overall TIEP on the mean comparison tests.

**Keywords:** <FLEISHMASN CONTAMINATION>, <GOODNESS OF FIT>, <EQUIVALENCE TEST>, <SIMULATION>, <FUNCTION>, <INFERENCE ZONE>

# CAPÍTULO I

## 1 INTRODUCCIÓN

### 1.1 Antecedentes

Las pruebas estadísticas para comparar dos tratamientos son muy utilizadas en varios ámbitos de la investigación, (Dipp et al., 1992), (Alcántara Vasconcellos, 2010). En estadística inferencial (Wackerly et al., 2010; Ronald Walpole, Raymond Myers & Myers, 2012) estas pruebas funcionan bajo el cumplimiento de los supuestos de normalidad y homocedasticidad los cuales son comprobables a través de otras pruebas de hipótesis a las que denominaremos pretest como Shapiro Wilk, Anderson Darling, Kolmogorov Smirnov, etc (Anderson & Darling, 1954; Shapiro & Wilk, 1965; Smirnov, 1948) en el caso de normalidad o F, Bartlet, Cochran, Levene, etc (Brown & Forsythe, 1974; Snedecor & Cochran, 1989) en el caso de la homocedasticidad. Investigaciones realizadas demuestran que existe una afectación global de la probabilidad de cometer un error tipo I (TIEP-Probability of a type I error, por sus siglas en inglés) al realizar los pretest tradicionales.

Refiriéndose al supuesto de normalidad Rasch et al (2011), mediante un estudio de simulación sistemática demostró que pre-testear este supuesto, por medio de las pruebas tradicionales Shapiro Wilks, Kolmogorov Smirnov, etc, dicha prueba previa conduce a riesgos finales desconocidos en la probabilidad global de cometer un error de tipo I y tipo II, por lo que no aconseja utilizarlas.

De acuerdo con lo mencionado por Wellek (2010), el problema en la alteración de la TIEP no es ocasionado por el proceso de pretesteo, sino porque al parecer, la forma como tradicionalmente se prueban los supuestos no es adecuada. Tradicionalmente, los test para probar normalidad y homocedasticidad establecen en su hipótesis nula el cumplimiento perfecto de este supuesto, lo cual nos deja muchas dudas estructurales en el camino. Por ejemplo, respecto a la normalidad, George Box (1979) menciona *“en la naturaleza no existe una distribución perfectamente normal, sin embargo, con suposiciones normales, que se sabe que son falsas, a menudo se puede derivar resultados que coinciden, con una aproximación útil a los que se encuentran en el mundo real”*, lo cual nos lleva a cuestionar la validez de probar algo que teóricamente sabemos que no existe. Por otra parte, en una metodología tradicional, sabemos que, cuando la hipótesis nula  $H_0$  (hipótesis donde se evidencia el perfecto cumplimiento del supuesto), en el caso de homocedasticidad, al contrastar las hipótesis de manera tradicional ( $H_0: \sigma_1^2 = \sigma_2^2$  vs  $H_1: \sigma_1^2 \neq$



$\sigma_2^2$ ), se presenta un inconveniente lógico, no rechazar la hipótesis nula al comparar igualdad de varianzas no significa que exista homocedasticidad perfecta, no concluimos afirmando la veracidad de esta conjetura, lo cual no necesariamente prueba que se cumple el supuesto deseado, o como lo dijo Altman y Bland (1995): “Ausencia de evidencia no es evidencia de ausencia” en su lugar, lo único que podemos probar es que no existe evidencia suficiente para decir que la hipótesis alternativa  $H_1$  (hipótesis que niega el perfecto cumplimiento del supuesto) es falsa, así como lo menciona Wellek (Wellek, 2010) “Una diferencia no significativa no debe ser confundida con una significativa de homogeneidad”.

Con el fin de superar estas contradicciones teóricas, propias de las pruebas de hipótesis tradicionales, Wellek (2010), desarrolló un enfoque, basado en equivalencia. Según el autor, se entiende por equivalencia a *igualdad excepto para desviaciones irrelevantes*, las cuales están dadas por un límite de equivalencia, el cual representa un umbral alrededor del cual es permitido desviarse sin causar efectos significativos. Además, es importante mencionar que este planteamiento tiene una estructura inversa a los test de hipótesis tradicionales, de tal manera que, en la hipótesis alternativa se establece la equivalencia, lo cual permite superar las contradicciones mencionadas en el párrafo anterior, bajo este enfoque, rechazar  $H_0$ , nos da la evidencia necesaria para afirmar que el supuesto se cumple.

Aunque el enfoque de equivalencia, teóricamente supera las dificultades presentadas en el enfoque teórico, el inconveniente principal es que no existe, debido a la naturaleza propia de los test, un criterio técnico que nos permita establecer un umbral de equivalencia apropiado, el valor de esta irrelevancia, tal como menciona el autor queda a juicio de la experiencia y necesidades del investigador, lo cual, a nuestro criterio, no es una argumentación válida, al menos en los procesos de pretesteo que son objeto de la presente investigación. Para el caso de la homocedasticidad (Flores & Ocaña) desarrollaron una metodología de una manera más técnica, que permitió encontrar los límites de equivalencia  $\omega_1^2$  y  $\omega_2^2$  a partir de distribuciones normales con media común y varianzas  $\sigma_1^2$  y  $\sigma_2^2$  respectivamente, en donde la TIEP al comparar medias a través de la prueba t de student sobre muestras generadas por estas distribuciones se encuentran en los límites de permisividad  $\alpha \pm \delta$  definidos por el usuario. Dicho límite permite determinar la distancia máxima entre la verdadera TIEP y el valor nominal  $\alpha$ .

La verdadera TIEP es un parámetro desconocido, donde se estimó este valor por medio de un proceso de simulación y los límites de equivalencia se determinaron en función del nivel de significancia  $\alpha$ , el grado de permisividad  $\delta$  y número de observaciones de la muestra definidos por el usuario. Luego al aplicar el pretest de equivalencia con estos límites, se demostró que,

contrario a lo que sucedía cuando se pretesteaba con los test tradicionales, el enfoque de equivalencia usado sí controlaba la probabilidad de error tipo I.

Si bien es cierto, el trabajo realizado por (Flores & Ocaña) es una guía que nos permite sin temor a equivocarnos, pretestear el supuesto de homocedasticidad en una prueba de comparación de medias, nada se ha hecho aún por estudiar una metodología que nos permita comprender la afectación de los pretest de normalidad, ya sean en el enfoque tradicional o de equivalencia. La presente investigación, tiene como objetivo estimar la probabilidad global de cometer un error de tipo I y la potencia de las pruebas de hipótesis para comparar medias, cuando se pretestea el supuesto de normalidad, bajo un enfoque tradicional la prueba de bondad de Ajuste (Chi Cuadrado) y bajo el enfoque a través de una prueba de falta de ajuste. Como ya habíamos mencionado uno de los mayores retos es encontrar los límites de equivalencia válidos para el pretest de normalidad equivalente, los cuales deben hallarse en función del tamaño de la muestra y nivel de significancia.

En la siguiente investigación en el primer capítulo se empieza describiendo los antecedentes, planteamiento del problema, justificación y los objetivos. En el segundo capítulo se analiza los conceptos de las diversas pruebas y metodologías necesarias para desarrollar un nuevo test. El tercer capítulo describe la metodología utilizada para desarrollar el algoritmo donde se define cada una de las funciones que permitieron determinar la TIEP global en cinco procedimientos planteados. En la cuarta sección, se muestra los resultados obtenidos de la simulación y finalmente en la última sección se describen las principales conclusiones y se identifican las áreas para futuras investigaciones.

## **1.2 Formulación del problema**

¿Cómo el proceso de pretestear normalidad en una prueba de comparación de medias afecta la TIEP global?

## **1.3 Preguntas directrices o específicas de la investigación**

- ¿Qué afectación de la probabilidad global de error tipo I en las pruebas de comparación de medias cuando se pretestea el supuesto de normalidad usando la prueba tradicional de bondad de ajuste?

- ¿Cuál es la finalidad de diseñar una función algorítmica de iteración numérica y determinar límites adecuados para las pruebas de hipótesis, para el caso de equivalencia cuando se pretestea el supuesto de normalidad en las pruebas de comparación de medias?
- ¿Qué metodología se puede aplicar para medir la Probabilidad de error tipo I cuando se pretestea el supuesto de normalidad bajo el enfoque tradicional y de equivalencia?
- ¿Qué afectación de la TIEP global se presenta al comparar los diferentes escenarios estudiados aplicados cuando se pretestea el supuesto de normalidad usando pruebas de bondad de ajuste y falta de ajuste?

#### 1.4 Justificación de la Investigación

Las pruebas de comparación de medias son de gran importancia, utilizados en varios ámbitos de la investigación como: en el área de bioestadística (Oca et al., 2004), educación (Dipp et al., 1992), etc. Las pruebas de comparación de medias son usadas para determinar si existen diferencias significativas en la eficacia de dos tratamientos, es decir hacer inferencias acerca de parámetros poblacionales para saber si un tratamiento (metodología o proceso, etc..) es o no mejor que otro. Por ejemplo: en el ámbito de la medicina se puede plantear hipótesis para probar si un nuevo fármaco genérico es más eficaz para combatir con algún padecimiento frente a un medicamento comercial, el investigador debe verificar si se cumplen las hipótesis planteadas en base al promedio que tienen los dos tratamientos, si en promedio el nuevo fármaco genérico tiene un efecto mejor o no en los pacientes, siempre será necesario realizar pruebas de comparación de medias para tomar una decisión acertada.

En la mayoría de investigaciones manifiestan que existe una alteración en la probabilidad global de cometer un error tipo I (TIEP, por sus siglas en inglés) cuando se pretestean los supuestos (normalidad y homocedasticidad) de las pruebas de comparación de medias bajo un enfoque tradicional, es decir que al trabajar con una probabilidad de error tipo I elevada corremos un gran riesgo al tomar una decisión, es decir se comete un error al rechazar (tomar como falsa) la hipótesis nula  $H_0$  que en realidad es verdadera, teniendo en cuenta que en la hipótesis nula se encuentra el cumplimiento del supuesto de normalidad/ homocedasticidad al menos en el enfoque tradicional. El problema radica en la falta de una metodología adecuada para verificar el supuesto de normalidad.

Para el caso de homocedasticidad (Flores et al., 2018) desarrollaron un pretest bajo el enfoque de equivalencia de dispersión para distribuciones Gaussianas, donde la zona de inferencia (límites de equivalencia) no está formalmente definida, por lo cual propusieron un procedimiento iterativo

en base del número de observaciones, el nivel de significancia y grados de permisividad definidos por el investigador, además determinaron la afectación de la probabilidad global de cometer un error tipo I, demostrando que el uso del enfoque de equivalencia para pretestear homocedasticidad hace que las afectaciones que sufre la TIEP desaparezcan, razón por la cual en base a la investigación previa (Flores et al., 2018) pensamos que para el caso de normalidad debemos realizar el mismo trabajo, se pretende generar un aporte metodológico y práctico, construir y probar límites adecuados de irrelevancia que permitan controlar la afectación de la TIEP de las pruebas de comparación de medias cuando se pretestea normalidad.

## **1.5 Objetivos**

### ***1.5.1 Objetivo General***

Estudiar la afectación de la Probabilidad de error tipo I para comparar medias, cuando se pretestea el supuesto de normalidad bajo los enfoques tradicionales y de equivalencia.

### ***1.5.2 Objetivos Específicos***

- Diagnosticar la afectación de la Probabilidad global de error tipo I en las pruebas de comparación de medias cuando se pretestea el supuesto de normalidad usando una prueba tradicional de bondad de ajuste.
- Diseñar una función algorítmica que determine el límite de equivalencia adecuado de una prueba de falta de ajuste para pretestear normalidad en las pruebas de comparación de medias.
- Crear una función para verificar el supuesto de normalidad mediante una prueba de falta de ajuste, usando los límites de equivalencia determinados por la función algorítmica anterior.
- Comparar la efectividad, en términos de la afectación del error tipo I, que tienen las pruebas de comparación de medias cuando se pretestea el supuesto de normalidad usando pruebas de bondad de ajuste y falta de ajuste.

## **1.6 Hipótesis**

El pretest de normalidad bajo el enfoque de equivalencia controla la probabilidad global de cometer un error tipo I, en las pruebas de comparación de medias.

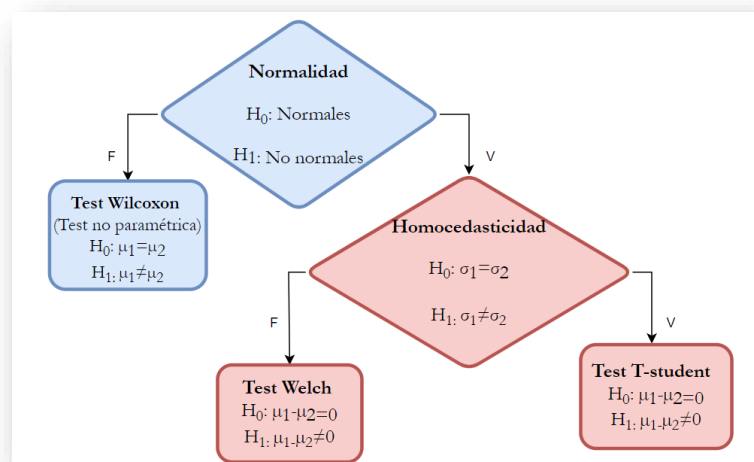
## CAPÍTULO II

### 2 MARCO TEÓRICO

#### 2.1. Antecedentes del problema

Antes de realizar la elección de la prueba de hipótesis apropiada (Wilcoxon, t-Student o Welch) (Wilcoxon, 1945; Student, 1908; Welch, 1947) usada para determinar posibles desigualdades entre 2 medias poblacionales, se debe verificar el cumplimiento de los supuestos de normalidad y homocedasticidad, los mismos que se verifican realizando una prueba previa como por ejemplo Shapiro-Wilk o Kolmogoroff-Smirnov (Shapiro & Wilk, 1965; Kolmogorov, 1933) en el caso de normalidad.

Cuando la hipótesis nula de estas pruebas se rechaza se asume que los datos no provienen de una distribución normal y se adopta el método no paramétrico de Wilcoxon, mientras la hipótesis nula de normalidad no se rechaza, entonces se asume el cumplimiento de este supuesto y el siguiente paso es probar el cumplimiento o no de homocedasticidad. Pruebas como Levene, Bartlett, etc, tradicionalmente nos han ayudado a pretestear este supuesto, en las cuales, al rechazar la hipótesis nula de igualdad de varianzas, se asume heterocedasticidad y el método más adecuado para comparación de medias es el test de Welch, mientras que, si esta hipótesis no es rechazada, se asume homocedasticidad y la prueba t-Student se asume como la adecuada. En la Figura 1, se resume esta dinámica adoptada tradicionalmente en los procesos de inferencia estadística.



**Figura 1-2.** Pretest Tradicional

Realizado por: María Palacios, 2022

Usualmente para verificar los supuestos se usan procedimientos previos a los que se conocen como pretest, aunque usarlos no es lo más adecuado. Varios estudios (Hsu (1938), Overall, J. E., Atlas, R. S., & Gibson, J. M. (1995)) muestran que aplicar estos procesos alteran la Probabilidad global de cometer un error tipo I, especialmente con tamaños de muestras pequeñas y desbalanceadas.

Rasch, Kubinger y Moder (Kubinger et al., 2011) afirman que realizar pruebas previas para validar los supuestos de normalidad (Shapiro Wilks, Kolmogorov Smirnov) y homocedasticidad (Levene, Bartlett, Cochran) en la pruebas de comparación de medias, conduce a riesgos desconocidos provocando una alteración de la probabilidad global de cometer un error de tipo I y tipo II. Por lo cual estos autores recomiendan nunca utilizar este procedimiento de pretesteo.

Acerca de estas alteraciones de la TIEP debido al proceso de pretesteo, Wellek en su libro “Testing statistical hypotheses of equivalence and noninferiority” (Wellek, 2010) menciona que el problema no radica en pretestear los supuestos sino más bien en la forma como estos tradicionalmente han sido planteados, razón por la cual propone un nuevo enfoque que se basa en el concepto de equivalencia. En este tipo de test la hipótesis alternativa establece equivalencia es decir igualdad (de parámetros como por ejemplo varianza) excepto para desviaciones irrelevantes, mientras que la hipótesis nula establece la no equivalencia. Uno de los aspectos más importante es determinar el  $\epsilon$  óptimo que permita establecer un criterio comparativo y fiable el cual lleve a una decisión acertada de la normalidad o no normalidad de los datos.

En el caso del supuesto de homocedasticidad, (Flores & Ocaña, 2018) propusieron un test para probar equivalencia entre varianzas, el cual contrario a todos los procedimientos tradicionales, demostró que no existe alteración de la probabilidad de error tipo I al simular de las pruebas de comparación de medias. Es por ello, que la presente investigación busca construir un pretest basado en ese mismo enfoque, pero ahora para probar normalidad, lo cual creemos, también controlará la TIEP y permitirá procedimientos de pruebas y análisis de datos confiables.

## **2.2. Bases Teóricas**

### ***2.2.1. Pruebas de Hipótesis para comparación de medias***

#### ***2.1.1.1 Pruebas de hipótesis***

Cuando se realiza una investigación estadística con la finalidad de tomar una decisión sobre la población, se determina una afirmación o suposición acerca de un parámetro de una o más

poblaciones cuya veracidad se prueba, a esta afirmación se le conoce como hipótesis y el procedimiento basado en la evidencia de la muestra para determinar si esa hipótesis es verdadera o falsa se llama prueba de hipótesis, hay pruebas unilaterales y bilaterales. Una prueba estadística de acuerdo a Wackerly et al. (2010, p.37) tiene cuatro componentes esenciales se deben plantear dos hipótesis: la hipótesis nula es cualquier hipótesis que desea probar se denota  $H_0$ , la hipótesis alternativa es la hipótesis que se acepta cuando la hipótesis nula es rechazada se denota  $H_1$ , estadística de la prueba y región de rechazo, las hipótesis se comparan seleccionando el estadístico adecuado para cada situación y de este modo se decide en favor de la hipótesis nula o de la alternativa. El estadístico de prueba es una función de las mediciones muestrales en las que la decisión estadística estará basada. La región de rechazo define los valores del estadístico de prueba para el cual la  $H_0$  ha de ser rechazada a favor de la  $H_1$ .

### 2.1.2 *Distribución muestral*

Por lo general necesitamos estudiar las propiedades de una determinada población, para lo cual podemos calcular algunos parámetros. Un parámetro es una caracterización numérica de la distribución de la población de manera que describe, parcial o completamente, la función de densidad de probabilidad de la característica de interés, es decir establece una característica real de dicha población por ejemplo la media, desviación estándar, etc...Por ejemplo cuando se especifica el valor del parámetro de escala exponencial  $\theta$  se describe de manera completa la función de densidad de probabilidad:

$$f(X; \theta) = \frac{1}{\theta} \exp(-X/\theta) \quad (1)$$

Una vez que se conoce el valor de  $\theta$  entonces puede formularse cualquier hipótesis estadística de interés, estos parámetros a veces no son fáciles de calcular ya que la población es demasiado numerosa como para analizar a todos los individuos que la componen. Por tal motivo, para estimar los parámetros de determinada población recurrimos a extraer una muestra de la misma, se puede calcular un estadístico, así obtener un estimado de dicho parámetro. y de este modo inferir su comportamiento (Otzen & Manterola, 2017).

#### 2.1.2.1 *Estadístico de Prueba*

Un estadístico de prueba que es una expresión matemática sigue una función de probabilidad distribución muestral, es una función que sirve para calcular un parámetro de la población, a partir los datos de la muestra que se tome. El estadístico de prueba se utiliza en las pruebas de hipótesis para comparar los datos con lo que se espera bajo la hipótesis nula y así determinar si se rechaza o no la hipótesis nula. Las pruebas de hipótesis son diferentes y utilizan diferentes estadísticos de

prueba según el modelo de probabilidad asumido en la hipótesis nula. El estadístico de prueba calcula el valor p. (Clark, 2004).

### 2.1.2.2 P valor

El p valor es el área acumulada bajo la curva como se muestra en la figura 2, y se obtiene a partir de la diferencia entre la unidad y la función de distribución acumulada que especifica la probabilidad de que una variable aleatoria sea menor o igual que un valor dado, como el valor obtenido del estadístico de prueba. La fórmula de la función de distribución es la siguiente:

$$F(x) = p(X \leq x) = \int_{-\infty}^x f(x)dx \quad (2)$$

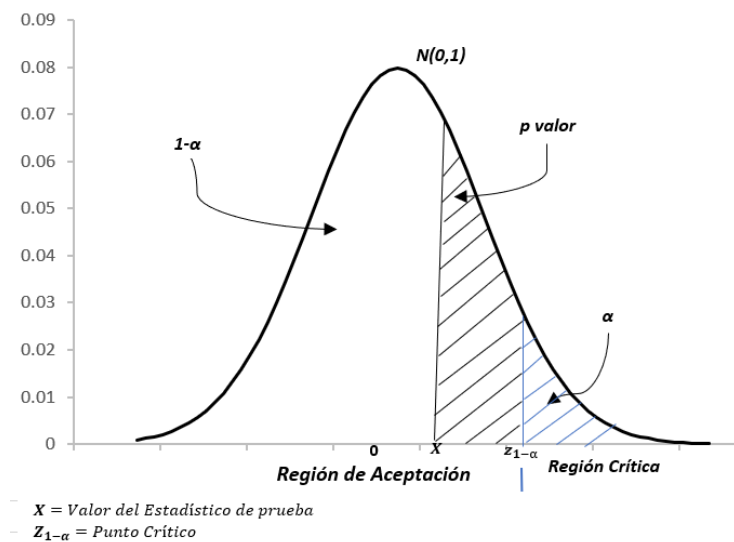
Dónde:

X: Variable aleatoria continua

f(x): función de densidad de la probabilidad

F(x): función de distribución acumulada

Entonces el  $p - valor = 1 - F(x)$



**Figura 2-2.** p valor área bajo la curva  
**Realizado por:** María Palacios, 2022

El p valor viene dado por el estadístico de prueba calculado a partir de los datos de la investigación, es la probabilidad o área que deja la cola del valor crítico, el valor obtenido cambia aleatoriamente de una muestra a otra diferente, es la probabilidad que permite determinar el grado de credibilidad de la hipótesis nula. El p valor es la probabilidad calculada suponiendo que H0 es



verdadera, el  $p$  valor es el área capturada por el valor calculado en el estadístico de prueba.(Devore et al., 2021).

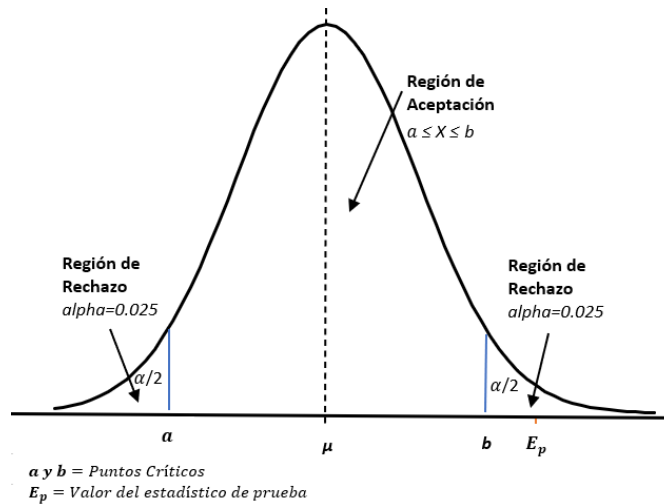
Para lo cual se utiliza el nivel de significancia que es un valor establecido por el investigador no existe un valor estándar o universal para probar la hipótesis, en algunos casos utilizan el nivel de significancia del 5%. El nivel de significancia se denota  $\alpha$  es la probabilidad de rechazar la hipótesis nula cuando es verdadera, el mismo que permite establecer con base en probabilidades un criterio para determinar si existe la evidencia suficiente para rechazar la hipótesis nula. (Clark, 2004).

Si el  $p$  valor  $< \alpha$  no se rechaza  $H_0$  y se concluye que no existe evidencia para decir que  $H_1$  es cierta, Si el  $p$  valor  $> \alpha$  se rechaza  $H_0$  la hipótesis y se dice que existe evidencia para decir  $H_1$  es cierta.

### **2.1.2.3 Valores Críticos**

Los valores críticos o también conocidos como región crítica o de rechazo de una prueba, es el conjunto de todos los valores del estadístico de prueba obtenidos para los cuales la hipótesis nula será rechazada, la parte de la distribución del estadístico cuya área se denota  $\alpha$  se llama región de rechazo y el complemento de esta región se llama región de aceptación y su área es  $(1 - \alpha)$ . Un valor crítico es un punto que divide la región de aceptación y rechazo de la hipótesis nula, a través de los valores críticos se comprueba matemáticamente si la hipótesis nula debe ser rechazada o no.

El punto crítico en el dominio de la distribución que deja a la derecha un área  $\alpha$ , cuando el estadístico de prueba cae a la derecha o dentro de la zona de rechazo que se observa en la figura # se rechaza  $H_0$  a favor de la hipótesis  $H_1$ , si el estadístico de prueba cae en la región de aceptación no rechazo la hipótesis  $H_0$  (Gmurman V E - Teoria De Las Probabilidades Y Estadistica Matematica.PDF, n.d.).



**Figura 3-2.** Región de rechazo y aceptación de  $H_0$  en la distribución normal para dos colas  
**Realizado por:** María Palacios, 2022

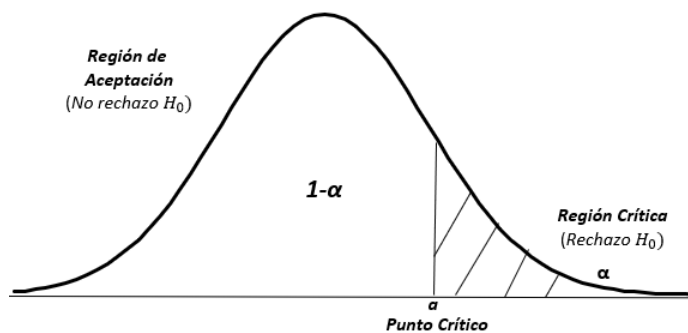
**Prueba Unilateral (una cola)**

Esta prueba tiene un valor crítico y la región de rechazo es única a un lado de la distribución. Dónde sí la media de la muestra es mayor que la media hipotética de la población:

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0$$

La región crítica para la hipótesis alternativa estará en la cola derecha de la distribución del estadístico de prueba como se muestra en la figura 4.



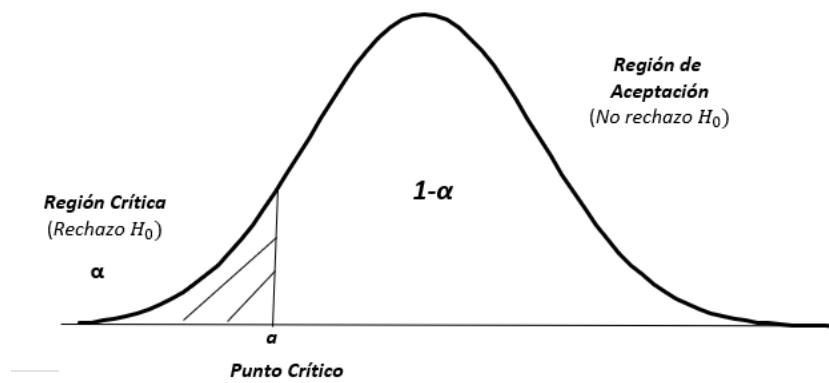
**Figura 4-2.** Prueba de uno sola cola a la derecha  
**Realizado por:** María Palacios, 2022

Ahora si la media de la muestra es menor que la media hipotética de la población:

$$H_0: \mu = \mu_0$$

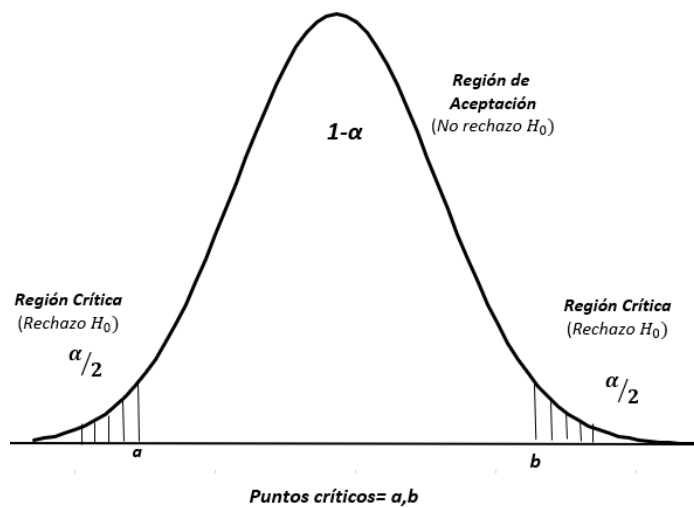
$$H_1: \mu < \mu_0$$

La región crítica para la hipótesis alternativa estará en la cola izquierda de la distribución del estadístico de prueba, como se muestra en la figura 5.



**Figura 5-2.** Prueba de uno sola cola a la izquierda  
**Realizado por:** María Palacios, 2022

*Prueba Bilateral (dos colas)*



**Figura 6-2.** Prueba de dos colas  
**Realizado por:** María Palacios, 2022

Si la media de la muestra es significativamente mayor o menor que la media hipotética de la población, Tiene dos valores críticos la región de rechazo se determina dividiendo el nivel de significancia en dos partes iguales, es muy común que tengan probabilidades iguales en cada cola de la distribución del estadístico de prueba, como se observa en la figura 4.

Dónde:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ H_1: \mu &\neq \mu_0 \end{aligned}$$

#### 2.1.2.4 Tipos de Errores

No obstante rechazar o no la hipótesis nula no significa que esta sea falsa, sino que no existe evidencia suficiente para aceptarla, sin embargo, al tomar una decisión se puede cometer errores por lo cual existen dos tipos de error:

*Probabilidad de error tipo I:* Significa rechazar la  $H_0$  cuando esta es verdadera, este valor coincide con el nivel de significancia  $\alpha$  que es un valor que se suele proporcionar por lo general de 0,05 , es la probabilidad que el punto muestral (estadístico) caiga en región de rechazo R bajo  $H_0$ , esto es,

$$\begin{aligned} \alpha &= P[\text{rechazar } H_0 / H_0 \text{ es verdadera}] \\ &= \int_R f(\hat{\theta} / H_0) d\hat{\theta} \\ &= \int_c^\infty f(\hat{\theta} / H_0) d\hat{\theta} \end{aligned} \quad (3)$$

Dónde  $f(\hat{\theta} / H_0)$  es la distribución muestral de  $\hat{\theta}$  bajo la hipótesis  $H_0$ .

*Probabilidad de error tipo II:* Significa no rechazar  $H_0$  cuando esta es falsa se denotado por  $\beta$  es la probabilidad que el punto muestral caiga en la región de aceptación A bajo  $H_1$ , esto es;

$$\begin{aligned} \beta &= P[\text{aceptar } H_0 / H_0 \text{ es falsa}] \\ &= P[\text{rechazar } H_1 / H_1 \text{ es verdadera}] \\ &= \int_A f(\hat{\theta} / H_1) d\hat{\theta} \\ &= \int_{-\infty}^c f(\hat{\theta} / H_1) d\hat{\theta} \end{aligned} \quad (4)$$

Dónde  $f(\hat{\theta} / H_1)$  es la distribución muestral de  $\hat{\theta}$  bajo la hipótesis  $H_1$ . (Quispe Caray, 2018)

Mientras que  $1 - \beta$  es la probabilidad de rechazar una hipótesis que es falsa es conocida como “potencia de la prueba”, la potencia de una prueba estadística permite determinar la fiabilidad de dichas pruebas, como se indica en la Tabla 1-2

**Tabla 1-2.** Toma de decisiones Tipo de Error

<b>Decisión</b>	<b><math>H_0</math> Verdadera</b>	<b><math>H_0</math> Falsa</b>
Rechaza $H_0$	Acierta $1 - \alpha$ Nivel de confianza	Error Tipo II $\beta$
Acepta $H_0$	Error Tipo I	Acierto $1 - \beta$ Potencia de Prueba

**Realizado por:** María Palacios, 2022

### 2.1.2.5 *Lema de Neyman y Pearson*

Las Pruebas de Significación fueron desarrolladas por Fisher para hacer inferencias a partir de un proceso inductivo que conduciría al conocimiento, analizado como teoría matemática, su procedimiento tenía algunas fallas o limitaciones al ser construido sobre nociones inespecíficas. Estas falencias condujeron a Jerzy Neyman y Egon Pearson a desarrollar otro criterio para abordar el problema de la validez de las Hipótesis Estadísticas de este modo se dio el origen de la Estadística Matemática actual entre 1928 y 1933 donde redactaron una serie de documentos como “El problema de la forma más eficiente de las pruebas de hipótesis estadística”, su desarrollo teórico fue denominado por ellos Pruebas de Hipótesis, partía de una sólida aplicación de las Matemáticas y fueron desarrolladas como una teoría matemática derivada directamente de la Teoría de Probabilidades. esta solución al problema se identifica como un mecanismo de comportamiento inductivo, opuesto en su esencia a la inferencia inductiva de las Pruebas de Significación de Fisher. Las Pruebas de Hipótesis de Neyman y Pearson se construyen considerando junto a la hipótesis  $H_0$  de las Pruebas de Significación, otra hipótesis,  $H_1$  o alternativa, que incorpora el proceso a seguir en el caso en que  $H_0$  sea rechazada.

cuando estos supuestos son pre-testeados mediante las pruebas tradicionales, existe un alto riesgo de cometer un error Tipo I

Con esta tendencia se transformó en un problema de decisión, con dos posibles alternativas con los datos de los que dispone el investigador solo se puede rechazar una hipótesis y por lo tanto aceptar la otra, se establecen dos tipos de error: Error tipo 1, que se comete cuando se acepta cuando se rechaza la hipótesis nula que es cierta y Error tipo II aceptar una hipótesis nula que es falsa.

En la teoría de *Neyman-Pearson* lo que se pretende es acotar la máxima probabilidad ( $\max_{\theta \in \theta_0} P_\theta(R)$ ) de error tipo I por  $\alpha \in ]0,1[$ , o sea, se busca una región de rechazo R tal que:

$$P_0(R) < \alpha \text{ entonces } \max_{\theta \in \theta_0} P_\theta(R) \leq \beta$$

Y, por otro lado, minimizar la probabilidad de error tipo II, para lo cual, se busca una región de rechazo R tal que:

$$P_0(R) > \beta \text{ entonces } \min_{\theta \in \theta_1} P_\theta(R) \geq \beta$$

Como se observa los test de hipótesis están diseñados para controlar la probabilidad máxima de rechazar  $H_0$  cuando es cierta. En algunas ocasiones se observa que es posible rechazar  $H_0$  con los mismos datos para un nivel de significancia  $\alpha = 0.05$  que aceptar con  $\alpha = 0.01$ .

Sea  $\alpha \in (0,1)$  y sean  $\theta_0$  y  $\theta_1$  dos valores distintos del parámetro desconocido  $\theta$ . La región de rechazo más potente de tamaño  $\alpha$  para el contraste de hipótesis.

$$H_0: \theta = \theta_0 \quad \text{Vs} \quad H_1: \theta = \theta_1$$

esta dada por  $C = \{(X_1, \dots, X_n) = \frac{L(X_1, \dots, X_n; \theta_1)}{L(X_1, \dots, X_n; \theta_0)} > c\}$  (5)

En dónde C: Conjunto de valores de la muestra aleatoria

c es una constante que hace que esta región de rechazo sea de tamaño  $\alpha$

### 2.1.3 Test de hipótesis para comparar medias

Para comparar medias bajo  $H_0: \mu_1 - \mu_2 = 0$  y  $H_1: \mu_1 - \mu_2 \neq 0$ , donde  $\mu_1$  es la media poblacional del tratamiento 1 y  $\mu_2$  es la media poblacional del tratamiento 2, para probar estas hipótesis nosotros tenemos bajo el enfoque paramétrico T de Student donde veo la variable implicada, el estadístico de prueba, la distribución de probabilidades del estadístico de prueba, los valores críticos y bajo el enfoque no paramétrico como es el test de Wilcoxon donde tendremos una distribución estadística libre, la variable implicada y el estadístico de prueba.

#### 2.1.3.1 Test de Wilcoxon

Al verificar el supuesto de Normalidad y determinar que los datos no siguen una distribución normal sino una distribución libre, se procede a utilizar el test Wilcoxon es una prueba no paramétrica es la versión alternativa de la prueba t de Student, aplicada para comparar dos muestras independientes, también conocida como test de Wilcoxon, se basa en  $x_1, \dots, x_{n_1}$  y  $y_1, \dots, y_{n_2}$  que son dos muestras aleatorias independientes con distribuciones libres, compara cada observación de  $x_i$  con cada observación en la segunda muestra de  $y_i$ . (WILCOXON, 1946). Las hipótesis para contrastar se utilizan reglas de decisión, como se indica en la Tabla 2-2.

**Tabla 2-2.** Regla de decisión

<i>Hipótesis Nula:</i>	<b><math>H_0: \mu_1 = \mu_2</math></b>
<i>Hipótesis Alternativa:</i>	$H_0: \mu_1 \neq \mu_2$ Contraste bilateral
	$H_0: \mu_1 > \mu_2$ Contraste unilateral La distribución de la población x se desplaza a la derecha de la población y.
	$H_0: \mu_1 < \mu_2$ Contraste unilateral La distribución de la población x se desplaza a la izquierda de la población y.

**Fuente:** Wilcoxon,1945

**Realizado por:** María Palacios, 2022

Mientras que el estadístico de prueba dado por:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_2 \quad (6)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \quad (7)$$

$$U = \min(U_1, U_2) \quad (8)$$

Para lo cual, existe la regla de decisión que se indica en la Tabla 3-2.

**Tabla 3-2.** Región de Rechazo de la prueba U de Mann Withney

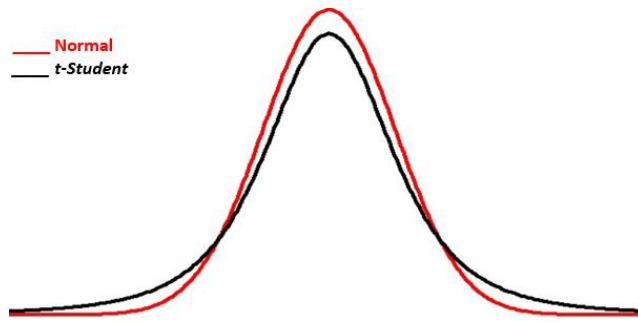
Unilateral izquierda	<b><math>U \geq n_1 n_2 - U_0</math></b>
Unilateral derecha	<b><math>U \leq U_0</math></b>
Bilateral	<b><math>U \leq U_0</math> o <math>U \geq n_1 n_2 - U_0</math></b>

**Fuente:** Wilcoxon,1945

**Realizado por:** María Palacios, 2022

### 2.1.3.2 Test t- Student

La distribución de probabilidad de la t- Student sigue una distribución normal, aunque t- Studente presenta unas colas más pesadas que la curva de Gauss, como se muestra en la figura 7.



**Figura 7-2.** Distribución t-Student vs distribución normal  
**Realizado por:** María Palacios, 2022

Suponga que se seleccionaron muestras aleatorias independientes de distribuciones normales  $X_1, X_2, \dots, X_{n1} \sim N(\mu_1, \sigma_1^2)$  y  $Y_1, Y_2, \dots, Y_{n2} \sim N(\mu_2, \sigma_2^2)$ , con medias desconocidas y que poseen varianzas iguales ( $\sigma_1^2 = \sigma_2^2$ ). Para la construcción de la prueba t-Student que compara las medias  $\mu_1 = \mu_2$ , las hipótesis se plantean a continuación (CEPAL, 2018).

La hipótesis nula es:

$$H_0: \mu_1 - \mu_2 = 0$$

Mientras que la hipótesis alternativa se define en la Tabla 4-2.

**Tabla 4-2.** Hipótesis alternativa de la prueba t -Student

<b>Unilateral a la izquierda</b>	<b>Unilateral derecha</b>	<b>Bilateral (2 colas)</b>
$H_1: \mu_1 - \mu_2 > 0$	$H_1: \mu_1 - \mu_2 < 0$	$H_1: \mu_1 - \mu_2 \neq 0$

**Fuente:** CEPAL (2018)

**Realizado por:** María Palacios, 2022

El estadístico de prueba es:

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (9)$$

Dónde:

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1)}{n_1 + n_2 - 2} \quad (10)$$

t sigue una distribución t de Student con  $n_1 + n_2 - 2$  grados de libertad, al crecer el número de grados de libertad la distribución t de Student tiende hacia la normal estándar (Bowen et al., 1986).

La región de rechazo viene dada por los siguientes parámetros que muestran en la Tabla 5-2.



**Tabla 5-2.** Región de rechazo de la prueba t-Student a un nivel de significancia  $\alpha$ .

RR Izquierda	RR Derecha	RR Dos colas
$t > t_{\alpha, n_1 + n_2 - 2}$	$t < -t_{\alpha, n_1 + n_2 - 2}$	$ t  > t_{\frac{\alpha}{2}, n_1 + n_2 - 2}$

**Fuente:** Bartoszynski & Niewiadomska-Bugaj, 19

**Realizado por:** María Palacios, 2022

## 2.1.4 Test para verificar el supuesto de normalidad

### 2.1.4.1 Prueba Chi cuadrado de bondad de ajuste

Una prueba de bondad de ajuste Chi Cuadrado o también conocida como distancia Chi Cuadrado, es calculada a partir de la tabla de contingencia es una prueba de libre distribución (no paramétrica) se emplea para decidir cuándo un conjunto de datos se apega a una distribución de probabilidad dada. Considérese una muestra aleatoria de tamaño  $n$  de la distribución de una variable aleatoria  $X$  dividida en  $K$  clases exclusivas y mutuamente excluyentes, y sea  $N_i, i = 1, 2, \dots, k$ , el número de observaciones en la  $i$ -ésima clase (Bowen et al., 1986).

Considérese la verificación de la hipótesis nula

$$H_0: \pi(x) = \pi_0(x)$$

En donde el modelo de probabilidad propuesto  $\pi_0(x)$  se encuentra especificado, de manera completa, con respecto a todos los parámetros. De esta forma la hipótesis nula es sencilla. Dado que se especifica  $\pi_0(x)$  de manera completa, se puede obtener la probabilidad  $p_i$  de obtener una observación en la  $i$ -ésima clase bajo  $H_0$ , en donde necesariamente  $\sum_{i=1}^k p_i = 1$ .

Sea  $n_i$  la realización de  $N_i$  para  $i = 1, 2, \dots, k$  de manera tal que  $\sum_{i=1}^k n_i = n$ . La probabilidad de tener, de manera exacta,  $n_i$  observaciones en la  $i$ -ésima clase es  $p_i^{n_i}$  para  $i = 1, 2, \dots, k$ . Dado que existen  $k$  categorías mutuamente excluyentes con probabilidades  $p_1, p_2, \dots, p_k$ , entonces bajo la hipótesis nula la probabilidad de una distribución multinomial determinada.

Para deducir una prueba estadística adecuada para  $H_0$ , considérese el caso en el que  $K=2$ . Esta es la distribución binomial con una función de probabilidad dada por:

*Definición 4.1.-* Sea  $X$  una variable aleatoria que representa el número de éxitos en  $n$  ensayos y  $p$  la probabilidad de éxitos con cualquiera de éstos. Se dice entonces que  $X$  tiene una distribución binomial con función de probabilidad.

$$p(x; n, p) = \left\{ \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n. \right. \quad (11)$$

O para cualquier otro valor  $0 \leq p \leq 1$ , para  $n$  entero

y en la que  $x = n_1, p = p_1, n - x = n_2$  y  $1 - p = p_2$ .

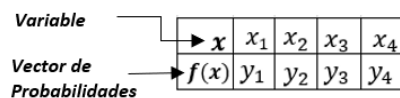
Considérese la variable aleatoria estandarizada  $y = \frac{N_1 - np_1}{\sqrt{np_1(1-p_1)}}$  (12)

Si el valor de  $n$  es suficientemente grande, la distribución  $Y$  es aproximadamente igual a la normal estándar, se sabe que el cuadrado de una variable aleatoria normal estándar tiene una distribución chi-cuadrado con un grado de libertad. Entonces, el estadístico de prueba:

$$\begin{aligned} \frac{(N_1 - np_1)^2}{np_1(1-p_1)} &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_1 - np_1)^2}{np_2} \\ &= \frac{(N_1 - np_1)^2}{np_1} + \frac{[n - N_2 - n(1-p_2)]^2}{np_2} \\ &= \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2} \\ &= \sum_{i=1}^2 \frac{(N_i - np_i)^2}{np_i} \end{aligned} \quad (13)$$

Tiene aproximadamente una distribución chi-cuadrado con un grado de libertad conforme  $n$  va tomando valores cada vez más grandes. La principal característica es que la prueba Chi Cuadrado  $X^2$  toma valores entre cero e infinito y no tiene valores negativos porque es la suma de valores elevados al cuadrado. (Cobo et al., 2014).

Existen dos tipos de variables discretas o continuas, digamos la variable  $x$  y la función de probabilidad sería  $f(x)$ , para definir que función de probabilidad puede seguir  $x$  (Binomial, Poisson, Normal, Chi cuadrado, etc...) depende de las probabilidades que toma:



**Figura 8-2.** Función de probabilidad

**Realizado por:** María Palacios, 2022

Este vector de probabilidades son los valores asociados a los diferentes valores que puede tomar la variable  $x$ . De acuerdo a como sea este vector de probabilidades sabemos qué tipo de función

de probabilidad sigue mi variable aleatoria, la función de probabilidad es el comportamiento de las probabilidades.

**Distribución Objetivo (normal).** - Denotado por  $\pi^0$ , son las probabilidades asociadas a las distribuciones que nosotros queremos obtener, que en nuestro caso específico sería a la distribución normal, entonces  $\pi^0$  sería un conjunto hecho vector de probabilidades donde cada uno de los elementos deberían estar asociados a las áreas bajo la curva de la normal en cada intervalo.

Esta distribución se caracteriza porque los valores se distribuyen formando una campana de Gauss, en torno a un valor central que coincide con el valor medio de la distribución. Sea X una variable aleatoria, se dice que se distribuye como una normal:

$$X \rightarrow N(\mu, \sigma); \quad \mu \in \mathcal{R} \quad \sigma > 0$$

Dónde se verifica que  $-\infty < x < +\infty$ ,  $\mu$ , es el valor medio de la distribución y es precisamente donde se sitúa el centro de la curva (de la campana de Gauss), y  $\sigma$  es cualquier valor entre  $-\infty$  y  $+\infty$ , si su función de densidad viene dada por:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (14)$$

**Distribución observada de los datos.** - Denotado por  $\hat{\pi}$ , lo que vamos a tener son datos donde yo puedo sacar las frecuencias relativas en los mismos intervalos, se quiere ver si las frecuencias en los datos  $\hat{\pi}$  se aproximan o no a las distribuciones reales  $\pi$  que hay en la distribución teórica.

**Distribución verdadera.** - Denotada por  $\pi$  es la distribución teórica que se supone que siguen los datos, la cual nunca se va a conocer, queremos ver si se ajusta a  $\pi^0$  que es la distribución objetivo la normal,  $\hat{\pi}$  es la que se aproxima a  $\pi$ .

Para verificar si los datos vienen o no de una distribución normal en la prueba de Bondad de ajuste de acuerdo a nuestra investigación se han identificado las siguientes hipótesis a contrastar:

$$\begin{aligned} H_0 : d_c(\pi, \pi^0) &= \mathbf{0} && \text{Los datos son normales} \\ H_1 : d_c(\pi, \pi^0) &\neq \mathbf{0} && \text{Los datos no son normales} \end{aligned}$$

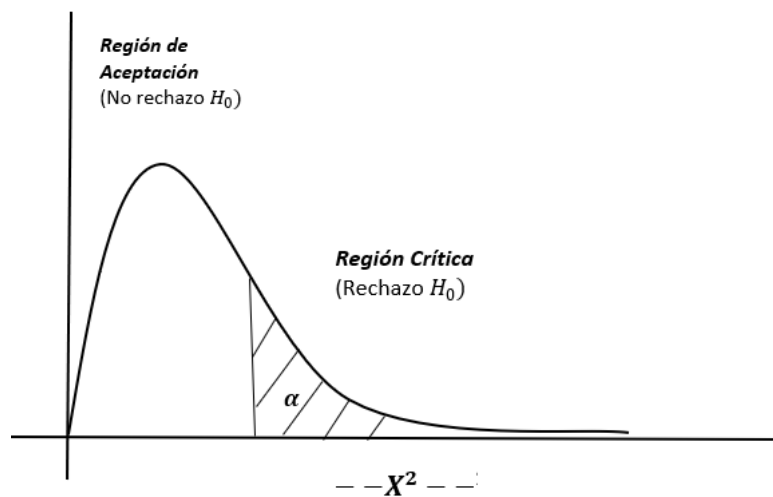
Dónde  $d_c$  es la distancia Chi cuadrado entre la distribución teórica  $\pi$  y la distribución observada de los datos  $\pi^0$ .

El estadístico de prueba es:

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño mayor que 5, el estadístico de prueba es:

$$X^2 = \sum_{i=1}^k \frac{(\hat{\pi}_i - \pi_i^0)^2}{\pi_i^0} \quad (15)$$

Dónde  $k$  es el número de clases existentes  $\hat{\pi}$  es la frecuencia de distribuciones observadas en cada clase y  $\pi^0$  es la frecuencia distribuciones objetivo. La hipótesis se rechaza cuando  $X^2 > X_\alpha^2$  o cuando el p-valor es menor que el nivel de significancia  $\alpha$ , donde  $X_\alpha^2$  es el valor crítico de una distribución chi cuadrado con  $k-1$  grados de libertad que deja a la derecha un área de  $\alpha$ . Cuyo gráfico dependerá de los grados de libertad.



**Figura 9-2.** Chi cuadrado es una prueba Unilateral  
**Realizado por:** María Palacios, 2022

Como se observa en la figura 9-2 la prueba es lateral derecha y se utilizan los valores de la cola superior de  $X^2$  para ubicar la zona de rechazo, debido a que desviaciones grandes de los valores observados con los valores esperados, tienden a contradecir la hipótesis nula respecto a las probabilidades asociadas a las categorías (valores o clases). Por lo tanto, la hipótesis nula se rechazará cuando el estadístico de la prueba  $X^2$  (valor observado bajo el supuesto de hipótesis nula cierta) asuma un valor grande.

#### 2.1.4.2 Prueba de Equivalencia para falta de Ajuste

En su forma más simple los problemas sobre probar bondad de ajuste en lugar de falta de ajuste implican una distribución multinomial completamente especificada junto con las de la misma clase.

El conjunto de datos primarios consiste de  $n$  vectores mutuamente independientes:

**Tabla 6-2.** Vectores mutuamente independientes

$\mathbf{v}_1$	$\mathbf{v}_2$	...	$\mathbf{v}_n$
$y_{11}$	$y_{21}$		$y_{n1}$
$y_{12}$	$y_{22}$		$y_{n2}$
.	.		.
.	.		.
.	.		.
$y_{1k}$	$y_{2k}$		$y_{nk}$

con  $k \geq 2$

**Realizado por:** María Palacios, 2022

Cada vector  $\mathbf{v}_i = (y_{i1}, y_{i2}, \dots, y_{ik})$  consiste de  $k - 1$  ceros y un uno y la probabilidad de que un elemento no nulo aparezca en la  $j$ -ésima posición, se asume que el mismo número  $\pi_j \in [0,1]$  para cada elemento de la muestra. Como es usual en el análisis de datos categóricos estos vectores se agregan a las correspondientes celdas  $(x_1, x_2, \dots, x_k)$  que se forman definiendo  $x_j = \#\{i/i \in \{1, \dots, n\}, y_{ij} = 1\}$  para  $j = 1, \dots, k$ .

*Ejemplo:*

Suponga un conjunto de datos {a,b,c} con frecuencias {2,3,4} entonces  $k = 3, n = 2 + 3 + 4 = 9$

**Tabla 7-2.** Datos, vectores y frecuencias

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$	$\mathbf{v}_5$	$\mathbf{v}_6$	$\mathbf{v}_7$	$\mathbf{v}_8$	$\mathbf{v}_9$
<b>a</b>	1	1	0	0	0	0	0	0	0
<b>b</b>	0	0	1	1	1	0	0	0	0
<b>c</b>	0	0	0	0	0	1	1	1	1

$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

**Realizado por:** María Palacios, 2022

Tabla de vectores

$$\pi_j = \frac{1}{9}; \quad j = 1, \dots, 9$$

Entonces dado  $(x_1, x_2, \dots, x_k)$ , se deduce que  $\sum_{j=0}^k x_i = n$  y  $x_1, x_2, \dots, x_k \sim Mult(n, \pi)$  con  $\pi = (\pi_1, \pi_2, \dots, \pi_k)$ . Entonces la distribución de probabilidad es:

$$P(x_1 = x_1, \dots, x_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} (\pi_1^{x_1} \pi_2^{x_2} \dots \pi_k^{x_k}) \quad (16)$$

La notación que usaremos es  $Mult(n; \pi_1, \pi_2, \dots, \pi_k)$ , son las probabilidades teóricas de la multinomial.

La distribución de referencia al cual se quiere probar que los datos  $(x_1, x_2, \dots, x_k)$  si se ajustan suficientemente bien, se establece por el simple hecho de escoger el parámetro adecuado, digamos  $\pi^0 = (\pi_1^0, \pi_2^0, \dots, \pi_k^0)$  que son las probabilidades teóricas de la distribución objetivo (por ejemplo, la normal) que queremos ajustar.

Con el fin de cuantificar el grado de disimilitud entre la verdadera distribución (implica  $\pi$ ) y la distribución objetivo (implica  $\pi^0$ ) de  $x_1, x_2, \dots, x_k$ , parece razonable usar la ordinaria distancia euclídea  $d_e$  entre los parámetros asociados (vectores  $\pi$  y  $\pi^0$ ), lo cual deja la siguiente formulación del problema de probar el ajuste de  $Mult(n; \pi_1, \dots, \pi_k)$  a  $Mult(n; \pi_1^0, \dots, \pi_k^0)$

$$H_0: d_e^2(\pi, \pi^0) \geq \varepsilon^2 \quad \text{Falta de Ajuste (No normalidad)}$$

$$H_1: d_e^2(\pi, \pi^0) < \varepsilon^2 \quad \text{Ajuste Normalidad}$$

Con  $d_e^2(\pi, \pi^0) = \sum_{j=1}^k (\pi_j - \pi_j^0)^2$  que es la distancia euclídea entre los vectores de la distribución verdadera y objetiva.

A partir de la teoría asintótica propuesta por Bishop (1975), Teorema 14.3 se sabe que el vector normalizado

$$\sqrt{n}(\hat{\pi}_1 - \pi_1, \dots, \hat{\pi}_k - \pi_k) = \sqrt{n}\left(\frac{x_1}{n} - \pi_1, \dots, \frac{x_k}{n} - \pi_k\right) \quad (17)$$

$\hat{\pi}_j = \frac{x_j}{n}$  de frecuencias relativas que converge en ley (conforme  $n \rightarrow \infty$ ) a una variable aleatoria normal multivariante ( $k$ ) con media cero y matriz de covarianzas dado por:

$$\Sigma = D_\pi - \pi' \pi = \begin{matrix} k \times k \\ \begin{pmatrix} \pi_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \pi_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \pi_k \end{pmatrix} \end{matrix} - \begin{matrix} k \times k \\ \begin{pmatrix} \pi_1^2 & \pi_1 \pi_2 & \pi_1 \pi_k \\ \pi_2 \pi_1 & \pi_2^2 & \pi_2 \pi_k \\ \pi_k \pi_1 & \pi_k \pi_2 & \pi_k^2 \end{pmatrix} \end{matrix} \quad (18)$$

Aún más ya que  $\pi \rightarrow d_e^2(\pi, \pi^0)$  es diferenciable sobre  $\pi$  con vector gradiente  $\nabla d_e^2(\pi, \pi^0)$ . Todo esto implica que podemos usar el llamado *método  $\delta$*  (que es el polinomio de grado 1 de Taylor) para encontrar la distribución asintótica de  $d_e^2(\hat{\pi}, \pi^0)$ .

“Distancia estimada entre el vector verdadero y el vector objetivo”.

Nos queda, entonces:

$$\sigma^2[\sqrt{n} d_e^2(\hat{\pi}, \pi^0)] = 4(\pi - \pi^0)' (\pi - \pi^0) \quad (19)$$

Y al desarrollar el método  $\delta$  nos damos cuenta que este estadístico de prueba  $d_e^2(\pi, \pi^0) = \sum_{j=1}^k (\pi_j - \pi_j^0)^2$  sigue una distribución normal como se observa en la fórmula 20 :

$$(20) \quad \begin{array}{ccc} \text{Distribución} & \text{Distribución} & \\ \text{estimada} & \text{teórica} & \\ \downarrow & \downarrow & \\ \text{conforme } n \rightarrow \infty & \text{Fórmula (19)} & \end{array} \quad \sqrt{n}(d_e^2(\hat{\pi}, \pi^0) - d_e^2(\pi, \pi^0)) \leq N(0; \sigma^2[\sqrt{n} d_e^2(\hat{\pi}, \pi^0)])$$

Aún más el estimador de la varianza es:

$$v^2(\hat{\pi}, \pi^0) = 4[\sum_{j=1}^k (\hat{\pi}_j - \pi_j^0)^2 \hat{\pi}_j - \sum_{j_1=1}^k \sum_{j_2=1}^k (\hat{\pi}_{j_1} - \pi_{j_1}^0)(\hat{\pi}_{j_2} - \pi_{j_2}^0) \hat{\pi}_{j_1} \hat{\pi}_{j_2}] \quad (21)$$

Por tanto

$$d_e^2(\hat{\pi}, \pi^0) \sim N(\theta, \frac{v^2(\hat{\pi}, \pi^0)}{n}) \quad (22)$$

**Regla de decisión:**

$$\begin{aligned} H_0: d_e^2(\pi, \pi^0) &\geq \varepsilon^2 \\ H_1: d_e^2(\pi, \pi^0) &< \varepsilon^2 \end{aligned}$$

Dónde  $\varepsilon^2$  es la irrelevancia o equivalencia a cero, con base a la distribución asintótica normal de  $d_e^2(\pi, \pi^0)$ , el límite superior del intervalo de confianza es **luup**.

Rechazar  $H_0$  si el límite superior es menor que  $\varepsilon^2$ .

$$luup = d_e^2(\hat{\pi}, \pi^0) + Z_\alpha \frac{v(\hat{\pi}, \pi^0)}{\sqrt{n}} \quad (23)$$

Entonces: Rechazo  $H_0$  (falta de ajuste de  $Mult(n, \pi) = Mult(n; \pi_1, \dots, \pi_k)$  a  $Mult(n, \pi^0) = Mu; lt(n, \pi_1^0, \dots, \pi_k^0)$ ) si:

$$luup < \varepsilon^2$$

Así es como construyo la función normal test.

### 2.1.4.3 Comparación Bondad Vs Falta de Ajuste

**Tabla 8-2.** Diferencia entre la prueba de bondad de ajuste y falta de ajuste

Bondad de Ajuste	Falta de Ajuste
$H_0: d_c(\pi, \pi^0) = 0$ ( <i>Normalidad Perfecta</i> ) $H_1: d_c(\pi, \pi^0) \neq 0$	$H_0: d_e^2(\pi, \pi^0) \geq \varepsilon^2$ $H_1: d_e^2(\pi, \pi^0) < \varepsilon^2$ ( <i>Normalidad Irrelevante</i> )
No rechazar $H_0$ no me dice existe evidencia para decir que la distribución es distinta de una normal, lo cual no necesariamente implica normalidad.	Rechazar $H_0$ si me permite decir que existe evidencia para decir que la distribución es normal. Si existe evidencia para decir que $H_1$ es normal.

**Realizado por:** María Palacios, 2022

Basados en la propuesta de Chi Cuadrado el test tradicional utilizado en nuestra investigación para probar normalidad, donde se establece como hipótesis nula que si la distancia entre  $\pi$  y  $\pi^0 = 0$ , decimos que hay normalidad si la distancia es exactamente igual a cero en este sentido está muy mal ya que como George Box (1979) lo menciona “*en la naturaleza no existe una distribución perfectamente normal, sin embargo, con suposiciones normales, que se sabe que son falsas, a menudo se puede derivar resultados que coinciden, con una aproximación útil a los que se encuentran en el mundo real*”, entonces en la vida real lo que tenemos son datos como podríamos tratar de probar algo que teóricamente sabemos que no existe, no deberíamos decir que es igual a cero sino equivalente a cero, de este modo al utilizar la metodología tradicional para contrastar las hipótesis ( $H_0 = d_c(\pi, \pi^0) = 0$  y  $H_1 = d_c(\pi, \pi^0) \neq 0$ ), presenta un inconveniente lógico ya que no rechazar la hipótesis nula al comparar la distancia entre la distribución objetivo y la distribución observada, solo indica que no hay las pruebas suficientes para determinar la veracidad de este supuesto o como lo dijo Altman y Bland (1995): “*Ausencia de evidencia no es evidencia de ausencia*”, de este modo lo único que podemos probar es que no existe evidencia suficiente para decir que la hipótesis alternativa  $H_1$  (hipótesis que niega el perfecto cumplimiento del supuesto) es falsa, Wellek (Wellek, 2010) dice que el problema radica en la forma como han sido planteados los test tradicionales, ya que no deberíamos probar que existe una perfecta normalidad sino algo que sea irrelevante aproximado a la normalidad.

De acuerdo a los autores podemos concluir que en la vida real no existen datos perfectamente normales, por lo cual no tiene sentido utilizar estas técnicas tradicionales para contrastar hipótesis



de normalidad perfecta. En su lugar Wellek propone un nuevo enfoque basado en equivalencia, entendiendo por equivalencia a una forma dilatada de una igualdad, donde se determina una zona que se considera irrelevante como para determinar que existen diferencias significativas entre los parámetros comparados, además se plantea una nueva forma de contrastar las hipótesis, se establece como la hipótesis nula la no equivalencia y la hipótesis alternativa la equivalencia, en este sentido, el criterio de Cochran (1942) establece que una prueba de hipótesis se puede considerar adecuada si su TIEP está alejada máximo a un 20% por encima y por debajo del nivel de significancia [ $\alpha \pm 0.20 \alpha$ ].

## 2.2 Método de Fleishman

En inferencia estadística muchos de los enunciados de probabilidad son basados en el supuesto de normalidad por ejemplo el rechazo de la hipótesis nula, al probar la igualdad de varianzas provenientes de valores no normales poblacionales, al analizar datos al usar estadística se hace implícitamente la suposición de que los datos son muestreados de una población normal, por eso es muy importante considerar la no normalidad para estudiar su efectos y de este modo incluso determinar la robustez de los test utilizados tradicionalmente.

Existen muchos procedimientos útiles para generar datos no normales, aunque en el ámbito de las ciencias sociales y de la salud cabe destacar el método de transformación lineal de Fleishman y el método propuesto por (Berrocal, 2016) que se basa en la generalización de la distribución lambda de Tukey.

El método desarrollado por Allen Fleishman en el año de 1978, es uno de los más utilizados en estudios de simulación, y tiene la ventaja de ser simple, rápido y fácilmente generalizable a la generación de datos multivariados no normales, mediante el procedimiento descrito por (Ciarla, 2018).

El procedimiento propuesto por Fleishman (1978) utiliza una transformación polinomial para generar datos no normales. Específicamente, toma la suma de una combinación lineal de una variable aleatoria normal, su cuadrado y su cubo (con cuatro constantes) (Agut, 2018) . Como se muestra en la fórmula 18 donde Y es una variable de distribución desconocida y con parámetros:  $\mu = 0, \quad \sigma^2 = 1, \quad \gamma_1, \gamma_2$

$$Z = a + bX + cX^2 + dX^3 \quad (24)$$

O

$$\mathbf{Z} = ((\mathbf{dX} + \mathbf{c})\mathbf{X} + \mathbf{b})\mathbf{X} + \mathbf{a} \quad (25)$$

X es una variable aleatoria normalmente distribuida con media 0 y varianza 1,  $X \sim N(0,1)$  y tendrá una distribución dependiente sobre las constantes, el valor de la constante  $a$  será igual a  $-c$  cómo se observa a continuación.

Para resolver las constantes, es necesario encontrar los diversos momentos de Z. El primer momento (media) se puede encontrar por:

$$\mathcal{E}(Z) = a + b\mathcal{E}(X) + c\mathcal{E}(X^2) + d\mathcal{E}(X^3) \quad (26)$$

Dado que  $X \approx N(0,1)$ , luego  $\mathcal{E}(X) = 0$ ,  $\mathcal{E}(X^2) = 1$ ,  $\mathcal{E}(X^3) = 0$ ,  $\mathcal{E}(X^4) = 3$ ,  $\mathcal{E}(X^5) = 0$ ,  $\mathcal{E}(X^6) = 15$ ,  $\mathcal{E}(X^8) = 105$ ,  $\mathcal{E}(X^9) = 0$ ,  $\mathcal{E}(X^{10}) = 945$ ,  $\mathcal{E}(X^{11}) = 0$ ,  $\mathcal{E}(X^{12}) = 10395$ ,  $\mathcal{E}(X^{13}) = 0$ ,  $\mathcal{E}(X^{14}) = 135135$

Al sustituir en (26) encontramos

$$\mathcal{E}(Z) = a + c \quad (27)$$

Para que la distribución tenga media cero, lo cual es conveniente y suficiente para tener:

$$a = -c \quad (28)$$

Para encontrar la varianza de Z se necesita conocer el segundo momento de Z:

$$\mathcal{E}(Z^2) = a^2 + 2ab\mathcal{E}(X) + (2ac + b^2)\mathcal{E}(X^2) + 2(ad + bc)\mathcal{E}(X^3) + (2bd + c^2)\mathcal{E}(X^4) + 2cd\mathcal{E}(X^5) + d^2\mathcal{E}(X^6); \quad (29)$$

Sustituyendo los diversos valores esperados se obtiene:

$$\mathcal{E}(Z^2) = a^2 + 2ac + b^2 + 6bd + 3c^2 + 15d^2 \quad (30)$$

Resolviendo y sustituyendo los valores esperados para el tercer y cuarto momento se obtiene:

$$\mathcal{E}(y^2) = a^2 + 3a^2c + 3ab^2 + 18abd + 9ac^2 + 45ad^2 + 9b^2c + 90bcd + 15c^3 + 315cd^2 \quad (31)$$

$$\mathcal{E}(y^4) = a^4 + 4a^3c + 6a^2b^2 + 36a^2bd + 18a^2c^2 + 90a^2d^2 + 36ab^2c + 360abcd + 60ac^3 + 1260acd^2 + 3b^4 + 60b^3d + 90b^2c^2 + 630b^2d^2 + 1260bc^2d + 3780bd^3 + 105c^4 + 5670c^2d^2 + 10395d^4 \quad (32)$$

La varianza se define como:

$$Var(Z) = b^2 + 6bd + 2c^2 + 15d^2 \quad (33)$$

En la tabla, la varianza se fijó arbitrariamente en 1. Con medidas de asimetría:

$$\gamma_1 = \frac{k_3}{k_2^{3/2}} \quad (34)$$

y curtosis

$$\gamma_2 = \frac{k_4}{k_2^2} \quad (35)$$

Dónde ( $k_i$ ) son los cumulantes dados por Kendall&Stewart,1969.

$$k_2 = \mathcal{E}(Z^2) - [\mathcal{E}(Z)]^2 \quad (36)$$

$$k_3 = \mathcal{E}(Z^3) - 3\mathcal{E}(Z^2)\mathcal{E}(Z) + 2[\mathcal{E}(Z)]^3, \quad (37)$$

Y

$$k_4 = \mathcal{E}(Z^4) - 4\mathcal{E}(Z^3)\mathcal{E}(Z) - 3[\mathcal{E}(Z^2)]^2 + 12\mathcal{E}(Z^2)[\mathcal{E}(Z)]^2 + 6[\mathcal{E}(Z)]^4, \quad (38)$$

Resolviendo (34) y (35) en términos de coeficientes y haciendo que la media sea cero ( $c=-a$ ) da:

$$\gamma_1 = 2c(b^2 + 24bd + 105d^2 + 2) \quad (39)$$

Y

$$\gamma_2 = 24(bd + c^2[1 + b^2 + 28bd] + d^2[12 + 48bd + 141c^2 + 225d^2]) \quad (40)$$

Resolviendo (39) para c, da:

$$c = \frac{\gamma_1}{2(b^2 + 24bd + 105d^2 + 2)} \quad (41)$$

Al establecer la varianza en 1.0 y usar (33) y (41) se obtiene:

$$2 = 2b^2 + 12bc + \frac{\gamma_1^2}{(b^2 + 24bd + 105d^2 + 2)^2} + 30d^2 \quad (42)$$

Por lo tanto, los cuatro coeficientes (a,b,c y d) se pueden encontrar con la restricción de que la media, la varianza, asimetría y la curtosis son 0,1, $\gamma_1$  y  $\gamma_2$  respectivamente. Esto se hace resolviendo (40) y (42), obteniendo b y d, siendo c y a dados por (41) y (37). Dado que (40) y (42), son expresiones no lineales bastante complicadas, una solución iterativa se utilizó en base a una modificación del método de Newton.

Este procedimiento calcula los coeficientes a, b, c y d mediante una transformación polinómica de acuerdo a valores del tercero y cuarto momento es decir la asimetría ( $\gamma_1$ ) y curtosis  $\gamma_2$ . Sabiendo que la asimetría y la curtosis son los indicadores más importantes de la extensión de la curva y definen los grados de separación o el grado de contaminación de la normal (Bendayan et al., 2014).

### 2.3 Potencia de una prueba

Actualmente existen herramientas estadísticas de fácil manejo que nos permiten determinar la fiabilidad de nuestras pruebas estadísticas y el tamaño muestral necesario. El cálculo de la

potencia de una prueba estadística y su estrecha relación con el tamaño muestral (Bono & Arnau, 1995).

La potencia de una prueba estadística se calcula como  $1-\beta$  y se entiende como la probabilidad de afirmar que la  $H_1$  es verdadera (Quesada & Figuerola, 2010). También se puede entender como la probabilidad de rechazar  $H_0$  cuando de hecho es falsa (Calvillo & Ramos Quiroga, 2017). El interés de la potencia de una prueba estadística estriba en que los análisis estadísticos más comúnmente usados están diseñados para “controlar” el error de tipo I obviándose tradicionalmente el error de tipo II, por lo que el cálculo de la potencia de una prueba estadística también se puede entender como una medida de confianza del análisis que hemos realizado, principalmente cuando no hemos obtenido un resultado significativo. Así, si disminuimos el error de tipo II estaremos aumentando la potencia de la prueba estadística.

## CAPÍTULO III

### DESARROLLO DE LA INVESTIGACIÓN

#### 3.1 Metodología

En la presente investigación se realizó pruebas bajo el enfoque de equivalencia y Chi Cuadrado con el fin de estimar la TIEP global de las pruebas de comparación de medias cuando se pretestea el supuesto de normalidad, para ello mediante el lenguaje de programación R se genera muestras aleatorias normales, donde lo primero que se debe encontrar son los límites de equivalencia denominado épsilon cuadrada  $\varepsilon^2$  válido.  $\varepsilon^2$  es el umbral de equivalencia respecto a cero, es decir, es la distancia máxima que puede existir entre cero y un punto para que la desigualdad sea irrelevante (la diferencia a cero sea irrelevante). Para lo cual, se realizará un test de normalidad de acuerdo a la definición del criterio de Cochran, donde la hipótesis nula corresponde a los datos que no son equivalentemente normales, mientras que la hipótesis alternativa hace referencia a los datos equivalentemente normales. Si el p valor es menor al nivel de significancia se rechazaría  $H_0$ , porque existe evidencia para decir que  $H_1$  es cierta, y se procederá a aplicar la prueba t de Student y si no, se rechazaría la hipótesis nula y se realizaría la prueba de Wilcoxon. Finalizado este proceso, se procederá a comparar la TIEP y Potencia de los distintos métodos de comparación de medias para identificar el test más adecuado.

Entonces primero se planteó la regla de decisión utilizando de Criterio de Cochran:  $\alpha \pm 20\% * \alpha$ , donde los parámetros determinados son  $\alpha = 0.05$  como nivel de significancia,  $\text{diffMean } \bar{x} - \bar{y} = 0$  como la diferencia de medias. Se desarrolló un algoritmo de simulación donde se creó una función `opt_eps` en base al tamaño de la muestra y se contaminó la misma a través de los coeficientes de Fleishman, los mismos que fueron generados por medio de saltos iterativos de la asimetría y curtosis, generando así 6 niveles de contaminación, respecto a la muestra mientras menos simétrica sea y de acuerdo a la curtosis que determina la curva, se fue verificando su validez dentro del criterio de Cochran en caso de no cumplir con el criterio la simulación se detiene y acoge los últimos valores que se encontraban dentro del límite planteado, hasta este punto se puede considerar que la muestra se asemeja a una distribución normal.

Una vez determinados estos límites se procede al cálculo de la distancia euclídea al cuadrado entre la distribución observada  $\pi$  de acuerdo a los coeficientes de Fleishman y la distribución objetivo  $\pi^0$ , con su respectivo intervalo de confianza donde el límite superior será considerado como el  $\varepsilon^2$  el mismo que toma valores aleatorios de acuerdo al tamaño de la muestra. Para

determinar el  $\epsilon$  óptimo se calculó la raíz cuadrada del  $\epsilon^2$ , determinado como el valor máximo que se puede alejar de cero y poder considerar la muestra aproximada a una distribución normal.

Se desarrolló la función NormalTesT donde se aplicaran los test de normalidad y los parámetros a usar son la muestra contaminada con los últimos coeficientes de Fleishman y su respectivo  $\epsilon$  óptimo calculado, aquí se aplica primero el método tradicional que en nuestra investigación es el test Chi Cuadrado en base a la muestra contaminada se obtiene el p -valor para determinar si la muestra es o no es normal, el siguiente proceso es aplicar el test bajo el enfoque de equivalencia (test de Wellek) con los mismo parámetros ingresados se calcula la distancia euclídea al cuadrado entre el vector de la distribución observada y el vector de la distribución teórica del valor calculado se determina el intervalo de confianza donde el límite superior será considerado para rechazar o no la hipótesis nula. Al final la función devuelve la decisión de cada uno de los test.

En la función rejectH0 de acuerdo al resultado obtenido en la función anterior, se determina el test de comparación de medias a utilizar, dependiendo si los datos siguen una distribución normal se aplica t-Student sino se aplicaría el test no paramétrico de Wilcoxon, luego se utilizó el método de simulación de Montecarlo para estimar la probabilidad de error tipo I, donde se estableció un total de 10000 simulaciones con diferentes tamaños de pares de muestra establecidos en la tabla 9-3, en la cual contabiliza el número de veces que cada test comete un error de tipo I, seguidamente se implementó la función TIEP\_POT que permite calcular la TIEP y la potencia utilizando la sintaxis de la función antes mencionada rejectH0, el parámetro diffMean toma un valor de 0 para calcular la TIEP y un valor de 3 para calcular la potencia de los 6 niveles de contaminación planteados.

Finalmente, se gráfica los resultados obtenidos con el propósito de realizar una comparativa visual de la potencia y además verificar si existe una alteración en la probabilidad de error tipo I al realizar los diferentes pretest, para cada uno de los pares de muestra y niveles de contaminación.

### ***3.1.1 Tamaños de las muestras***

Mediante R se generan muestras aleatorias normales, con el fin de analizar y observar posibles cambios que se generan al variar el tamaño de la muestra, la contaminación en la muestra normal, TIEP y POTENCIA al someter al pretest, por tanto, en el siguiente trabajo se utilizó los siguientes pares de muestras.

**Tabla 9-3.** Muestras balanceada y desbalanceada

<b>n1</b>	<b>n2</b>
5	5
3	7
7	3
10	10
6	14
14	6
5	10
10	5

**Realizado por:** María Palacios, 2022

En este apartado se detalla los diferentes pasos y sus diversas etapas hasta lograr al objetivo principal de la investigación. Para esto se plantea diversas funciones que son detalladas a continuación.

### **3.1.2 Caracterización de funciones**

#### **3.1.2.1 Función rnonorm**

En la presente investigación se utilizó el método de Fleishman que nos permite crear muestras no normales, es un método muy utilizado en estudios de simulación, su ventaja es que muy sencillo y rápido de utilizar, siendo muy útil ya que por medio de este procedimiento se puede analizar casos posibles que se presentan al momento de procesar datos reales donde su distribución presente un alejamiento irrelevante y se confirma su proximidad a una distribución normal.

Entonces por medio del método de Fleishman se generan muestras de distribución desconocida, independientemente del tamaño muestral y con un alejamiento de la normal medible, que permite detectar y controlar los grados de contaminación de la normal que se pueden considerar irrelevantes, dichos coeficientes varían de acuerdo a los valores asignados de asimetría y curtosis, se puede establecer un nivel de contaminación. Al buscar de forma iterativa estos valores existen casos en los cuales la combinación de valores asignados a la asimetría y curtosis no son los adecuados, para hallar los coeficientes de Fleishman ya que se obtienen valores indeterminados, de acuerdo a este suceso buscar de manera manual dichos coeficientes se torna complejo, por lo cual se crea una función que revise todos los casos posibles al calcular los coeficientes de Fleishman, esta función propone un salto de error para los valores de asimetría y curtosis con las cuales no se puede calcular dichos coeficientes.

Se aplicó el método de Fleishman con la ayuda de funciones existentes en el software estadístico R, a partir de la función “fleishman.coef” implementada en la librería BinNonNor cuyos parámetros se establecieron en base a distintas combinaciones de la simetría y curtosis, según

corresponda la intensidad requerida es decir grados diferentes de contaminación de la normalidad. Para lo cual se genera el siguiente código:

La función `rnonorm` utiliza el parámetro `n` para identificar los tamaños de muestra y `f.coef` ingresa valores de coeficientes de Fleishman calculados mediante la función `fleishman.coef` dichos valores son estimados de acuerdo a las iteraciones de asimetría y curtosis.

Los parámetros de la función son:

```
rnonorm(n, f.coef)
```

Argumentos

`n` Entero que determina el tamaño de la muestra  
`f.coef` Vector de coeficientes de Fleishman

Valores de retorno

`x` Valores contaminados de la muestra

### 3.1.2.2 *Error de tipo I de t-Student (TIEP)*

Mediante la simulación de Monte Carlo que consta de 10000 repeticiones se verifica las veces que el test t-Student se equivoca al comparar los distintos pares de medias, el resultado final de la función será la proporción de veces que se equivoca al comparar el valor teórico de diferencia de medias igual a cero (`diffMean=0`). Este algoritmo se aplica para los diferentes tamaños de muestras con sus respectivos niveles de contaminación, a continuación, se detalla los parámetros de la función.

#### 3.1.2.2.1 *Función tiepT*

Calcula la proporción de cometer un error tipo 1 al realizar la prueba t-Student.

```
tiepT(nSim, n1, n2, f.coef, diffMean = 0, alpha = 0.05)
```

Argumentos

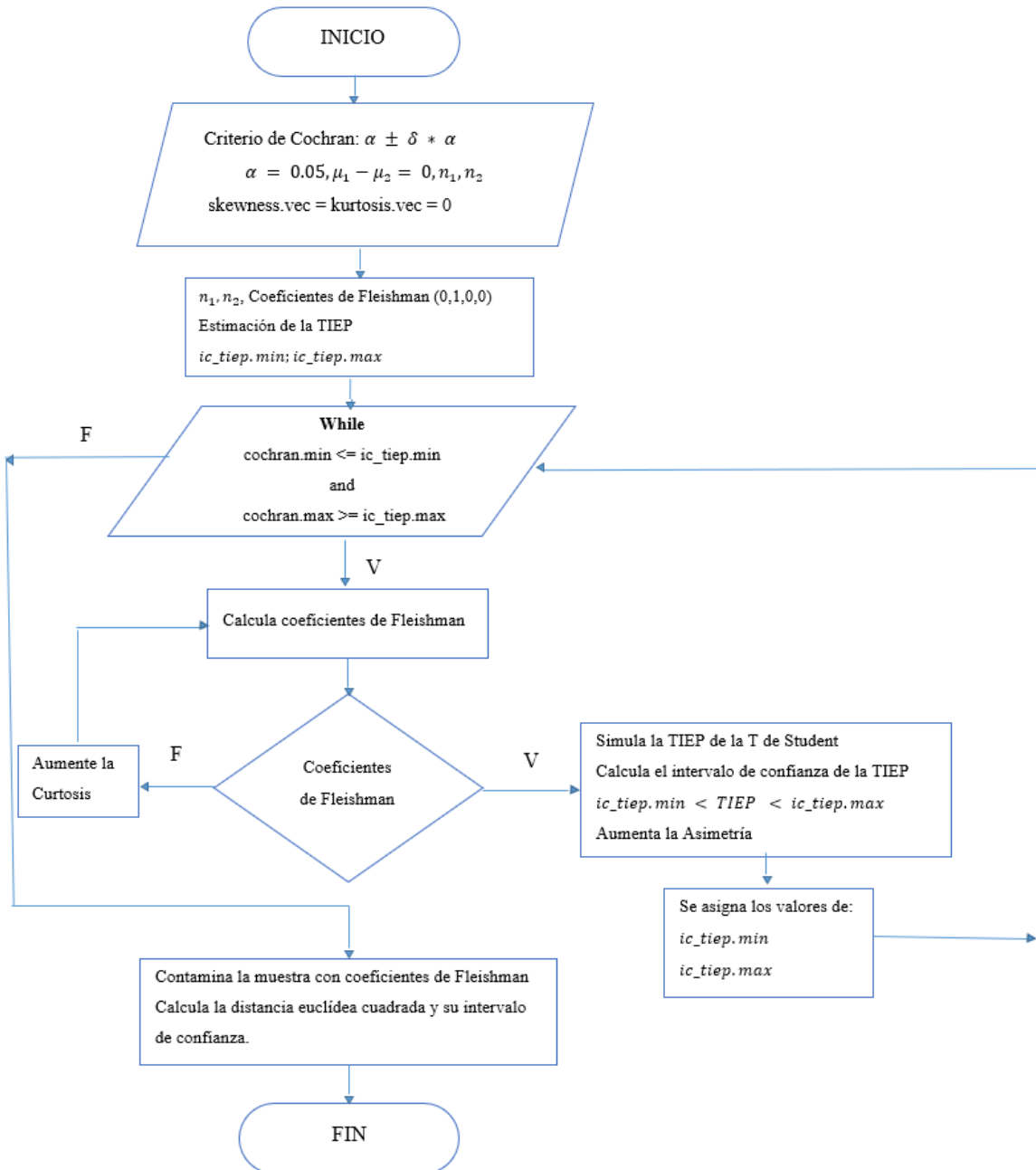
`nSim` Entero que determina el número de simulaciones  
`n1, n2` Valor entero que determina el tamaño de la muestra 1 y 2  
`f.coef` Vector de coeficientes de Fleishman  
`diffMean` Diferencia de medias por default es cero  
`alpha` Nivel de significancia de la prueba por default es 0,05



Valores de retorno

x Probabilidad de cometer el error tipo 1 de la prueba T

### 3.1.2.3 Épsilon o Irrelevancia



**Figura 10-3.** Proceso de hallar épsilon  
**Realizado por:** María Palacios, 2022

El  $\epsilon^2$  es la equivalencia a cero o irrelevancia, se utilizó como la contraparte de la prueba de bondad de ajuste (Prueba de desajuste), en este sentido se tiene como hipótesis nula la no

normalidad, el gran reto es encontrar el valor de  $\varepsilon^2$ , para ello se utiliza la función que genere muestras aleatorias contaminadas con los coeficientes de fleishman, empezando con una muestra normal y aumentando la contaminación verificando la validez de la TIEP al aplicar t-Student con un número de simulaciones de 10000 veces, de este modo verificar cuantas veces rechaza  $H_0$  a pesar de ser verdadera, estimar la TTIEP mediante un intervalo de confianza para una proporción p que estima una probabilidad:

$$ic\_tiep = \widehat{TIEP} \pm Z_{\frac{\alpha}{2}} * \sqrt{\frac{\widehat{TIEP} * (1 - \widehat{TIEP})}{n}}$$

Analizo si esta estimación está dentro del criterio de Cochran,  $\alpha = 0.05 \rightarrow [0.04 - 0.06]$  contaminando tanto con los coeficientes de Fleishman=(a,b,c,d) hasta que llegue a un punto donde la TIEP se encuentre en los valores extremos del criterio de Cochran, identificando así la contaminación máxima a ser utilizada para calcular la distancia euclídea al cuadrado de la diferencia entre las distribuciones verdaderas  $\pi$  y las distribuciones observadas  $\pi^0$ .

Para estimar el épsilon óptimo se procede a calcular el intervalo de confianza de la distancia euclídea al cuadrado, el límite superior será la irrelevancia  $\varepsilon^2 = \mathbf{lupp}$

$$\mathbf{lupp} = d_e^2(\hat{\pi}, \pi^0) + Z_{\alpha} \frac{v(\hat{\pi}, \pi^0)}{\sqrt{n}}$$

Ahora se procede a calcular la raíz cuadrada de  $\mathbf{lupp}$  que es el épsilon óptima, valor que se utilizara en el test de equivalencia. Los parámetros de la función se detallan a continuación:

### 3.1.2.3.1 Función *opt\_eps*

Determinar la irrelevancia  $\varepsilon^2$ .

```
opt_eps(nSim, n1, n2, diffMean = 0, alpha = 0.05, delta = 0.2,
salto.skew, salto.kurt, epsilon = 0.35)
```

Argumentos

nSim Entero que determina el número de simulaciones

n1, n2 Valor entero que determina el tamaño de la muestra 1 y 2

diffMean	Diferencia de medias por default es cero
alpha	Nivel de significancia de la prueba por default es 0,05
delta	Valor aceptado para el intervalo de Cochran de 0,2
salto.skew	Determina el valor de variación de la asimetría
salto.kurt	Determina el valor de variación de la Curtosis
epsilon	Valor de referencia de la irrelevancia por default 0,35

Valores de retorno

- x Irrelevancia de la prueba  $\varepsilon^2$  o límite de contaminación de la normal

### 3.1.2.4 Falta de ajuste

La función `normequiv` fue elaborado con el fin de analizar si una muestra de datos proviene de una distribución normal aplicando el test de equivalencia donde sus parámetros de ingreso son un vector de muestra contaminada y su valor  $\varepsilon$  óptimo encontrado o hallado en la sintaxis de `opt_eps`.

El test de equivalencia calcula los intervalos de clase de la muestra contaminada y calcula su probabilidad teórica de acuerdo a los intervalos de clase calculados, una vez obtenidos estos valores se procede al cálculo de la distribución empírica u observada. Se procede al cálculo de la distancia euclídea al cuadrado de la diferencia entre los vectores de distribución teórica y vectores de distribución observada, como se observa en la expresión:

$$d_e^2(\pi, \pi^0) = \sum_{j=1}^k (\pi_j - \pi_j^0)^2$$

Seguidamente se procede a estimar el intervalo de confianza donde el límite superior será utilizado para comparar con el  $\varepsilon$  óptimo como se observa a continuación:

$$H_0: d_e^2(\pi, \pi_0) \geq \varepsilon^2$$

$$H_1: d_e^2(\pi, \pi_0) < \varepsilon^2$$

Como resultado final de la función será la decisión de rechazar o no rechazar la hipótesis nula. A continuación, se detalla los parámetros de ingreso de la función:

### 3.1.2.4.1 *Función normequiv*

Analizar si una muestra de datos proviene de una distribución normal usando la irrelevancia óptima.

```
normequiv(x, alpha = 0.05, epsilon = 0.15, ...)
```

Argumentos

- x Conjunto de datos
- alpha Nivel de significancia de la prueba por default es 0,05
- epsilon Valor de referencia de la irrelevancia por default 0,35

Valores de retorno

- Euclidean distance Es la distancia euclídea al cuadrado calculada
- Upper bound Límite superior de la distancia euclídea
- epsilon^2 es el valor de épsilon optimo elevado al cuadrado utilizada en comparación con el upper bound
- dec Muestra la decisión final del test de equivalencia, lo ideal rechazar Ho y asumir normalidad

### 3.1.2.5 *Comparación de test*

#### 3.1.2.5.1 *Función normalTest*

En esta función se desarrolla un algoritmo donde aplique tanto el test de equivalencia y la bondad de ajuste Chi Cuadrado, la sintaxis del test de equivalencia se detalló en la función normequiv, la prueba Chi Cuadrado se desarrolla calculando la distribución teórica y la respectiva distribución observada de la muestra contaminada, seguidamente se procede al cálculo del estadístico Chi Cuadrado que es:

$$X^2 = \sum_{i=1}^k \frac{(\hat{\pi}_i - \pi_i^0)^2}{\pi_i^0}$$

Utilizando el estadístico Chi Cuadrado se estima su probabilidad con k-1 grados de libertad. La hipótesis se rechaza cuando el p-valor es menor que el nivel de significancia  $\alpha$ . Los argumentos de la función se detallan a continuación:

```
normalTest(x, alpha = 0.05, epsilon, ...)
```

Argumentos

x Conjunto de datos  
alpha Nivel de significancia de la prueba por default es 0,05  
epsilon Valor de referencia de la irrelevancia por default 0,35

Valores de retorno

X2Test Decisión de normalidad según Chi cuadrado  
equivTest Decisión de normalidad según el test de equivalencia

### 3.1.2.5.2 *Función rejectH0*

La función rejectH0 utiliza las funciones monorm y normalTest, las pruebas de comparación de medias t-Student y Wilcoxon, se plantea una simulación de Monte Carlo con 10000 réplicas donde se establece estimar la TIEP de los pretest y la comparación de medias de acuerdo a una muestra contaminada y el  $\epsilon$  óptimo, en esta función el parámetro diffMean=0 es la diferencia de medias teóricas y f.coef son los coeficientes de Fleishman hallados con la función opt\_eps.

A continuación, se detalla los parámetros de la función:

```
rejectH0(nSim, n1, n2, epsilon, diffMean = 0, f.coef = c(0, 1, 0, 0),  
alpha = 0.05)
```

Argumentos

nSim Entero que determina el número de simulaciones  
n1, n2 Valor entero que determina el tamaño de la muestra 1 y 2  
diffMean Diferencia de medias por default es cero  
alpha Nivel de significancia de la prueba por default es 0,05  
epsilon Valor de referencia de la irrelevancia por default 0,35  
f.coef Coeficientes de fleishman

Valores de retorno

- x TIEP para el test de equivalencia, Chi-cuadrado, Wilcoxon y tstudent.

### **3.1.2.6 Comparación de TIEP y POTENCIA**

Para poder comparar la variación de la TIEP, al contaminar muestras normales mediante los coeficientes de fleshman y basado en el  $\varepsilon^2$  calculado, se desarrolla una función que permita conocer este valor mediante la simulación de Montecarlo (Berrocal, 2016). Comparar la variación de la Potencia, al contaminar muestras normales mediante los coeficientes de fleshman basados en el  $\varepsilon^2$  calculado, para hallar la potencia se utiliza la misma función para el caso de la TIEP pero con una variante que en la diferencia de medias se utiliza 3 y no 0. Para generar datos no normales, se utilizó una transformación de la función polinomial, a partir de la suma combinada lineal de una variable aleatoria normal ( $z = a + bX + cX^2 + dX^4$ ). En donde Z es una variable con distribución desconocida con media 0 y varianza 1, mientras que X es una variable aleatoria normalmente distribuida con media 0 y varianza 1 (Ciarla, 2018). Debido a que en este procedimiento se calcularon los coeficientes a, b, c y d mediante transformaciones polinómicas de acuerdo a valores del tercero y cuarto momento es decir la asimetría y curtosis.

Básicamente, la función “TIEP\_POT”, estima la Probabilidad global de Cometer un Error Tipo I en una prueba de comparación de medias cuando se pre-testea sus supuestos. El proceso consiste en generar muestras con los distintos niveles de contaminación de normalidad. A partir de estas muestras se aplican los diferentes procedimientos establecidos para determinar en cada una de ellas la proporción de veces que se rechaza la hipótesis de igualdad de medias, cuando esta es cierta. Esta proporción es el estimador deseado de la TIEP global.

#### **3.1.2.6.1 Función TIEP\_POT**

La función TIEP\_POT determina la TIEP y la potencia de los pretest Chi Cuadrado y Equivalencia y de los test de comparación de medias t-Student y Wilcoxon, la sintaxis de esta función utiliza el parámetro diffMean=0 para poder estimar la TIEP mientras que si se cambia el valor de diffMean=3 el resultado será la potencia de cada una de estas pruebas, empleando la función rejectH0. Para esta simulación se plantea realizar 10000 réplicas para cada uno de los pares de muestra y los diferentes niveles de contaminación.

Estos resultados servirán para comparar de manera visual con la ayuda de gráficos y de este modo poder concluir cual es la prueba más eficiente de acuerdo a la probabilidad de error tipo I y la potencia de cada uno de los test empleados en esta investigación. A continuación, se detalla los parámetros de la función:

```
TIEP_POT(nSim, muestra, epsilon.vec, level.cont, level = 0, diffMean = 0, Potencia = 3, alpha = 0.05)
```

#### Argumentos

nSim	Entero que determina el número de simulaciones
muestra	Matriz de valoraciones del tamaño de la muestra 1 y 2
epsilon.vec	Irrelevancia referente al tamaño de la muestra
level.cont	Nivel de contaminación de la muestra
level	Valor de referencia de contaminación por default es cero (sin contaminación)
diffMean	Diferencia de medias para el test
Potencia	Valor de cálculo de potencia del test
alpha	Nivel de significancia.

Además, se utilizaron los siguientes valores de retorno:

'Nivel.Contaminacion'	TIEP para el test de equivalencia, Chi-cuadrado, Wilcoxon y tstudent.
Tamaño.muestra	Tamaño de la muestra incluido en el test
TIEP.t	TIEP de la prueba t
TIEP.Wilcoxon	TIEP de la prueba de Wilcoxon
TIEP.Pretest.JiCUadrado	TIEP de la prueba Chi cuadrado
TIEP.Pretest.Equiv	TIEP de la prueba de Equivalencia
POT.Pretest.Equiv	POT de la prueba t
POT.Pretest.Equiv	POT de la prueba de Wilcoxon
POT.Pretest.Equiv	POT de la prueba Chi cuadrado
POT.Pretest.Equiv	POT de la prueba de Equivalencia

## CAPÍTULO IV

### 4. ANÁLISIS Y RESULTADOS

#### 4.2. Niveles de contaminación

El estudio se plantea un total de 6 niveles de contaminación, donde el nivel cero indica que la muestra no está contaminada y el valor 6 indica una contaminación severa. Para el cálculo de los coeficientes de Fleishman se plantea la asimetría y Curtosis que se observa en la tabla 10-4.

**Tabla 10-4.** Niveles de contaminación

	<i>Niveles de Contaminación</i>	<i>Asimetría <math>\gamma_1</math></i>	<i>Curtosis <math>\gamma_2</math></i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<b>Sin Contaminación</b>	0	0	0	0	1	0	0
<b>Contaminación Baja</b>	1	0.25	0.7	-0,0368	-0,9334	0,0368	-0,0213
<b>Contaminación Moderada</b>	2	0.75	1	-0,1191	0,9559	0,1191	0,0098
<b>Contaminación Alta</b>	3	1.3	2	-0,2491	0,9843	0,2491	-0,0164
<b>Contaminación Severa</b>	4	2	6	-0,3137	0,8263	0,3137	0,0227
<b>Contaminación Severa</b>	5	3	15	-0,3457	0,5883	0,3457	0,0861
<b>Contaminación Severa</b>	6	4	30	-0,3347	0,3077	0,3347	0,1603

Realizado por: María Palacios, 2022

#### 4.3. Épsilon Óptimo

La tabla 10-4 muestra los resultados obtenidos para muestras balanceadas y desbalanceadas donde la decisión final fue tomada de acuerdo al test de equivalencia, tomando al épsilon óptimo como el límite de aceptación, ya que si el valor supera o excede este límite la muestra contaminada ya no es considerada como una distribución normal.

**Tabla 11-4.** Épsilon óptimo para cada par de muestra

<i>i</i>	<i>Muestra</i>	<i>Épsilon Óptimo (<math>\epsilon</math>)</i>	<i>Decisión</i>
1	n1 = 5 n2 = 5	0.614276	Reject H0 - Normality
2	n1 = 3 n2 = 7	0.517345	Reject H0 - Normality
3	n1 = 7 n2 = 3	0.614276	Reject H0 - Normality
4	n1 = 10 n2 = 10	0.640997	Reject H0 - Normality



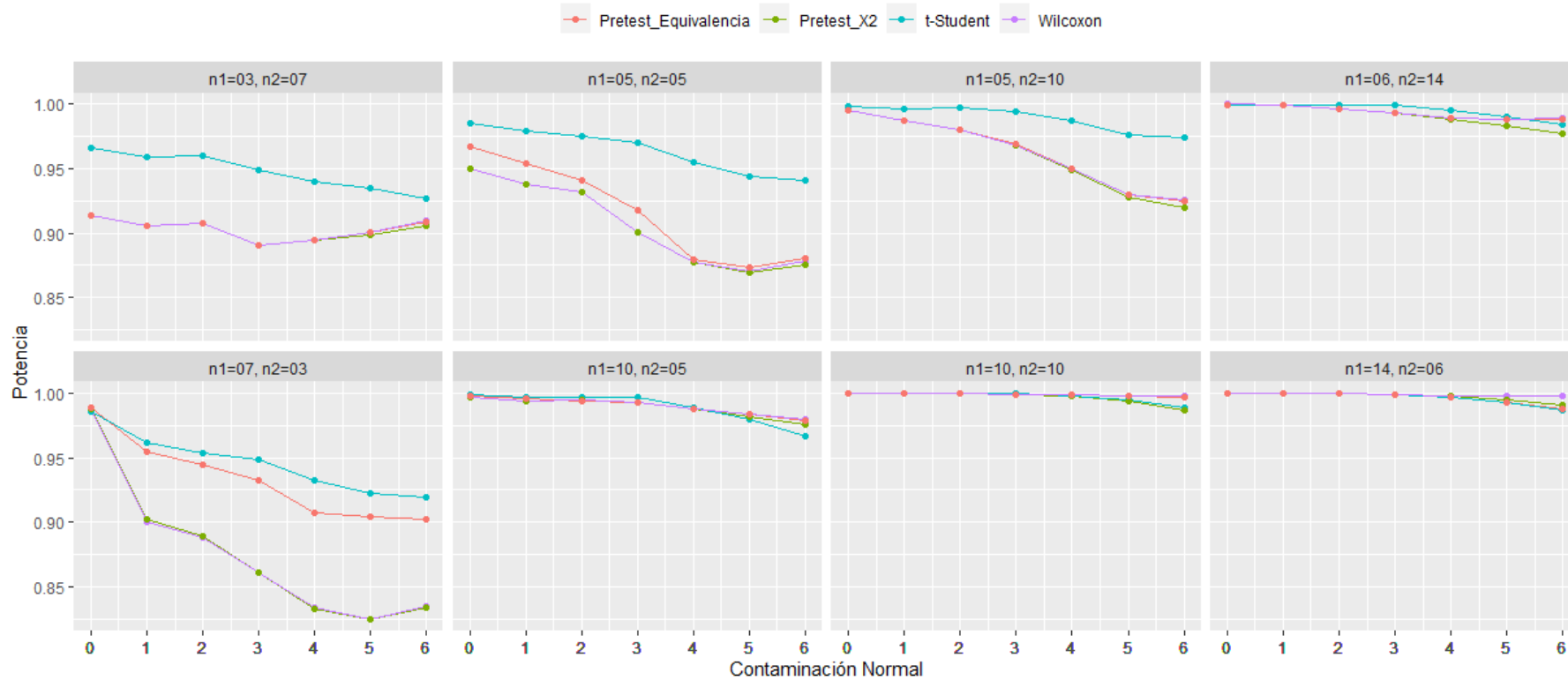
5	n1 = 6 n2 = 14	0.643935	Reject H0 - Normality
6	n1 = 14 n2 = 6	0.643935	Reject H0 - Normality
7	n1 = 5 n2 = 10	0.590953	Reject H0 - Normality
8	n1 = 10 n2 = 5	0.46073	Reject H0 - Normality
9	n1 = 10 n2 = 20	0.459871	Reject H0 - Normality
10	n1 = 20 n2 = 30	0.412614	Reject H0 - Normality

**Realizado por:** María Palacios, 2022

#### 4.4. Potencia de las pruebas de normalidad

Se analizaron 6 distintos niveles de contaminación (0-6) de la normal, los resultados obtenidos de la simulación para cada uno de los pares de muestra y niveles de contaminación y se encuentran en el anexo F, se identificó de acuerdo a lo observado que en la potencia simulada t de Student es la prueba con mayor potencia para todos los niveles de no normalidad y tamaños de muestra, mientras que el pretest de Equivalencia cuando las muestras son pequeñas y balanceadas, bueno en la mayoría de los casos se ve que el pretest de equivalencia tiene mucha mejor potencia que el pretest chi cuadrado.

4.4.1. *Resultados finales*

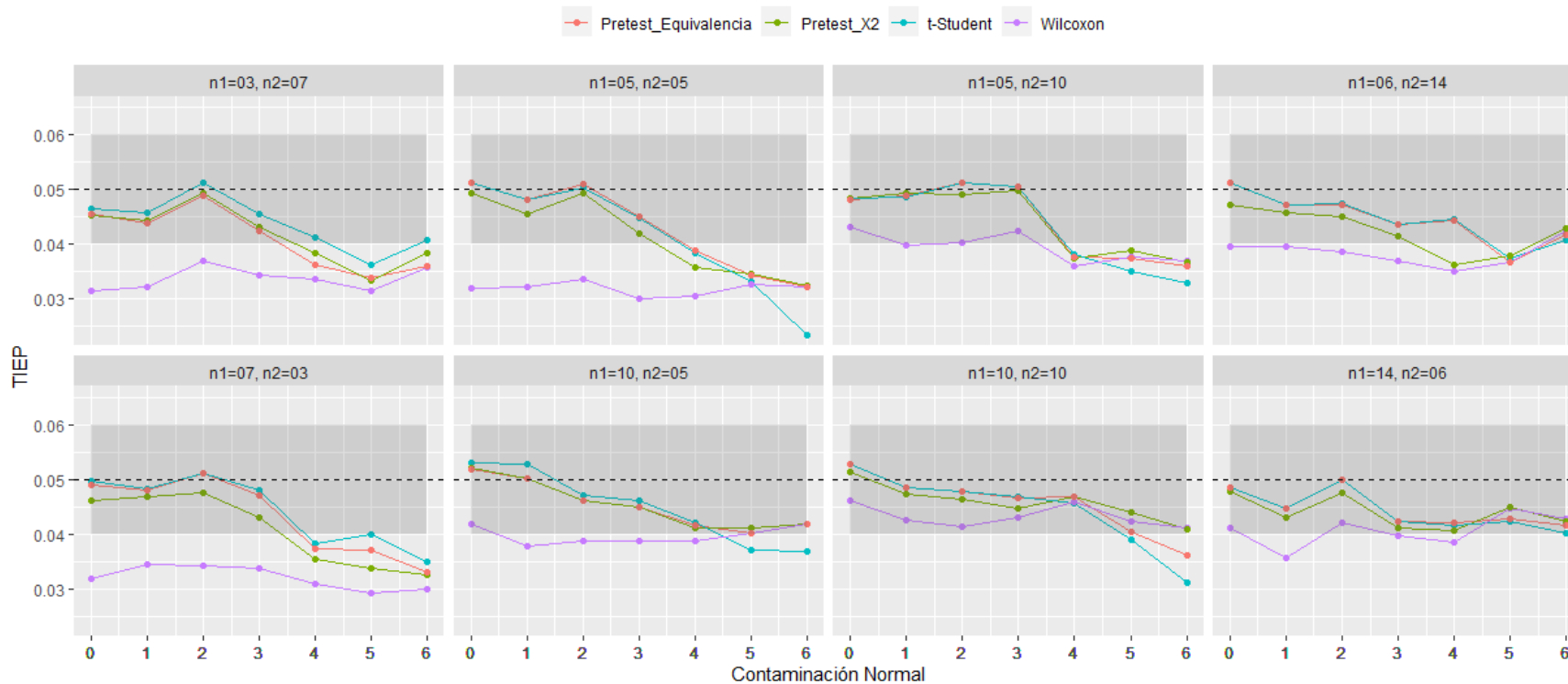


**Gráfico 1-4.** Estimación de la potencia de una prueba de comparación de medias cuando se pre-testean el supuesto de normalidad usando distintos tamaños muestrales con una significancia  $\alpha = 0.05$ .

**Realizado por:** María Palacios, 2022

En la estimación de la potencia cuando dos medias poblacionales son comparadas en tamaños muestrales desbalanceados por ejemplo  $n_1 = 3$  y  $n_2 = 7$  se observa que la potencia para los procedimientos directos, en estos casos, se identificó que el test t-Student fue aquel que evidencio los mejores resultados en todos los casos, como se indica en la Figura 4-1. En el análisis comparativo entre el pretest de equivalencia y chi cuadrado en el caso de tamaños de muestrales  $n_1 = 3$  y  $n_2 = 7$  se determinó que la potencia de los pretest se asemeja, ahora en tamaños muestrales iguales a 5 la potencia del pretest de equivalencia mejora significativamente en comparación con el pretest chi cuadrado, lo mismo ocurre con el tamaño muestral  $n_1 = 7$  y  $n_2 = 3$ , claramente se observa que el pretest de equivalencia cuando las muestras al menos son pequeñas y balanceadas, presenta una mejor potencia que el pretest Chi Cuadrado.

Entonces en la potencia se determinó, que lo más importante es una prueba que cometa menos proporción de errores de tipo I lo que determina la eficiencia de aplicar un test, con menos probabilidad de tomar una decisión incorrecta, por tanto, cuando las muestras al menos fueron pequeñas y balanceadas claramente se observó datos erróneos en los gráficos donde, se concluyó que el pretest de equivalencia tiene mejor potencia que el pretest Chi cuadrado.



**Gráfico 2-4.** Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean el supuesto de normalidad, usando distintos tamaños muestrales con una significancia  $\alpha = 0.05$ .

**Realizado por:** María Palacios, 2022

Luego en el análisis de TIEP global se identificaron las siguientes dispersiones que se observan en la Figura 4-2, parece ser que en la mayoría de las técnicas dan resultados similares, pero al observar punto por punto en casos desbalanceados  $n_1 = 7$  y  $n_2 = 3$  para los niveles 4 y 5 el pretest de equivalencia está más cerca de la línea entrecortada que es el alfa teórico y el pretest chi cuadrado no, para el caso de  $n_1 = 6$  y  $n_2 = 14$  el pretest de equivalencia está casi equivalente a t-Student, tiene una ligera mejora pero no es significativa porque van al a par, no se puede negar que hacer el pretest de equivalencia con el chi cuadrado casi

es igual con una mínima diferencia pero esto es mínimo y pasa en ciertos puntos, por lo que se identificó que para ciertos niveles la prueba de equivalencia que se estudió presenta una relevancia. Por otra parte, la prueba Wilcoxon muestran una TIEP totalmente descontrolada en la mayoría de los casos representados.

En estos casos, se puede observar que los tres procedimientos se mantienen robustos dado que presentan una buena estimación de la TIEP global (no perfecto, pero dentro de los límites establecidos por el criterio de Cochran) con valores muy cercanos al nivel de significancia  $\alpha = 0.05$ , para cualquier nivel de contaminación de la normalidad.

Cuando hay perfecta normalidad se supone que la prueba t Student es una buena opción, el suceso que se presenta es este estudio es que incluso cuando existe contaminaciones Leve, Alta y Severa de la normalidad la prueba t-Student parece ser la mejor, nuevamente es tan buena como para aplicarla directamente. Aunque en la realidad no se puede aplicar t-student directamente, debido a que la teoría estadística indica que se debe probar primero normalidad o verificar los supuestos, es decir nadie lo hace siempre habrá un pretest de por medio.

Finalmente, para contaminaciones Alta y Severa de la normalidad analizando el pretest de Equivalencia y Chi Cuadrado, de acuerdo a las gráficas el pretest de equivalencia resulta ser una mejor opción para verificar el supuesto de normalidad puesto que en la mayoría de escenarios la TIEP se encuentra dentro de los límites de robustez propuestos por Cochran y se demostró tener una mejor potencia que el pretest Chi Cuadrado.

De acuerdo a los resultados obtenidos se puede concluir que la hipótesis planteada si se cumple es decir el pre test normalidad bajo el enfoque de equivalencia si controlo la probabilidad de cometer un error tipo I de igual manera lo hizo el pre test tradicional como efecto de la robustez de la prueba t Student haciendo que la probabilidad de error tipo I se mantenga controlada para ambos métodos, sin embargo es importante hacer notar que las pruebas no paramétricas no controlan de la mejor forma esta probabilidad de error tipo I

#### 4.5. Discusión

Según Wellek (2010), el problema en la alteración de la TIEP no es ocasionado por el proceso de pretesteo, si no, más bien por la forma en como están planteados los test tradicionales. Wellek desarrolló un enfoque, basado en equivalencia, donde se entiende por equivalencia a *igualdad excepto para desviaciones irrelevantes*, las cuales están dadas por un límite de equivalencia, el cual representa un umbral alrededor del cual es permitido desviarse sin causar efectos significativos., Por otro lado Box en su criterio de normalidad indica que “en la naturaleza no existe una distribución perfectamente normal, sin embargo con suposiciones normales, que se sabe que son falsas, a menudo se puede derivar resultados que coinciden, con un aproximación útil a los que se encuentran en el mundo real” (Chimbo & Cruz, 2015). En el presente estudio se plantearon desviamientos de la normalidad leve, moderada, alta y severa, en base al desarrollo del pretest motivo de esta investigación, se obtuvo como resultado que la probabilidad global de cometer un error tipo I parece ser igual en todas las técnicas, inclusive en la que se ha desarrollado, de ahí que sería pertinente analizar punto por punto para determinar si se respeta la probabilidad de error identificada dentro del criterio de Cochran. Al realizar la comparación con el pretest Chi Cuadrado se determina no respeta la probabilidad de error o no controla la TIEP, esto se manifiesta gráficamente mostrando una ligera mejora como se puede observar en la figura 4.1.

La potencia hallada en la experimentación respecto al pretest de equivalencia resulta ser óptima, ya que en la mayoría de los casos se estabiliza en un rango entre 0.9 a 0.95 a diferencia del pretest Chi Cuadrado que con ciertos tamaños muestrales presenta una potencia decreciente en función del tamaño de la muestra.

Finalmente se demuestra que resulta ser más robusto el pretest de equivalencia desarrollado en la verificación del supuesto de normalidad para comparación de medias.

## CONCLUSIONES

- Se realizó el proceso de pre testeo para los diferentes niveles de contaminación con varios pares de tamaños muestrales, donde se verificó el supuesto de normalidad usando la prueba tradicional de bondad de ajuste en este caso específico la prueba Chi Cuadrado, estimando la probabilidad global de cometer un error tipo I.
- Se desarrolló un algoritmo iterativo en función del tamaño de la muestra y el nivel de significancia, donde se generó diferentes niveles de contaminación por medio de los coeficientes de Fleishman, de este modo se obtuvo la máxima contaminación permitida en base a la cual se calculó los límites de irrelevancia de acuerdo al criterio de Cochran.
- Se desarrolló la función normequiv para verificar el supuesto de normalidad bajo el enfoque de equivalencia usando como parámetro de entrada el límite superior de la irrelevancia encontrado, caracterizado por  $\varepsilon$ , mediante la simulación de Montecarlo con 10000 repeticiones se estimó la probabilidad global de cometer un error tipo I.
- Se comparó la efectividad, en términos de la afectación del error tipo I, que tienen las pruebas de comparación de medias cuando se pre-testea el supuesto de normalidad bajo el enfoque tradicional como es la prueba de Bondad de ajuste Chi Cuadrado, comparándolo con el pretest de falta de ajuste desarrollado el pretest de equivalencia, se obtuvieron valores semejantes en el control de la TIEP, en tanto que, respecto a la potencia, el pretest de equivalencia presentó mejores resultados, en el análisis comparativo entre los pretest de equivalencia y Chi Cuadrado se estableció el pretest de equivalencia como la prueba con menor probabilidad de cometer un error tipo I, por lo cual se considera como la opción adecuada para verificar el supuesto de normalidad en las pruebas de comparación de medias.
- De acuerdo a los resultados se concluye que el test de comparación de medias t-Student es una prueba muy robusta ante desviaciones de la normalidad, la teoría estadística dice que la T-Student sirve solo para las muestras normales y Wilcoxon para muestras no normales, sin embargo, aquí se observa todo lo contrario que la T-Student permanece robusta para casos de no normalidad en el control de las alteraciones en la TIEP y Wilcoxon es muy débil, incluso la potencia de t-Student está siempre por encima de la potencia de Wilcoxon. Además, aplicar directamente t-Student resultó ser el mejor de los procedimientos analizados en todos los escenarios propuestos de contaminación de la normal analizada, pero en la vida real no se puede aplicar t-student directamente, nadie lo hace, ya que la teoría estadística dice que se debe probar primero normalidad o verificar los supuestos.

## **RECOMENDACIONES**

Es importante determinar límites de irrelevancia para pre testear normalidad, pues presenta un buen control de la TIEP, además tiene una mejor potencia que los procedimientos bajo el enfoque tradicional.

Antes de realizar las simulaciones se debe considerar una computadora de buen rendimiento con un procesador mayor o igual a una core i5, esto es recomendado ya que, al momento de realizar las simulaciones con 10000 repeticiones, el tiempo de procesamiento fue muy extenso.

Se recomienda que nunca use el test de Wilcoxon para muestras no normales (aunque esto contradice la teoría) es una muy mala alternativa debido a que se tendría una alta probabilidad global de cometer un error de tipo I.

Es mejor no pre testear (bajo ningún método) y en su lugar usar directamente la t-Student.

Posteriormente se sugiere realizar una investigación usando pretest de equivalencia para verificar los supuestos y comparándolo con técnicas directas como por ejemplo t-Student.



## BIBLIOGRAFÍA

- Mendes, M. & Pala, A. (2003). Type I Error Rate and Power of Three Normality Tests. *Information Technology Journal*, 2(2), 135–139. <https://doi.org/10.3923/itj.2003.135.139>
- Agut, S. (2018). *Análisis de necesidades de competencias en directivos de organizaciones turísticas. El papel de la formación.* 479.
- Alcántara Vasconcellos, E. (2010). Análisis de la movilidad urbana. Espacio, medio ambiente y equidad. In *Bogotá, Colombia*.  
<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Análisis+de+la+movilidad+urbana+Espacio+,+medio+ambiente+y+equidad#0>
- Anderson, T. W. & Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268), 765–769.  
<https://doi.org/10.1080/01621459.1954.10501232>
- Bendayan, R., Arnau, J., Blanca, M. J. & Bono, R. (2014). Comparison of the procedures of fleishman and ramberg et al. for generating non-normal data in simulation studies. *Anales de Psicología*, 30(1), 364–371. <https://doi.org/10.6018/analesps.30.1.135911>
- Berrocal, F. (2016). *Análisis comparativo de tres métodos de valoración de puestos de trabajo.* 377. <https://eprints.ucm.es/38833/1/T37653.pdf>
- Bono, R. & Arnau, J. (1995). Consideraciones generales en torno a los estudios de potencia. *Anales de Psicología*, 11(2), 193–202.
- Bowen, J. Van, Canavos, G. C., Hoel, P. G. & Kemp, K. W. (1986). Applied Probability and Statistical Methods. In *Journal of the American Statistical Association* (Vol. 81, Issue 393, p. 258). <https://doi.org/10.2307/2288011>
- Box, G. E. P. (1979). Robustness in the Strategy of Scientific Model Building. In *Robustness in Statistics*. ACADEMIC PRESS, INC. <https://doi.org/10.1016/b978-0-12-438150-6.50018-2>
- Brown, M. B. & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346), 364–367.  
<https://doi.org/10.1080/01621459.1974.10482955>
- Calvillo, E. M. & Ramos Quiroga, R. (2017). *Validación y confiabilidad de pruebas de opción múltiple para la evaluación de habilidades Que para obtener el grado de.* 1–131.
- CEPAL. (2018). Las pruebas de hipótesis. In *Universidad de Guanajuato* (p. 12).  
<https://blogs.ugto.mx/enfermeriaenlinea/unidad-didactica-3-las-pruebas-de-hipotesis/>
- Chimbo, D. & Cruz, D. (2015). *LEVALUACIÓN DEL POTENCIAL FORRAJERO Y COMPOSICIÓN NUTRICIONAL DEL PASTO HÍBRIDO CUBA OM-22 (Pennisetum purpureum Schumacher x Pennisetum glaucum L.) A TRES EDADES DE CORTE.*
- Ciarla, G. A. (2018). *El método de análisis de inteligencia en el modelo de planeamiento militar*

- para la defensa argentino . Tutor : Dr Jorge Claudio Szeinfeld Autor : Gabriel Andrés Ciarla Fecha :  
[http://sedici.unlp.edu.ar/bitstream/handle/10915/32313/Documento\\_completo\\_\\_.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/32313/Documento_completo__.pdf?sequence=1)
- Clark, M. L. (2004). The P values and the confidence intervals: Which to trust? *Revista Panamericana de Salud Publica/Pan American Journal of Public Health*, 15(5), 293–296.  
<https://doi.org/10.1590/s1020-49892004000500001>
- Cobo, E., Cortés, J., Gonzales, J. A., Riba, L., Pealez, R., Vilaro, M. & Bielsa, N. (2014). Prueba de Significación y Contraste de Hipótesis. *Bioestadística Para No Estadísticos*, 44.
- Devore, J. L., Berk, K. N. & Carlton, M. A. (2021). *Overview and Descriptive Statistics*.  
[https://doi.org/10.1007/978-3-030-55156-8\\_1](https://doi.org/10.1007/978-3-030-55156-8_1)
- Dipp, A., Serrano, Á. & González, C. (1992). *Estudio comparativo de la calidad educativa entre dos modelos pedagógicos a nivel preescolar a.* 1–11.
- Fallas, J. (2014). Pruebas de Hipotesis. *Creative Commons*, 161–167.  
[http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/prueba\\_hipotesis\\_2012.pdf](http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/prueba_hipotesis_2012.pdf)
- Flores, P., Ocaña, J. & Sánchez, T. (2018). Verificación De Supuestos En Las Pruebas De Comparación De Medias. Una Revisión. *Ciencia Digital*, 2(4.1.), 5–22.  
<https://doi.org/10.33262/cienciadigital.v2i4.1..187>
- Gmurman V E - *Teoria De Las Probabilidades Y Estadística Matematica.PDF*. (n.d.).
- Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Inst. Ital. Attuari*, 4, 83–91.
- Kubinger, K. D., Rasch, D. & Yanagida, T. (2011). A new approach for testing the Rasch model. *Educational Research and Evaluation*, 17(5), 321–333.  
<https://doi.org/10.1080/13803611.2011.630529>
- Molina, M., Ochoa, C. & E, O. (2018). 16.
- Moral Peláez, I. (2016). Comparación de medias. *Métodos Estadísticos Para Enfermería Nefrológica*, 165–184.
- Oca, J., Llu, B. J. & Jord, C. (2004). *Facultad de Biología Departamento de Estadística Programa de Doctorado en Estadística , Análisis de Datos y Bioestadística Bienio : 2003-2004 Análisis de Procedimientos para la Evaluación de Medicamentos : Bioequivalencia y Farmacogenética Doctoranda : Ma. 2003–2004.*
- Otzen, T. & Manterola, C. (2017). Técnicas de Muestreo sobre una Población a Estudio. *International Journal of Morphology*, 35(1), 227–232. <https://doi.org/10.4067/S0717-95022017000100037>
- Flores, P. & Ocaña, M. (2018). Heteroscedasticity irrelevance when testing means difference. *Sort*, 42(1), 59–72. <https://doi.org/10.2436/20.8080.02.69>

- Quesada, J. & Figuerola, J. (2010). Potencia de una prueba estadística: aplicación e interpretación en ecología del comportamiento. *Etología*, 22(2010), 19–37.  
<http://www.ebd.csic.es/jordiplataforma/subidas/Etologia2010.pdf>
- Quispe Caray, C. R. (2018). Universidad Nacional De San Universidad Nacional De San Agustín De Arequipa. E“Efecto De La Implementación De La Gestión De Logística Inversa En Los Resultados Económicos Y Medioambientales De La Empresa Industrial Reyemsa Periodo 2017, 164.  
<http://repositorio.unsa.edu.pe/handle/UNSA/10883%0Ahttp://repositorio.unsa.edu.pe/handle/UNSA/4057%0Ahttp://repositorio.unsa.edu.pe/handle/UNSA/8014%0Ahttp://repositorio.unsa.edu.pe/handle/UNSA/6899>
- Rodríguez Carrasco, G. V., Asmat Marín, M. K., Orellana Dextre, Z. M. de los Á. & Salazar Espinoza, K. M. (2017). Análisis del impacto de la modalidad de contratación de personal en la atención al usuario en el servicio de certificados literales, en las oficinas de los distritos de Surco, San Isidro, San Borja y Jesús María (sede Rebagliati) de la Superintendencia Na. *Universidad Peruana De Ciencias Aplicadas*, 89.
- Ronald Walpole, Raymond Myers, S. & Myers. (2012). Probabilidad y estadística para ingeniería y ciencias. In *Probabilidad y estadística para ingeniería y ciencias* (Novena, Vol. 29, Issue 1).
- Sánchez, T. (2015). *Escuela Superior Politécnica De Chimborazo*.
- Shapiro, A. S. S. & Wilk, M. B. (1965). Biometrika Trust An Analysis of Variance Test for Normality ( Complete Samples ) Published by : Oxford University Press on behalf of Biometrika Trust Stable. *Biometrika*, 52(3), 591–611.  
<https://pdfs.semanticscholar.org/1f1d/9a7151d52c2e26d35690dbc7ae8098beee22.pdf>
- Smirnov, N. (1974). Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, 19(2), 279–281.  
<https://doi.org/10.1214/aoms/1177730256>
- Student. (1908). “The Probable Error of a Mean.” *Biometrika*, 6(2984), 1–25.  
<https://doi.org/10.1038/119040a0>
- Tapia-Flores, C. E. & Cevallos-Flores, K. L. (2021). Pruebas Para Comprobar La Normalidad De Datos En Procesos Productivos : Anderson- Darling , Ryan-Joiner , Shapiro-Wilk Y Kolmogórov-Smirnov Tests To Verify the Normality of Data in Production Processes : Anderson- Darling , Ryan-Joiner , Shapiro-Wilk and. *Societas*, 23(2), 83–97.  
<http://portal.amelica.org/ameli/jatsRepo/341/3412237018/index.html>
- Vergara, M. & Babativa, G. (2010). El supuesto de normalidad: ¿mito o realidad? *Equidad y Desarrollo*, 1(13), 127–131. <https://doi.org/10.19052/ed.211>
- Wackerly, D., Mendenhall, W., Schaffer & Richard. (2010). *Estadística Matemática* (S. Cervantes (Ed.); 7ma Edición). Cengage Learning Editores, S. A.

[http://www.economia.unam.mx/biblioteca/Pdf/Estadística matemática con aplicaciones  
7edi\\_MENDENHALL.pdf](http://www.economia.unam.mx/biblioteca/Pdf/Estadística%20matemática%20con%20aplicaciones%207edi_MENDENHALL.pdf)

Welch, B. L. (1947). The Generalization of 'Student's' Problem when Several Different Population Variances are Involved. In *Biometrika* (Vol. 34, Issue 1/2, p. 28).

<https://doi.org/10.2307/2332510>

Wellek, S. (2010). Testing Statistical Hypotheses of Equivalence and Noninferiority. In *Testing Statistical Hypotheses of Equivalence and Noninferiority*.

<https://doi.org/10.1201/ebk1439808184>

Wilcoxon, F. (1945). *Some Uses of Statistics in Plant Pathology* Author ( s ): Frank Wilcoxon  
Published by : International Biometric Society. 1(4), 2–6.

WILCOXON, F. (1946). Individual comparisons of grouped data by ranking methods. *Journal of Economic Entomology*, 39(6), 269. <https://doi.org/10.1093/jee/39.2.269>

## ANEXOS

### ANEXO A.- *Muestras contaminadas*

Se genera muestras normales aleatorias contaminadas mediante el método de fleishman, para lo cual se genera el siguiente código:

```
rnonorm <- function(n, f.coef){  
  z <- rnorm(n)  
  f.coef[1] + (f.coef[2] * z) + (f.coef[3] * z^2) + (f.coef[4] * z^3)  
}
```

```
fleishman.coef(n.NN = 1, skewness.vec, kurtosis.vec)
```

Donde skewness.vec especifica la asimetría y kurtosis.vec define la Curtosis

## ANEXO B.- Probabilidad de cometer un error de tipo I (TIEP)

La probabilidad de cometer un error tipo 1, se realizó mediante una simulación de Monte Carlo donde se analiza cual es el porcentaje de veces que se equivoca el test, mediante el siguiente código.

```
tiepT <- function(nSim, n1, n2, f.coef, diffMean = 0, alpha = 0.05){  
  n <- n1 + n2  
  facts <- factor(rep(1:2, c(n1, n2)))  
  mu <- rep(c(0, diffMean), c(n1, n2))  
  pvals <- replicate(nSim, {  
    muestra <- rnonorm(n, f.coef)  
    t.test(muestra ~ facts, var.equal = T, mu = 0, conf.level = 1 -  
alpha)$p.value  
  })  
  sum(pvals < alpha) / nSim  
}
```

## ANEXO C.- Épsilon o Irrelevancia

Para determinar la irrelevancia se elaboró el siguiente código fuente en R.

```
opt_eps <- function(nSim, n1, n2, diffMean = 0, alpha = 0.05, delta =
0.2, salto.skew, salto.kurt, epsilon = 0.35){

  # criterio de cochran
  cochran.max <- alpha + delta*alpha
  cochran.min <- alpha - delta*alpha

  z <- qnorm(alpha/2, lower.tail = F)
  skewness.vec <- 0
  kurtosis.vec <- 0

  coeini <- try(round(as.vector(fleishman.coef(n.NN = 1, skewness.vec,
kurtosis.vec)),4), silent = T)
  coeini
  ltiep <- tiepT(nSim, n1, n2, diffMean = diffMean, f.coef = coeffle)
  ltiep
  leei <- z * sqrt((ltiep * (1 - ltiep)) / nSim)
  leei
  ic_tiep.min <- ltiep - leei
  ic_tiep.max <- ltiep + leei

  dftiep <-c(1:9) %>% setNames(c('tiep.min', 'tiep.max',
'tiep', 'skewness', 'kurtosis'))

  while (cochran.min <= ic_tiep.min && cochran.max >= ic_tiep.max ) {

    coeffle <- try(round(as.vector(fleishman.coef(n.NN = 1,
skewness.vec, kurtosis.vec)),4), silent = T)
    coeffle

    if(is.vector(coeffle) == T){
      set.seed(1234)
```

```

    ltiep <- tiepT(nSim, n1, n2, diffMean = diffMean, f.coef =
coeffle)
    ltiep
    lee <- z * sqrt((ltiep * (1 - ltiep)) / nSim)
    lee
    skewness.vec <- skewness.vec + salto.skew

    dftiep <- rbind(dftiep,
round(c(ic_tiep.min,ic_tiep.max,ltiep,skewness.vec,kurtosis.vec,
coeffle),4))

  }else{
    kurtosis.vec <- kurtosis.vec +salto.kurt
  }
  dftiep
  ic_tiep.min <- ltiep - lee
  ic_tiep.max <- ltiep + lee
}
epsopt <- as.array(dftiep[-1,])
print(epsopt) # Imprime Las Iteraciones realizadas dentro de ciclo
while
  coef_fle <- epsopt[nrow(epsopt),6:9] # Extrae Los coeficientes de
fleshman
  set.seed(1234)
  data <- rnorm(n1+n2)
  acont <- rnonorm(n1+n2, f.coef = coef_fle)
  atest <- normequiv(x = acont, epsilon, alpha)
  opt_eps <- sqrt(atest[2]) # Este serÃ-a el resultado de La funci3n.
  cat(' \n \n Test de falta de ajuste para mi data oficial \n \n')
  print(normequiv(x = data, epsilon = opt_eps))
  cat(' \n \n')
  data.frame('muestra'= paste0('n1 = ', n1, ' ', 'n2 = ', n2),
'epsilon'=opt_eps, 'Desicion'=normequiv(x = data, epsilon =
opt_eps)$upper.bound)
}

```



## ANEXO D.- Falta de ajuste

La función `normequiv` fue elaborado con el fin de analizar si una muestra de datos proviene de una distribución normal.

```
normequiv <- function(x, alpha = 0.05, epsilon = 0.15, ...){
  n <- length(x)
  h <- hist(x, plot = F)
  breaks <- h$breaks
  nbreaks <- length(breaks)
  theoretical <- NULL
  for(i in 1:(nbreaks-1)){
    theoretical[i] <- pnorm(breaks[i + 1], ...) - pnorm(breaks[i],
    ...)
  }
  theoretical <- c(pnorm(breaks[1], ...), theoretical,
    pnorm(breaks[nbreaks], lower.tail = F, ...))
  afr <- c(0, h$counts, 0)
  k <- length(afr)
  de <- afr / n # distribución empírica
  deuc <- sum((de - theoretical)^2) # distancia euclídea cuadrada

  # error estandar
  u <- qnorm(alpha, lower.tail = F)
  fac1 <- sum(((de - theoretical)^2) * de)
  fac2 <- 0
  for(j1 in 1:k){
    for(j2 in 1:k){
      fac2 <- fac2 + (de[j1] - theoretical[j1]) * (de[j2] -
theoretical[j2]) * de[j1] * de[j2]
    }
  }
  vn2 <- 4 * (fac1 - fac2)
  ee <- sqrt(vn2 / n)

  # Límite superior I.C
```

```

lupp <- deuc + (u * ee)

res <- c("Euclidean Distance" = deuc, "upper bound" = lupp, "eps^2"
= epsilon^2)
dec <- ifelse(lupp < epsilon^2, "Reject H0 - Normality", "Not Reject
H0 - Non Normality")
data.frame(t(res), "Decision" = dec)
#attr(,"Decision") <- dec

}

```

donde  $x$  es el conjunto de datos,  $\alpha$  es el nivel de significancia,  $\epsilon$  es el valor óptimo calculado con la función `opt_eps`.

## ANEXO E.- Comparación de test

Se elaboró la función normalTest la cual se pretende realizar comparaciones del test de equivalencia, Chi-cuadrado.

```
normalTest <- function(x, alpha = 0.05, epsilon, ...){
  n <- length(x)
  h <- hist(x, plot = F)
  breaks <- h$breaks
  nbreaks <- length(breaks)
  theoretical <- NULL
  for(i in 1:(nbreaks-1)){
    theoretical[i] <- pnorm(breaks[i + 1], ...) - pnorm(breaks[i],
    ...)
  }
  theoretical <- c(pnorm(breaks[1], ...), theoretical,
    pnorm(breaks[nbreaks], lower.tail = F, ...))
  afr <- c(0, h$counts, 0)
  k <- length(afr)
  # Igualdad (chi cuadrado)
  chi <- sum((afr - (theoretical*n))^2 / (theoretical*n))
  pvalchi <- pchisq(chi, k - 1, lower.tail = F)
  # Equivalencia (Test de Wellek)
  de <- afr / n # distribución empírica
  deuc <- sum((de - theoretical)^2) # distancia euclídea cuadrada
  # error estandar
  u <- qnorm(alpha, lower.tail = F)
  fac1 <- sum(((de - theoretical)^2) * de)
  fac2 <- 0
  for(j1 in 1:k){
    for(j2 in 1:k){
      fac2 <- fac2 + (de[j1] - theoretical[j1]) * (de[j2] -
theoretical[j2]) * de[j1] * de[j2]
    }
  }
  vn2 <- 4 * (fac1 - fac2)
```

```

ee <- sqrt(vn2 / n)
# LÃ-mite superior I.C
lupp <- deuc + (u * ee)
# Normality ???
c('X2Test' = ifelse(pvalchi >= alpha, T, F),
  'equivTest' = ifelse(lupp < epsilon^2, T, F))
}

```

donde x es la muestra, alpha es el nivel de significancia y epsilon es el valor optimo calculado con la funci3n opt\_eps.

La function rejectHO elabora pruebas comparativas de los test de equivalencia, Chi-cuadrado, test, Wilcoxon y t student.

```

rejectH0 <- function(nSim, n1, n2, epsilon, diffMean = 0,
  f.coef = c(0, 1, 0, 0), alpha = 0.05){
  n <- n1 + n2
  facts <- factor(rep(1:2, c(n1, n2)))
  mu <- rep(c(0, diffMean), c(n1, n2))
  tests <- replicate(nSim, {
    data <- rnonorm(n, f.coef = f.coef) + mu
    pretest <- normalTest(data, epsilon = epsilon)
    tstud <- t.test(data ~ facts, var.equal = T, conf.level = 1 -
alpha)$p.value
    wilc <- wilcox.test(data ~ facts, conf.level = 1 - alpha)$p.value
    c('t_direct' = tstud, 'wilcx_direct' = wilc,
      ifelse(pretest == T, tstud, wilc))
  })
  rowSums(tests < alpha, na.rm = T) / nSim
}

```

donde nSim es el n3mero de simulaciones, n1 es la muestra 1, n2 es la muestra 2, epsilon es el valor optimo, diffMean = 0, f.coef son los coeficientes de fleishman y alpha es el nivel de significancia.

## ANEXO F.- Comparación de TIEP y POTENCIA

Para ello se elabora la siguiente función.

```
TIEP_POT <- function(nSim, muestra, epsilon.vec, level.cont, level =
0, diffMean = 0, Potencia = 3, alpha = 0.05){

  library(BinNonNor)
  library(tidyverse)

  potiep <- data.frame()

  for(i in 1:nrow(muestra)){

    fleshc <- fleishman.coef(n.NN = 1, level.cont$simetria[level+1],
level.cont$Curtosis[level+1])

    tiepot <- c(rejectH0(nSim, muestra$n1[i], muestra$n2[i], epsilon
= epsilon$upper.bound[i], diffMean,
                    f.coef = fleshc, alpha),
              rejectH0(nSim, muestra$n1[i], muestra$n2[i], epsilon
= epsilon$upper.bound[i], diffMean = Potencia,
                    f.coef = fleshc, alpha))
    potiep <- rbind(potiep, data.frame('Gra_Cont' =
level.cont$Gr.Cont[level+1], 'muestra' = paste0('n1 = ',
muestra$n1[i], ' ', 'n2 = ', muestra$n2)[i], t(tiepot)))
  }

  colnames(potiep) <- c('Nivel.Contaminacion', 'Tamaño.muestra',
'TIEP.t', 'TIEP.Wilcoxon',
'TIEP.Pretest.JiCUadrado', 'TIEP.Pretest.Equiv', 'POT.t',
'POT.Wilcoxon', 'POT.Pretest.JiCUadrado', 'POT.Pretest.Equiv')
  potiep
}
```

## ANEXO G.- Resultados de la simulación

La siguiente tabla muestra los resultados obtenidos al realizar la simulación, donde se analiza la TIEP y Potencia de cada uno de los pretest y pruebas de comparación de medias.

<i>Nivel de Contaminación n</i>	<i>Tamaño de muestra</i>	<i>TIEP.t</i>	<i>TIEP Wilcoxon</i>	<i>TIEP Pretest Chi Cuadrado</i>	<i>TIEP Pretest Equivalencia</i>	<i>POT.t</i>	<i>Potencia Wilcoxon</i>	<i>Potencia Chi Cuadrado</i>	<i>Potencia Pretest Equivalencia</i>
0	n1 = 5 n2 = 5	0.0512	0.0321	0.0493	0.0512	0.9849	0.9504	0.9504	0.9674
0	n1 = 3 n2 = 7	0.0464	0.0313	0.0453	0.0454	0.9659	0.9133	0.9133	0.9136
0	n1 = 7 n2 = 3	0.0473	0.0323	0.0454	0.0469	0.9649	0.9057	0.9083	0.9606
0	n1 = 10 n2 = 10	0.0493	0.0432	0.0491	0.0493	1	0.9999	0.9999	0.9999
0	n1 = 6 n2 = 14	0.0518	0.0404	0.0507	0.0518	0.9999	0.9998	0.9998	0.9997
0	n1 = 14 n2 = 6	0.0512	0.0436	0.0505	0.0512	1	0.9995	0.9995	1
0	n1 = 5 n2 = 10	0.0501	0.0416	0.0492	0.0501	0.999	0.9968	0.9968	0.9969
0	n1 = 10 n2 = 5	0.0508	0.0412	0.0496	0.0495	0.9995	0.9967	0.9968	0.9978
0	n1 = 10 n2 = 20	0.0502	0.0482	0.0497	0.0503	1	1	1	1
0	n1 = 20 n2 = 30	0.0483	0.0475	0.0483	0.0483	1	1	1	1
1	n1 = 5 n2 = 5	0.0496	0.0332	0.0485	0.0495	0.977	0.9392	0.9392	0.9558
1	n1 = 3 n2 = 7	0.0468	0.032	0.046	0.0448	0.9591	0.9054	0.9054	0.9056
1	n1 = 7 n2 = 3	0.0471	0.0324	0.045	0.0471	0.9614	0.901	0.9022	0.9574
1	n1 = 10 n2 = 10	0.0485	0.0442	0.0487	0.0485	0.9999	1	1	0.9999
1	n1 = 6 n2 = 14	0.0479	0.0423	0.0475	0.0479	1	0.9991	0.9991	0.9992
1	n1 = 14 n2 = 6	0.0481	0.0403	0.0465	0.0481	0.9997	0.9996	0.9996	0.9997
1	n1 = 5 n2 = 10	0.0516	0.0404	0.0505	0.0515	0.9979	0.9892	0.9892	0.9895
1	n1 = 10 n2 = 5	0.0502	0.0405	0.0492	0.0498	0.9982	0.9957	0.9957	0.9972
1	n1 = 10 n2 = 20	0.0523	0.052	0.0527	0.0523	1	0.9999	0.9999	0.9999
1	n1 = 20 n2 = 30	0.0506	0.0472	0.0491	0.0506	1	1	1	1

2	n1 = 5 n2 = 5	0.0472	0.0288	0.0445	0.0467	0.9786	0.9292	0.9292	0.9421
2	n1 = 3 n2 = 7	0.0484	0.0333	0.0463	0.0468	0.9592	0.9014	0.9014	0.9015
2	n1 = 7 n2 = 3	0.0458	0.0312	0.0452	0.0458	0.9551	0.8885	0.8897	0.9473
2	n1 = 10 n2 = 10	0.0501	0.0455	0.0498	0.0502	1	0.9998	0.9998	0.9999
2	n1 = 6 n2 = 14	0.0496	0.0405	0.048	0.0496	0.9995	0.9981	0.9981	0.9982
2	n1 = 14 n2 = 6	0.0498	0.0422	0.0481	0.0498	0.9997	0.9999	0.9998	0.9997
2	n1 = 5 n2 = 10	0.0469	0.0391	0.0459	0.0469	0.997	0.9841	0.9841	0.9841
2	n1 = 10 n2 = 5	0.0496	0.0405	0.0488	0.0495	0.9977	0.9971	0.9971	0.9969
2	n1 = 10 n2 = 20	0.048	0.0468	0.0475	0.0481	1	1	1	1
2	n1 = 20 n2 = 30	0.048	0.0478	0.0485	0.048	1	1	1	1
3	n1 = 5 n2 = 5	0.0442	0.0303	0.0416	0.0433	0.9683	0.8959	0.8959	0.9134
3	n1 = 3 n2 = 7	0.0488	0.0345	0.0467	0.0463	0.95	0.8871	0.8871	0.8871
3	n1 = 7 n2 = 3	0.0449	0.0322	0.0433	0.0445	0.9475	0.8596	0.8599	0.9329
3	n1 = 10 n2 = 10	0.0476	0.0438	0.0472	0.0477	0.9997	0.9995	0.9995	0.9995
3	n1 = 6 n2 = 14	0.0467	0.0437	0.0461	0.0469	0.9995	0.9933	0.9933	0.9934
3	n1 = 14 n2 = 6	0.0449	0.0384	0.0428	0.0449	0.9995	0.9997	0.9997	0.9995
3	n1 = 5 n2 = 10	0.0468	0.0398	0.0464	0.0471	0.9941	0.9705	0.9705	0.9705
3	n1 = 10 n2 = 5	0.0473	0.0439	0.0459	0.0455	0.9965	0.9931	0.9931	0.9932
3	n1 = 10 n2 = 20	0.0486	0.0479	0.0501	0.0484	1	1	1	1
3	n1 = 20 n2 = 30	0.0457	0.0459	0.0449	0.0458	1	1	1	1
4	n1 = 5 n2 = 5	0.0342	0.0275	0.0329	0.0339	0.9546	0.8818	0.8818	0.8841
4	n1 = 3 n2 = 7	0.0419	0.0324	0.0364	0.0366	0.9384	0.8932	0.893	0.8932
4	n1 = 7 n2 = 3	0.0424	0.0331	0.0382	0.0411	0.9294	0.8347	0.8347	0.9054
4	n1 = 10 n2 = 10	0.0445	0.0466	0.0475	0.0454	0.9987	0.9987	0.9986	0.9986
4	n1 = 6 n2 = 14	0.0445	0.0383	0.0394	0.0443	0.9954	0.9891	0.9881	0.9891
4	n1 = 14 n2 = 6	0.0481	0.0456	0.0466	0.0477	0.9975	0.9992	0.999	0.9974
4	n1 = 5 n2 = 10	0.041	0.0378	0.0405	0.0401	0.9865	0.9512	0.9509	0.9512
4	n1 = 10 n2 = 5	0.0435	0.039	0.0414	0.0405	0.9866	0.9865	0.9861	0.9865

4	n1 = 10 n2 = 20	0.0444	0.05	0.0496	0.0467	0.9999	0.9998	0.9989	0.9998
4	n1 = 20 n2 = 30	0.0527	0.0533	0.0533	0.0516	1	1	0.9983	1
5	n1 = 5 n2 = 5	0.0352	0.034	0.039	0.0371	0.9438	0.8731	0.8726	0.8748
5	n1 = 3 n2 = 7	0.0387	0.0329	0.0351	0.0359	0.932	0.8952	0.8932	0.8951
5	n1 = 7 n2 = 3	0.0427	0.0336	0.0375	0.0409	0.9213	0.8243	0.8239	0.9015
5	n1 = 10 n2 = 10	0.0386	0.0424	0.0434	0.0395	0.9948	0.9975	0.9924	0.9974
5	n1 = 6 n2 = 14	0.0398	0.0392	0.0396	0.0406	0.9915	0.9878	0.9833	0.9878
5	n1 = 14 n2 = 6	0.0406	0.0377	0.038	0.0409	0.9922	0.9988	0.9948	0.993
5	n1 = 5 n2 = 10	0.042	0.0441	0.0453	0.044	0.98	0.9335	0.9313	0.9332
5	n1 = 10 n2 = 5	0.0409	0.0409	0.0416	0.0407	0.9801	0.9847	0.9827	0.9845
5	n1 = 10 n2 = 20	0.0425	0.0477	0.0474	0.0458	0.9988	0.9998	0.9921	0.9997
5	n1 = 20 n2 = 30	0.044	0.052	0.0519	0.0476	1	1	0.9896	1
6	n1 = 5 n2 = 5	0.0246	0.0321	0.0338	0.032	0.9388	0.8725	0.8693	0.8749
6	n1 = 3 n2 = 7	0.0367	0.0341	0.034	0.0343	0.9282	0.9071	0.9036	0.9067
6	n1 = 7 n2 = 3	0.0396	0.0343	0.0369	0.0369	0.922	0.8351	0.8342	0.9047
6	n1 = 10 n2 = 10	0.034	0.0424	0.0425	0.0394	0.9893	0.9976	0.9859	0.997
6	n1 = 6 n2 = 14	0.0391	0.0387	0.0387	0.0377	0.986	0.9882	0.9753	0.9881
6	n1 = 14 n2 = 6	0.0388	0.0395	0.0399	0.0396	0.9871	0.9978	0.9892	0.9883
6	n1 = 5 n2 = 10	0.0366	0.04	0.0413	0.0405	0.9724	0.9181	0.913	0.9166
6	n1 = 10 n2 = 5	0.036	0.042	0.0421	0.0415	0.9698	0.9799	0.9753	0.9787
6	n1 = 10 n2 = 20	0.0437	0.0486	0.0483	0.0464	0.9977	0.9996	0.9839	0.9994
6	n1 = 20 n2 = 30	0.0429	0.0491	0.0486	0.0483	1	1	0.9759	1