



**ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO**

**FACULTAD DE CIENCIAS**

**ESCUELA DE FÍSICA Y MATEMÁTICA**

**ALTERACIÓN DE LA PROBABILIDAD GLOBAL DE ERROR TIPO I  
EN LAS PRUEBAS DE COMPARACIÓN DE DOS MEDIAS CUANDO  
SE VERIFICAN SUS SUPUESTOS.**

**TRABAJO DE TITULACIÓN**

**TIPO: PROYECTO DE INVESTIGACIÓN**

Presentado para optar al grado académico de:

**INGENIERA EN ESTADÍSTICA INFORMÁTICA**

**AUTORA: TANIA JHOANNA SÁNCHEZ ACALO**

**TUTOR: ING. PABLO JAVIER FLORES MUÑOZ**

Riobamba-Ecuador

2019

**©2019, Tania Jhoanna Sánchez Acalo**

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**  
**FACULTAD DE CIENCIAS**  
**ESCUELA DE FÍSICA Y MATEMÁTICA**

El Tribunal de Tesis certifica que: El trabajo de investigación: “**ALTERACIÓN DE LA PROBABILIDAD GLOBAL DE ERROR TIPO I EN LAS PRUEBAS DE COMPARACIÓN DE DOS MEDIAS CUANDO SE VERIFICAN SUS SUPUESTOS**”, de responsabilidad de la señorita Tania Jhoanna Sánchez Acalo, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Titulación, quedando autorizada su presentación.

**FIRMA**

**FECHA**

Ing. Pablo Javier Flores Muñoz MSc.

**DIRECTOR DEL TRABAJO DE  
TITULACION**

\_\_\_\_\_

\_\_\_\_\_

Ing. Jorge Luis Hernández Ambato Ph.D.

**MIEMBRO DEL TRIBUNAL**

\_\_\_\_\_

\_\_\_\_\_

Yo, Tania Jhoanna Sánchez Acalo, declaro que el trabajo aquí descrito es de mi autoría, que los resultados del mismo son auténticos y originales.; y no han sido previamente presentados para ningún grado o calificación profesional; los textos constantes en el documento que provienen de otra fuente están debidamente citados y referenciados. Como autora asumo la responsabilidad legal y académica de los contenidos de este trabajo de titulación.

---

**Tania Jhoanna Sanchez Acalo**

**060487959-3**

## DEDICATORIA

*A mi padre Celestial y Jesucristo:* Por haberme permitido llegar hasta este momento tan importante de mi formación profesional, además por su infinito amor, bondad y fe que me brinda día con día para alcanzar mis propósitos terrenales y espirituales.

*A mi familia Sánchez Acalo:* Especialmente a mis padres Segundo y Aurora, a mis hermanos, Darwin, Henry, Jenny Gabriela, sobrinos Damian y Mayquel y a mis amigos miembros SUD por su apoyo incondicional, comprensión y amor que me han permitido ser una persona de bien.

*A mi compañero eterno Rafael:* Por haberme apoyado en todo momento, por la motivación, por siempre estar dispuesto a escucharme, pero más que nada, por su cariño y amor.

Tania Sánchez Acalo

## AGRADECIMIENTO

*Agradezco a:*

*Dios:* por ser mi guía en mi vida por darme dones y habilidades y más que todo por estar conmigo en todo momento.

*Mis padres Aurora y Segundo:* por su apoyo, motivación incondicional para alcanzar esta meta terrenal.

*A mis hermanos: Darwin, Henry, Jenny y Gabriela:* por motivarme y confiar en mis capacidades.

*A Rafael:* por su apoyo, amor y la motivación en todo momento.

*Ing. Pablo Flores Muñoz:* por brindarme su apoyo y orientación con sus conocimientos, convirtiéndose en un pilar para la culminación de este trabajo. Mi respectiva admiración por ser un excelente docente.

*Ing. Jorge Hernández:* por brindarme su apoyo y orientación con sus conocimientos al desarrollo de este trabajo.

A la Escuela Superior Politécnica de Chimborazo por brindarme la oportunidad para estudiar y convertirme en una profesional, al Dr Luis Vera director de la Escuela de Física y Matemática y todos los maestros que conforman la carrera de Ingeniería en Estadística Informática.

*Con cariño: Tania Sánchez Acalo*

## TABLA DE CONTENIDOS

RESUMEN.....	xii
ABSTRACT.....	xiii
INTRODUCCIÓN .....	1
CAPITULO I.....	4
1 <b>MARCO TEÓRICO REFERENCIAL.....</b>	<b>4</b>
1.1 <b>Antecedentes .....</b>	<b>4</b>
1.2 <b>Planteamiento del problema.....</b>	<b>7</b>
1.2.1 <i>Formulación del problema .....</i>	<i>8</i>
1.2.2 <i>Sistematización del problema .....</i>	<i>8</i>
1.3 <b>Justificación .....</b>	<b>9</b>
1.3.1 <i>Justificación Teórica.....</i>	<i>9</i>
1.3.2 <i>Justificación practica .....</i>	<i>10</i>
1.4 <b>Objetivos .....</b>	<b>11</b>
1.4.1 <i>Objetivo General .....</i>	<i>11</i>
1.4.2 <i>Objetivos Específicos.....</i>	<i>11</i>
CAPITULO II .....	12
2 <b>MARCO TEORICO.....</b>	<b>12</b>
2.1 <b>Pruebas de Hipótesis .....</b>	<b>12</b>
2.2 <b>Prueba de hipótesis para comparar dos medias poblaciones .....</b>	<b>13</b>
2.2.1 <i>Prueba T-Student .....</i>	<i>13</i>
2.2.2 <i>Prueba de Welch .....</i>	<i>14</i>
2.2.3 <i>Prueba de U de Mann Withney .....</i>	<i>15</i>
2.3 <b>Prueba de hipótesis para verificar los supuestos de normalidad .....</b>	<b>16</b>
2.3.1 <i>Test de Shapiro Wilk (SW).....</i>	<i>17</i>
2.4 <b>Test de hipótesis para verificar el supuesto de homocedasticidad.....</b>	<b>18</b>

2.4.1	<i>Enfoque clásico: Prueba Fisher-Snedecor (F)</i> .....	18
2.4.2	<i>Enfoque de equivalencia: prueba de dispersión de dos distribuciones Gaussiana</i> .....	20
2.5	Método de Variables de Control para reducir varianzas .....	22
2.6	Método de Fleishman .....	24
<b>CAPITULO III</b> .....		25
3	<b>MARCO METODOLOGICO</b> .....	25
3.1	Obtención del test más potente para probar normalidad.....	25
3.2	Proceso de simulación para la estimación del TIEP Global en pruebas de comparación de dos medias.....	27
3.3	Funciones previas usadas en TIEP.norm.....	30
<b>CAPITULO IV</b> .....		32
4	<b>MARCO DE RESULTADOS, DISCUSIÓN Y ANÁLISIS DE RESULTADOS</b> .....	32
4.1	Resultados Preliminares .....	32
4.1.1	<i>Potencia de las pruebas de normalidad</i> .....	32
4.2	Resultados Finales.....	33
4.3	Discusión .....	40
<b>CONCLUSIONES</b> .....		42
<b>RECOMENDACIONES</b> .....		43
<b>BIBLIOGRAFÍA</b>		
<b>ANEXOS</b>		

## ÍNDICE DE TABLAS

<b>Tabla 1-2:</b> Estado de la Naturaleza cuando se rechaza o no rechaza $H_0$ .....	12
<b>Tabla 2-2:</b> Hipótesis alternativas en l prueba t-Student .....	13
<b>Tabla 3-2:</b> Región de rechazo en la prueba t-Student .....	14
<b>Tabla 4-2 :</b> Hipótesis alternativas de la prueba U de mann withney .....	15
<b>Tabla 5-2 :</b> Región de rechazo de la prueba U de Mann Withney .....	16
<b>Tabla 6-2:</b> Tabla de contingencia 2*2 de las frecuencias relativas de .....	22
<b>Tabla 7-2:</b> Tabla de contingencia 2*2 de las probabilidades de los .....	22
<b>Tabla 8-3:</b> Coeficientes de Fleishman.....	24
<b>Tabla 9-3 :</b> Funciones utilizadas para las pruebas de normalidad.....	26
<b>Tabla 10-3:</b> Limites de irrelevancia .....	29

## ÍNDICE DE FIGURAS

<b>Figura 1-3: Procedimientos para la estimación del TIEP Global</b> .....	28
<b>Figura 2-4: Estimación de la potencia en varios de niveles de contaminación normal</b> .....	32
<b>Figura 3-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales <math>n_1 = 5, n_2 = 5</math> con una significancia <math>\alpha = 0.05</math>.</b> .....	33
<b>Figura 4-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales <math>n_1 = 10, n_2 = 10</math> con una significancia <math>\alpha = 0.05</math>.</b> .....	35
<b>Figura 5-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales <math>n_1 = 5, n_2 = 10</math> con una significancia <math>\alpha = 0.05</math>.</b> .....	36
<b>Figura 6-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales <math>n_1 = 10, n_2 = 5</math> con una significancia <math>\alpha = 0.05</math>.</b> .....	38

## ÍNDICE DE ANEXOS

<b>Anexo A :</b> Tablas de resultados de la Estimación de la Robustez de las Pruebas de Normalidad .....	7
<b>Anexo B:</b> Tablas de resultados de Estimación de la Potencia de las Pruebas de Normalidad.....	7
<b>Anexo C:</b> Tabla de resultados de la estimación de la Probabilidad de Error de Tipo I en las Pruebas de Comparación de Medias .....	9
<b>Anexo D:</b> Función power.norm .....	17
<b>Anexo E:</b> Funciones auxiliares para TIEP.norm .....	18
<b>Anexo F:</b> TIEP.norm .....	20

## RESUMEN

El presente estudio tuvo por objetivo estimar la probabilidad global de cometer un Error Tipo I de las pruebas para contrastar igualdad de dos medias (t-Student, Welch y Wilcoxon), cuando se pre-testean sus supuestos de normalidad y homocedasticidad. Investigaciones previas muestran que el proceso de pre-testeo tradicional produce graves alteraciones en la Probabilidad de Error Tipo I, y se recomienda eliminar este proceso de la enseñanza tradicional de estadística. Estudios posteriores muestran que el problema no es pre-testear los supuestos, sino la forma como estos son planteados. Un enfoque de equivalencia para pre-testear, podría eliminar los problemas descritos. En este trabajo se desarrollaron algoritmos de simulación estocástica que permitieron estimar la Probabilidad de Error Tipo I cuando se pre-testean los dos supuestos en conjunto, para muestras con diferentes contaminaciones o alejamientos de la normalidad y homocedasticidad. La generación de muestras no normales se lo realizó a partir del sistema de contaminación de Fleishman, el cual permitió generar muestras de distribución desconocida y con un alejamiento de la normalidad medible. El nivel de heterocedasticidad se establece a partir de la razón entre las varianzas poblacionales que generan las muestras. Los resultados hallados muestran que pre-testear bajo el enfoque tradicional altera significativamente la probabilidad global de Error Tipo I, sobre todo para muestras pequeñas y desbalanceadas y con una contaminación severa de la normalidad, por lo que su uso se desaconseja. Pre-testear bajo el enfoque de equivalencia, parece controlar de mejor manera esta probabilidad de error, por lo cual este enfoque se considera más confiable que el tradicional, incluso llama la atención que, para muestras con distribuciones desconocidas, esta técnica es más robusta que la prueba de Wilcoxon, la cual supone está diseñada exclusivamente para este tipo de muestras. Finalmente, algunas posibles investigaciones futuras sobre el tema se discuten.

**PALABRAS CLAVE:** ESTADISTICA, SIMULACION, PRE TESTEO, SUPUESTOS, PRUEBA DE EQUIVALENCIA, PRUEBA ESTADISTICA (FISHER), PRUEBA ESTADISTICA (T-STUDENT), PRUEBA ESTADISTICA (WELCH), PRUEBA ESTADISTICA (WILCOXON)

## ABSTRACT

The objective of this study was to estimate the global probability of committing a Type I error of the tests to compare equality of two measures (t-Student, Welch and Wilcoxon), when their assumptions of normality and homoscedasticity are pre-tested. Previous research shows that the traditional pre-testing process produces serious alterations in the Type I Error Probability, and it is recommended to eliminate this process from the traditional teaching of statistics. Later studies show that the problem is not to pre-test the assumptions, but the way they are posed. An equivalency approach to pre-test could eliminate the problems described. In this work, stochastic simulation algorithms were developed to estimate the Type I Error Probability when the two assumptions are pre-tested together, for samples with different contaminations or departures from normality and homoscedasticity. The generation of non-normal samples was made from the Fleishman contamination system, which allowed the generation of samples of unknown distribution and with a distance of measurable normality. The level of heteroscedasticity is established from the ratio between the population variances that generate the samples. The results show that pre-testing under the traditional approach significantly alters the overall probability in Error Type I, especially for small samples and Pre-testing under the equivalence approach, seems to better control this probability of error, that is why this approach is considered more reliable than the traditional one, and it is even striking that, for samples with unknown distributions, this technique is more robust than the Wilcoxon test, which supposedly is designed exclusively for this type of samples. Finally, some possible future research on the subject is discussed.

**KEY WORDS:** STATISTICS, SIMULATION, PERTEST, ASSUMPTIONS, EQUIVALENCE TEST, STATISTICAL TEST (FISHER), STATISTICAL TEST (T-STUDENT), STATISTICAL TEST (WELCH), STATISTICAL TEST (WILCOXON).

## INTRODUCCIÓN

Teóricamente, las pruebas para contrastar diferencias significativas entre dos medias están sujetas al cumplimiento del supuesto de normalidad y homocedasticidad. La mayoría de libros y software estadístico enseñan que para validar estos supuestos hay que utilizar otras pruebas de hipótesis (que llamaremos pre-test) como Shapiro Wilk, Anderson Darling, Kolmogorov Smirnov, etc.... en el caso de la normalidad o F, Bartlet, Cochran, Levene, etc en el caso de la homocedasticidad. Se suele asumir que cuando la hipótesis nula de normalidad es rechazada, una prueba no paramétrica como la de Wilcoxon (1946) es la más apropiada, por otra parte, si esta hipótesis no se rechaza, se suele asumir normalidad y a continuación el supuesto de homocedasticidad es pre-testeado. Cuando la hipótesis nula de igualdad de varianzas no se rechaza, se suele asumir el cumplimiento de este supuesto y la prueba t-Student supone ser el test más adecuado, mientras que cuando esta hipótesis es rechazada, se supone heterocedasticidad y una modificación de la t-Student conocida como el test de Welch (1947) es el procedimiento que se asume como más adecuado.

Este proceso, llamado pre-testeo, aunque es muy común, no es el más conveniente. Investigaciones realizadas (Hsu, 1938; Overall, Atlas y Gibson, 1995) demuestran que hacerlo conlleva a graves alteraciones en la probabilidad de cometer un Error Tipo I (TIEP, por sus siglas en inglés) (Flores et al., 2018, p. 2), por lo que se aconseja rotundamente eliminar este procedimiento y en su lugar usar directamente el test de Welch (sin pre-testear) como el único método para comparar medias, puesto que se ha comprobado que este procedimiento controla de manera adecuada la TIEP alrededor del nivel de significancia  $\alpha$  planteado en la prueba de diferencia de medias .

Wellek (2010, p.1-17) en su libro *“Testing Statistical Hypotheses of Equivalence and Noninferiority”*, sugiere que, el problema no radica en pre-testear los supuestos, sino en la forma en como han sido planteados en el enfoque tradicional, donde la hipótesis nula indica el cumplimiento de normalidad u homocedasticidad perfecta. El autor propone un nuevo enfoque basado en equivalencia. Llamaremos equivalencia a una forma dilatada de una igualdad, la cual queda definida por una zona que se considera irrelevante como para considerar la existencia de diferencias entre los parámetros comparados. Este enfoque plantea pruebas de hipótesis en que, la hipótesis nula establece la no equivalencia, mientras que la alterativa establece la equivalencia; en esta nueva forma de planteamiento se debe introducir una región que se encuentra limitada por dos valores

constantes positivos llamados límites de equivalencia, estos se ubican alrededor del parámetro y consideran las desviaciones irrelevantes bajo las cuales se establece la equivalencia.

Una deficiencia de las pruebas planteadas por Wellek (2010, pp. 29-31) es que, a pesar de la importancia que representa, usa un criterio subjetivo (basado en lo que él denomina "sentido estadístico común" o en "la experiencia del investigador), para determinar los límites de irrelevancia (Flores y Ocaña, 2018) desarrollan un método, que resulta menos subjetivo, se trata de una expresión algorítmica que en función del tamaño muestral y el nivel de significancia determina los límites de irrelevancia adecuados, que se deben usar en una prueba de equivalencia usada como pre-test para probar el supuesto de homocedasticidad. Estos límites son calculados iterativamente mediante un algoritmo numérico, como la razón entre varianzas por encima y por debajo de la perfecta homocedasticidad para las cuales se produce una TIEP dentro un límite de permisividad (en una prueba t-student de comparación de medias). Para más detalles sobre la construcción de este algoritmo, el lector puede remitirse <https://www.idescat.cat/sort/artpublished.html>.

Los estudios presentados hasta ahora (Mara et al., 2014; Kim y Cribbie, 2017; Flores Muñoz, 2017; Flores y Ocaña, 2018) sugieren que, cualquier investigador que garantice normalidad perfecta puede utilizar el Test de equivalencia de dispersión como pre-test para probar homocedasticidad, previo a una prueba de comparación de medias, sin temor a alterar la TIEP global. Aunque este resultado es bastante interesante, nada se ha investigado hasta el momento respecto a qué ocurre cuando no se puede asegurar la presencia de muestras normales y aún más, cuando este supuesto es pre-testeado. Es por esto que proponemos la presente investigación, con el fin de ampliar el estudio de tal manera que en un algoritmo realizado mediante un proceso de simulación estocástica se pueda obtener una estimación de la TIEP global de las pruebas de comparación de medias cuando se pre-testea los supuestos de normalidad y homocedasticidad a la vez. Los pre-test apropiados para probar cada una de los supuestos requeridos son; en el caso del supuesto de normalidad se utilizó la prueba Shapiro Wilks y en el caso de homocedasticidad la prueba F ya que su estadístico de prueba es el mismo que el usado en la prueba de equivalencia. Con el fin de experimentar muestras alejadas de una perfecta normalidad, se simularán muestras no normales, obtenidas por el método de Fleishman (1978), que podrían ser homocedásticas o con diferentes grados de heterocedasticidad definidas por la razón entre varianzas  $\frac{\sigma_1^2}{\sigma_2^2}$ . Finalmente, para reducir la variabilidad de la estimación de la TIEP, se usará la técnica denominada "Variables de Control", la cual permitirá mejorar la precisión de la simulación y para su validación el criterio de Cochran.

El primer capítulo empieza describiendo los antecedentes, planteamientos del problema, justificación y los objetivos. En el segundo capítulo se analiza los métodos y conceptos de las pruebas a utilizar. El tercer capítulo describe la metodología para construir el algoritmo que será utilizado para determinar la TIEP global en siete procedimientos planteados. En la cuarta sección, se analiza y discute los principales resultados hallados, y finalmente en la última sección se discuten las principales conclusiones y se identifican las áreas para futuras investigaciones.

## CAPITULO I

### 1 MARCO TEÓRICO REFERENCIAL

#### 1.1 Antecedentes

A pesar de que el proceso de pre-testeo previo a la elección de una prueba adecuada (Wilcoxon, t-Student o Welch) para contrastar igualdad de medias es muy común en la literatura estadística, estudios muestran que este procedimiento trae serios problemas relacionados con la probabilidad de cometer un Error de Tipo I, más aún cuando los tamaños muestrales son pequeños y desbalanceados. La reciente producción científica (Rasch et al., 2011; Flores Muñoz, 2017; Flores y Ocaña, 2018; Zimmerman, 2004) sugiere que pre-testear el supuesto de homocedasticidad mediante el pre-test (Levene, Bartlett, Cochran) conlleva a graves peligros desconocidos puesto que la Probabilidad global de Cometer un Error de Tipo I (TIEP) se ve severamente afectada, lo cual aumenta la probabilidad de tomar una decisión errónea; esto en el sentido de rechazar una igualdad de medias que teóricamente es verdadera. Por esta razón, los autores de estas investigaciones concluyen que este procedimiento de pre-testeo debe eliminarse por completo y en su lugar el uso directo del test de Welch (sin pre-test) parece ser un procedimiento adecuado, puesto que al hacerlo la TIEP queda muy bien controlada alrededor del nivel de significancia  $\alpha$  planteado para la prueba de comparación de medias. Refiriéndose al supuesto de normalidad algunos autores (Schucany y Tony Ng, 2006; Rochon y Kieser, 2011; Rasch et al., 2011) demostraron que pre-testear este supuesto mediante las pruebas (Shapiro Wilks, Kolmogorov Smirnov) usando muestras provenientes de distribuciones no normales no corrige la TIEP global.

Otro autor (Moder et al., 2009, pp. 258-264) demostró que cuando ambos supuestos (normalidad y homogeneidad) son pre-testeados mediante las pruebas de Kolmogorov Smirnov y Levene respectivamente, usando distintos tamaños de muestras, diferentes niveles de no normalidad y heterocedasticidad, conduce a graves riesgos desconocidos en las probabilidades de cometer un error de tipo I y tipo II, y que aplicar una t-Student o una prueba de Wilcoxon sin la verificación de los supuestos produce peores alteraciones a la TIEP, estos resultados demuestran nuevamente que Welch sigue siendo válido excepto para muestras pequeñas y con un grado severo de no normalidad.

Al respecto de los supuestos y refiriéndose específicamente al de normalidad Box, menciona 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 una aproximación útil a los que se encuentran en el mundo real” (Box, 1979; citados en Flores et al., 2018), entonces si no existe un conjunto de datos perfectamente normales, no tiene sentido usar modelos que sirven para contratar una hipótesis de normalidad perfecta, puesto que ya sabemos que esto no existe, en su lugar lo que nos debería interesar es conocer si los modelos usados para probar normalidad presentan una aproximación con la realidad, lo suficientemente buena como para ser utilizado. 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  $[\alpha \pm 0.20 \alpha]$ .

Al respecto de la homocedasticidad, el contraste tradicional viene planteado como  $H_0: \sigma_1^2/\sigma_2^2 = 1$ , no rechazar esta hipótesis, no significa que exista homogeneidad; simplemente significa no existe evidencia para inferir diferencias significativas entre las varianzas comparadas, lo cual no implica necesariamente homocedasticidad; mientras que rechazarla podría tan solo sugerir una heterocedasticidad irrelevante. En relación a lo anterior (Mara, et al., 2014; citados en Kim y Cribbie, 2017) afirma que no rechazar la hipótesis nula solo muestra que no hay pruebas suficientes para detectar las variaciones desiguales en la población, lo cual no implica necesariamente homocedasticidad o como en este mismo sentido “La ausencia de evidencia no es evidencia de ausencia” (Altman y Bland, 1995, p. 485).

En respuesta a la forma inadecuada de plantear la hipótesis (Wellek, 2010), propone un nuevo enfoque denominado equivalencia. Debido a que la noción de equivalencia ha formado parte importante de la terminología estándar en estadística aplicada y aunque el concepto preciso de equivalencia no es auto explicativo, el mismo autor en su libro populariza este término como la forma dilatada de una relación de identidad entre los parámetros analizados. Además, menciona que la dilatación en la hipótesis de equivalencia se induce por la ampliación de la hipótesis nula del problema (tradicional de prueba bilateral), mediante la adición de una zona de indiferencia alrededor de la región correspondiente o punto en el espacio de igualdad paramétrica, es decir, equivalencia significa igualdad excepto por desviaciones prácticamente irrelevantes (Wellek, 2010, p. 164). Las pruebas de equivalencia no realizan una comprobación de una perfecta igualdad de los parámetros a estimar, sino por el contrario plantean que  $H_0$  establece la “no equivalencia”, mientras que  $H_1$  establece “la equivalencia”. En esta nueva forma de plantear las hipótesis, se debe

añadir una región alrededor del parámetro para considerar las desviaciones irrelevantes bajo las cuales se establece la equivalencia, esta región se encuentra limitada por dos valores constantes llamados “*límites de equivalencia*”.(Flores y Ocaña, 2018, p. 59-72), además de proporcionar un método para calcular estos límites, demostraron que al usar esta prueba de equivalencia como un pre-test para probar homocedasticidad, la TIEP global de una prueba de comparación de medias no se ve afectada, al contrario, se observa que esta probabilidad queda muy bien controlada (de manera similar al procedimiento de Welch directo) alrededor del nivel de significancia  $\alpha$  planteado para la prueba de comparación de medias. Al contrario de las primeras investigaciones realizadas al respecto, estos autores concluyen que pre-testear homocedasticidad no conlleva a errores de tipo I, pero siempre y cuando el supuesto sea probado bajo un enfoque de equivalencia y se garantice normalidad.

A pesar de que los usos adecuados de las pruebas de equivalencia han demostrado resolver los inconvenientes relacionados al proceso de pre-testeo, creemos que este problema no ha sido resuelto en su totalidad, esto porque nada se sabe aún del valor que puede tomar la TIEP cuando el proceso de pre-testeo es completo, esto es cuando se contrastan los supuestos de homocedasticidad y normalidad a la vez. De hecho, la TIEP del estudio en referencia queda controlada solo cuando se puede asegurar normalidad, pero ¿Qué ocurre para poblaciones que presentan alejamientos de esta distribución? y aún más ¿Qué ocurre si se usa un pre-test para contrastar este supuesto?.

## 1.2 Planteamiento del problema

Ya se ha mencionado que el proceso de pre-testeo definirá la prueba adecuada para realizar la comparación de medias. También sabemos que cuando estos supuestos son pre-testeados mediante las pruebas tradicionales, existe un alto riesgo de cometer un error Tipo I, sin embargo, cuando este mismo proceso se lo realiza bajo el enfoque de equivalencia, el riesgo de cometer el error queda controlado, esto sin embargo funciona solo cuando se pre-testea homocedasticidad y se puede asegurar normalidad.

Queda claro que para conocer en su totalidad las consecuencias que implica el proceso de pre-testeo, debemos analizar lo que ocurre cuando no se puede asegurar la presencia de normalidad y además cuando todos estos conjuntos de muestras generadas son pre-testeados.

En este sentido el primer reto al que nos enfrentamos es la generación de muestras que podamos asegurar tengan distintos grados de contaminación o alejamientos de la normal. Una vez obtenidas estas muestras, existe la necesidad de hacerlas recorrer por todo el proceso de pre-testeo (el cual debe incluir la verificación de los dos supuestos en conjunto) para conocer lo que ocurre con la TIEP. El desconocimiento que hasta ahora tenemos de esta probabilidad, la cual se obtendrá mediante un proceso más completo que el estudiado hasta ahora, no ha permitido tener una medición de los posibles errores al que pudiera estar expuesto un investigador que al tener datos crudos se ve obligado a cumplir con todo el proceso de pre-testeo antes de decidir la prueba de comparación de medias adecuada para su estudio.

### ***1.2.1 Formulación del problema***

¿De qué forma se altera la TIEP global en las pruebas de comparación de medias cuando se realizan pre-test para verificar sus supuestos?

### ***1.2.2 Sistematización del problema***

¿De qué forma nos ayuda un algoritmo de simulación a obtener valores de la estimación de la TIEP global de las pruebas de comparación de medias cuando se verifican sus supuestos mediante los enfoques tradicionales y de equivalencia?

¿Qué ocurre al comparar los resultados de las estimaciones de la TIEP global producidas en los distintos procedimientos simulados?

¿Qué procedimientos estudiados son adecuados para realizar pruebas de comparación de medias?

## 1.3 Justificación

### 1.3.1 *Justificación Teórica*

Queda claro (al menos bajo el enfoque tradicional) que pre-testear los supuestos previo a una prueba de comparación de medias conlleva a una alteración en la probabilidad global de cometer un error de tipo I (TIEP). Actualmente muchos estadísticos teóricos (Hsu, 1938; Easterling y Anderson, 1978; Moser y Stevens, 1992; Overall et al., 1995; Zimmerman, 2004; Schucany y Tony Ng, 2006; Moder et al., 2009; Rasch et al., 2011; Flores Muñoz, 2017; Flores y Ocaña, 2018; Flores et al., 2018) proponen no aplicar pruebas previas para probar una comparación de medias de dos muestras porque conduce a riesgos innecesarios. Así también parece ser que el enfoque de equivalencia es una solución a dicho problema. El inconveniente es que el estudio está incompleto (falta estudiar alejamientos de la normalidad, pre-testeo de normalidad y uso de pruebas no paramétricas).

Al ser esta probabilidad un parámetro desconocido, mediante un proceso de simulación estocástica se puede estimar como la proporción de veces que se rechaza una hipótesis de igualdad de medias que teóricamente es cierta cuando se aplican los distintos procedimientos de pre-testeo. Al ser una estimación se producirá un error el cual a través de la creación de funciones óptimas y el uso de procedimientos de reducción de varianza disminuirá, de tal modo que conseguiremos una estimación lo más perfecta posible.

El cálculo de esta estimación permitirá cuantificar la alteración de la TIEP cuando se utilizan los diversos métodos de pre-testeo señalados, y de acuerdo al criterio de Cochran nos permitirá identificar si alguno de ellos es adecuado para realizar una prueba de comparación de medias que no produzca una alteración en la TIEP que conlleve a errores innecesarios.

### ***1.3.2 Justificación practica***

Una de las principales necesidades estadísticas, ya sea para investigación o para cualquier proceso que implique toma de decisiones en las áreas de: econometría (Greene, 2000), biología (Quinn y Keough, 2002), Ingeniería (Ryan y Wiley InterScience (Online service), 2007), medicina (Dhar, 2010; Fagerland, 2012) y muchas otras más, es comparar la eficiencia promedio que tienen dos tratamientos, es decir los analistas esperan confirmar la hipótesis planteada ( $H_0: \mu_1 - \mu_2 = 0$ ), con un control de los errores que se generan en todo proceso de inferencia, garantizando tomar una decisión correcta. La mayoría de los investigadores desconocen ciertas limitaciones de los métodos clásicos y no conocen otras alternativas actualmente investigadas y aplicadas, por tal motivo es importante garantizar las condiciones de los contrastes estadísticos apropiadas para la ejecución oportuna que establecerán conclusiones acertadas, dado que el poder de las pruebas estadísticas tiende efecto como tal en el estudio a realizarse así como en el proceso de toma de decisiones y conclusiones de un trabajo de investigación (Sedlmeier y Gigerenzer, 1989, p. 309).

## **1.4 Objetivos**

### ***1.4.1 Objetivo General***

Estimar la TIEP global de las pruebas de comparación de medias cuando sus supuestos son previamente probados.

### ***1.4.2 Objetivos Específicos***

- Realizar un algoritmo de simulación que permita obtener valores de la estimación de la TIEP global de las pruebas de comparación de medias cuando se pre-testea sus supuestos mediante los enfoques tradicionales y de equivalencia.
- Comparar los resultados de las estimaciones de la TIEP global que permitan identificar las alteraciones producidas en los distintos procedimientos simulados.
- Determinar bajo un criterio técnico (Criterio de Cochran) si alguno de los procedimientos estudiados es adecuado para realizar pruebas de comparación de medias.

## CAPITULO II

### 2 MARCO TEORICO

#### 2.1 Pruebas de Hipótesis

La hipótesis es una suposición, afirmación o conjetura de la distribución de una o más variables aleatorias, el resultado de una hipótesis puede ser probablemente verdadera o falsa para ello se toma como base las distribuciones de probabilidad (distribuciones muestrales). Según (Wackerly, Mendenhall y L., 2009) los elementos de una prueba estadística son: hipótesis nula  $H_0$ , hipótesis alternativa  $H_a$ , estadístico de prueba y región de rechazo.

Rechazar una hipótesis no significa que ésta sea falsa, como tampoco el no rechazarla significa que sea verdadera, la decisión tomada no está libre de error, por tanto, surge dos tipos de error: el Error Tipo I significa rechazar una hipótesis que es verdadera cuando no debería ser rechazada, se denota con  $\alpha$  y se elige antes de realizar la prueba estadística para compararla con el p-valor obtenido. El Error Tipo II significa no rechazar una hipótesis que es falsa cuando debiera ser rechazada, se denota como  $\beta$ , mientras que  $1 - \beta$  es la probabilidad de rechazar una hipótesis que es falsa es conocida como “potencia de la prueba”.

Es deseable que estas dos probabilidades de error ( $\alpha$  y  $\beta$ ) sean pequeñas, pero eso no es siempre posible.

**Tabla 1-2:** Estado de la Naturaleza cuando se rechaza o no rechaza  $H_0$

Decisión	Estado de la Naturaleza	
	$H_0$ verdadera	$H_0$ falsa
No rechazo $H_0$	$1 - \alpha$ Nivel de confianza	Error de Tipo II $\beta$
Rechazo $H_0$	Error de Tipo I $\alpha$ Nivel de significación	$1 - \beta$ Potencia de prueba

Realizado por: Tania Sanchez, 2019

## 2.2 Prueba de hipótesis para comparar dos medias poblaciones

### 2.2.1 Prueba t-Student

Suponga que se seleccionaron muestras aleatorias independientes de distribuciones normales  $X_1, X_2, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$  y  $Y_1, Y_2, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$  con medias desconocida y que poseen varianzas iguales  $\sigma_1^2 = \sigma_2^2$ . Para la construcción de una prueba t-Student que compara las medias  $\mu_1$ ,  $\mu_2$ , las hipótesis se plantea de la siguiente manera:

La Hipótesis nula a contrastar es

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

Mientras que la hipótesis alternativa se define en la siguiente tabla.

**Tabla 2-2:** Hipótesis alternativas en l prueba t-Student .

Unilateral a la Izquierda	Unilateral a la derecha	Bilateral (dos colas )
$H_1: \mu_1 - \mu_2 > 0$	$H_1: \mu_1 - \mu_2 < 0$	$H_1: \mu_1 - \mu_2 \neq 0$

Fuente: Bartoszynski & Niewiadomska-Bugaj, 1996

Realizado por: Tania Sánchez, 2019

El estadístico de prueba es:

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

Donde:

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

La región de rechazo viene dada como:

**Tabla 3-2:** Región de rechazo en la prueba t-Student

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, 1996

Realizado por: Sánchez, T

### 2.2.2 Prueba de Welch

Cuando se requiere contrastar la hipótesis ( $H_0: \mu_1 - \mu_2 = 0$ ) y no se tiene evidencia para suponer  $\sigma_1^2 = \sigma_2^2$  (homocedasticidad). Welch (1947, p. 28-35) propone el siguiente planteamiento.

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (3)$$

T sigue una distribución t-Student con v grados de libertad, donde:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 / (n_1 - 1) + \left(\frac{S_2^2}{n_2}\right)^2 / (n_2 - 1)} \quad (4)$$

Y su región de rechazo es cuando  $|T| > t_{\alpha/2, v}$ .

### 2.2.3 Prueba de U de Mann Withney

U de Mann Withney es una prueba alternativa no paramétrica de la prueba a la t-Student para dos muestras independientes (Wilcoxon, 1946) . En términos generales, la prueba Wilcoxon puede ser definido como:

Sean  $x_1, x_1, \dots, x_{n1}$  y  $y_1, y_1, \dots, y_{n2}$  dos muestras aleatorias independientes tomadas de dos poblaciones continuas X y Y.

Se define el siguiente contraste de hipótesis:

$$H_0: \widetilde{\mu}_X = \widetilde{\mu}_Y \text{ (Las distribuciones de las poblaciones X y Y son idénticas)}$$

Mientras que la hipótesis alternativa se define en la siguiente tabla.

**Tabla 4-2 :** Hipótesis alternativas de la prueba U de mann withney

Una cola		Dos colas
Izquierda	Derecha	
$H_1: \widetilde{\mu}_X > \widetilde{\mu}_Y$ La distribución de la población X se desplaza a la izquierda de la distribución de la población Y	$H_1: \widetilde{\mu}_X < \widetilde{\mu}_Y$ La distribución de la población X se desplaza a la derecha de la distribución de la población Y,	$H_1: \widetilde{\mu}_X \neq \widetilde{\mu}_Y$ Las distribuciones de las poblaciones X y Y tienen localizaciones diferentes

Fuente: Wilcoxon, 1945

Realizado por: Tania Sánchez, 2019

El estadístico de prueba es:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \tag{5}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \tag{6}$$

$$U = \min \{U_1, U_2\} \tag{7}$$

La región de rechazo viene dada como se observa en la Tabla N° 5-2.

**Tabla 5-2 :** Región de rechazo de la prueba U de Mann Withney

Una cola		Dos colas
Izquierda	Derecha	
$U \geq n_1 n_2 - U_0$	$U \leq U_0$	$U \leq U_0$ o $U \geq n_1 n_2 - U_0$

Fuente: Wilcoxon, 1945

Realizado por: Tania Sánchez, 2019

### 2.3 Prueba de hipótesis para verificar los supuestos de normalidad

El estudio realizado por (Dufour et al., 1998, p. 1-3) menciona casi 40 pruebas de normalidad disponibles en la literatura estadística, estas pruebas se diferencian en características tales como valores de asimetría y curtosis, distribución, función característica, relación lineal (entre la distribución de la variable y la variable normal estándar (Z)), otro hallazgos en este mismo sentido, presentado por (Seier, 2002, p. 3) sugiere que también difieren de la distribución empírica con la distribución normal, naturaleza de la distribución y la complejidad de la estadística de prueba. Mientras (Park, 2015) clasifica dichas pruebas en dos categorías que son: estadísticas descriptivas (asimetría, curtosis) y métodos basados en la teoría.

La pruebas se clasifica en cuatro subcategorías: asimetría y curtosis, prueba de distribución empírica, prueba de regresión y correlación y otra prueba especial (Seier, 2002, pp. 2-4), es decir, los procedimientos basados en:

- Regresión y correlación: Shapiro Wilk, Shapiro-Francia y Ryan-Joiner),
- Distribución empírica: Kolmogorov-Smirnov KS, LL, AD y Cramer-Von Mises CVM),
- Pruebas de momento es decir de asimetría y curtosis: D'Agostino y la prueba JB ,
- Prueba de espaciamento: Rao's prueba u prueba de Greenwood.
- Otras pruebas tales como chi-cuadrado (x2), Estandarizados 3 ° (Vb1), 4 ° momento (Vb2), prueba CSQ etc.

### 2.3.1 Test de Shapiro Wilk (SW)

Shapiro Wilks (SW) es una prueba de regresión y correlación más conocida, fue establecida por (Shapiro y Wilk, 1965) siendo la primera prueba capaz de detectar desviaciones de normalidad, debido al sesgo, curtosis o ambos (Althouse, 1998, p. 27-29). Años más tarde (Royston, 1982, p. 115) modifico esta prueba para ampliar la restricción del tamaño de la muestra a 2000 en un algoritmo AS181. Sin embargo, observo que aún eran incorrectos modificándolo a un rango  $3 \leq n \leq 5000$  implementado en el algoritmo AS R94 (Royston, 1992, p. 547).

Debido a sus buenas propiedades de potencia se ha convertida en la prueba preferida por los investigadores (Mendes y Pala, 2003, p. 135).

La hipótesis a contrastar se define como:

$H_0$ : La variable X sigue una distribución Normal

$H_1$ : La variable X no sigue una distribución Normal

- Si la prueba de normalidad es estadísticamente significativa ( $p \leq .05$ ), se concluye que la muestra proviene de una población no normal.
- Si la prueba de normalidad no es estadísticamente significativa ( $p > .05$ ), se tiene evidencia insuficiente para rechazar la hipótesis nula, y se procede como si la población fuera normal

SW se podría definir generalmente como:

Dada una muestra aleatoria ordenada Dada una muestra aleatoria ordenada,  $X_1 < X_1 < \dots < X_n$  la estadística de prueba Shapiro Wilk original se define como:

$$W = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

$$b = \sum_{i=1}^n a_i [X_{(n-i-1)} - x_i] \quad (9)$$

$$a_i = m'V^{-1}(m'V^{-1}m)^{-1/2} \quad (10)$$

Donde:

$[X_{(n-i-1)} - x_i]$  Son las diferencias sucesivas que se obtiene al restar el primer valor al último valor, el segundo al antepenúltimo y así sucesivamente.

$a_i$ : Son los coeficientes tabulados en la tabla de Shapiro Wilks. La ecuación N°10 calcula los coeficientes, donde  $m'$  son los valores esperados de las estadísticas de orden de variables aleatorias independientes e idénticamente distribuidas a partir de la distribución normal estándar y  $V$  es la matriz de covarianza.

Se rechaza  $H_0$  cuando  $W \leq W_{\alpha,n}$

El valor de SW se encuentra entre cero y uno, los valores pequeños conducen al rechazo de la hipótesis, mientras que un valor de uno indica la normalidad de los datos.

## 2.4 Test de hipótesis para verificar el supuesto de homocedasticidad

Para pre - testear el supuesto de homocedasticidad, existe una cantidad de pruebas tales como: Fisher, Levene, Bartlett, F max de Hartley, Fligner-Killeen, Cochran, Layard, Box-Andersen, Bootstrap, Bartlett modificado,  $\chi^2$ , BoxScheffé, entre otros. Dado que existen tantas pruebas, tan solo se menciona la prueba F como un test para probar este supuesto.

### 2.4.1 Enfoque clásico: Prueba Fisher-Snedecor (F)

Sea  $X_1, X_2, \dots, X_m$  una muestra aleatoria simple de una población  $N(\mu_1, \sigma_1^2)$  y  $Y_1, Y_2, \dots, Y_n$  una muestra aleatoria simple de una población  $N(\mu_2, \sigma_2^2)$  de tamaño  $m$  y  $n$ , suponga que las muestras se eligen de manera independiente. Los valores de las medias,  $\mu_1$  y  $\mu_2$  aquí son irrelevantes, se trata solo con varianzas  $\sigma_1^2$  y  $\sigma_2^2$ .

Para probar este supuesto donde dicho contraste viene dado por las siguientes hipótesis:

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

La prueba habitual para comparar dos varianzas es la prueba F basada en las dos varianzas de muestr

a en la

estadíst

ica.

$$F = \frac{S_1^2}{S_2^2} \quad (11)$$

El estadístico de prueba representa el cociente de las dos varianzas muestrales

Donde:

$$S_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2 \quad (12)$$

$$S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (13)$$

La distribución F tiene dos valores para los grados de libertad (GL): se asocia con el numerador y denominado ( $F_{gl1,gl2}$ )

Si  $s_1^2 > S_2^2$  se rechazar  $H_0$  si  $F > F_{(n_1-1, n_2-1, 1-\alpha)}$ , de lo contrario, el GL en el F-quantile tienen que intercambiarse.

### 2.4.2 Enfoque de equivalencia: prueba de dispersión de dos distribuciones Gaussiana

Wellek (2010, p. 164-167) presenta un test potente para probar el supuesto, llamado “Prueba de equivalencia de dispersión para dos distribuciones Gaussiana”, construido a partir del lema generalizado de Nyman y Pearson, donde la región crítica óptima es  $\{x | c_1 < T(X) < C_2\}$ .

Este pre-test podría definirse como:

Dado dos muestras aleatorias independiente  $X_1, X_2, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$  y  $Y_1, Y_2, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ , la zona de inferencia o en el caso región de tolerable heterogeneidad entre  $\sigma_1^2$  y  $\sigma_2^2$ , esta definida por el intervalo  $(\omega_1^2$  y  $\omega_2^2)$  la cual es suficientemente corta alrededor de 1, cubriendo el valor ratio  $\frac{\sigma_1^2}{\sigma_2^2}$ . Mediante la investigación realizada por (Flores y Ocaña, 2018, pp. 59-72) se definirá los valores para la zona de indiferencia  $\omega_1^2$  y  $\omega_2^2$ .

Las hipótesis a contrastar son:

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} < \omega_1^2 \wedge \frac{\sigma_1^2}{\sigma_2^2} < \omega_2^2 \quad \text{Hipótesis de no equivalencia (diferencia relevante de varianzas)}$$

$$H_1: \omega_1^2 < \frac{\sigma_1^2}{\sigma_2^2} < \omega_2^2 \quad \text{Hipótesis de equivalencia (diferencia no relevante)}$$

Para esta prueba, la hipótesis alternativa establece la “equivalencia” es decir ajuste perfecto igualdad de varianza excepto por desviaciones irrelevantes y las diferencias relevantes entre varianzas se establecen en la hipótesis nula.

Donde se satisface que  $\omega_1^2 < \frac{\sigma_1^2}{\sigma_2^2} < \omega_2^2$

Es test estadístico es:

$$Q = \frac{S_X^2}{S_Y^2} = \frac{n_2 - 1 \sum_{t=1}^{n_1} (X_t - \bar{X})^2}{n_1 - 1 \sum_{t=1}^{n_2} (Y_t - \bar{Y})^2} \quad (14)$$

La región de rechazo viene dada para:

Caso Balanceado

$$\left\{ \frac{1}{\tilde{C}_{\alpha, n-1}^{(0)}(\omega_2^2)} < Q < \tilde{C}_{\alpha, n-1}^{(0)}(\omega_2^2) \right\} \quad (15)$$

Caso desbalanceado

$$\left\{ \tilde{C}_{\alpha, n_1-1, n_2-1}^{(1)}(\omega_1^2, \omega_2^2) \right\} < Q < \left\{ \tilde{C}_{\alpha, n_1-1, n_2-1}^{(2)}(\omega_1^2, \omega_2^2) \right\} \quad (16)$$

Donde:

$\tilde{C}_{\alpha, n_1-1, n_2-1}^{(1)}(\omega_1^2, \omega_2^2)$ ,  $\tilde{C}_{\alpha, n_1-1, n_2-1}^{(2)}(\omega_1^2, \omega_2^2)$  son las constantes críticas se determinan mediante el

$$\begin{aligned} F_{n_1-1, n_2-1} \left( \frac{\tilde{C}_2}{\omega_1^2} \right) - F_{n_1-1, n_2-1} \left( \frac{\tilde{C}_1}{\omega_1^2} \right) &= \alpha \\ &= F_{n_1-1, n_2-1} \left( \frac{\tilde{C}_2}{\omega_2^2} \right) - F_{n_1-1, n_2-1} \left( \frac{\tilde{C}_1}{\omega_2^2} \right) \end{aligned} \quad (17)$$

sistema de ecuaciones:

Si el estadístico  $Q$  este dentro de la región establecida se rechaza el  $H_0$ .

## 2.5 Método de Variables de Control para reducir varianzas

El método de variables de control es una técnica alternativa al método de Monte Carlo el cual permite estimar el parámetro de interés y el error estándar. (Ocaña y Vegas, 1995, p. 631-632) proponen un método para obtener una mejor precisión en una simulación cuando el parámetro es una probabilidad y el estimador es una proporción. Considere en una simulación de Monte Carlo, la variable respuesta  $Y$  de tipo dicotómica (presencia o ausencia de un atributo). Por ejemplo para este estudio adaptamos la variable respuesta  $Y \rightarrow$  *rechazo/no rechazo de una hipótesis*, donde el parámetro de interés es la verdadera probabilidad de rechazo denotada como  $p_i = E(Y)$  y el estimador es la frecuencia relativa  $\widehat{p}_1$  del evento  $\{Y = 1\}$ . Sea  $C$  una variable dicotómica cuya esperanza conocida  $E(C) = p_{.1}$ , que al correlacionarse esta con la variable  $Y$ , se obtiene tablas de contingencia  $2 \times 2$  correspondiente a las frecuencias relativas y de las probabilidad de los pares  $(Y_i; C_i)$   $i = 0,1$  como se observa en las Tablas N° 6 -2 y Tabla N° 7-2.

.

**Tabla 6-2:** Tabla de contingencia  $2 \times 2$  de las frecuencias relativas de los pares  $(Y_i; C_i)$   $i = 0,1$

	<b>C=0</b>	<b>C=1</b>	
<b>Y=0</b>	$n_{00}$	$n_{01}$	$n_{0.}$
<b>Y=1</b>	$n_{10}$	$n_{11}$	$n_{1.}$
	$n_{.0}$	$n_{.1}$	$n$

Fuente: Ocaña & Vegas, 1995

Realizado por: Tania Sánchez, 2019

**Tabla 7-2:** Tabla de contingencia  $2 \times 2$  de las probabilidades de los pares  $(Y_i; C_i)$   $i = 0,1$

	<b>C=0</b>	<b>C=1</b>	
<b>Y=0</b>	$p_{00}$	$p_{01}$	$p_{0.}$
<b>Y=1</b>	$p_{10}$	$p_{11}$	$p_{1.}$
	$p_{.0}$	$p_{.1}$	<b>1</b>

Fuente: Ocaña & Vegas, 1995

Realizado por: Tania Sánchez, 2019

(Rothery, 1982, pp. 125-129) demostró que la estimación de máxima verosimilitud clásica del parámetro  $p_1$  es  $\tilde{p}_{.1}$  como se observa en la Ecuación N° 18 que al ser asintóticamente normal e insesgado, este estimador coincide con el estimador de variables de control ajustado por regresión propuesta por (Ocaña y Vegas, 1995, p. 163) cuando se utiliza una variable de control .

$$\tilde{p}_{.1} = p_{.0} \frac{n_{10}}{n_{00} + n_{10}} + p_{.1} \frac{n_{11}}{n_{01} + n_{11}} \quad (18)$$

El estimador de varianza para  $p_1$  viene dado por:

$$\hat{\sigma}^2 = \frac{\tilde{p}_{00} \tilde{p}_{10}}{p_{.0}(n+1) - 2} + \frac{\tilde{p}_{01} \tilde{p}_{11}}{p_{.1}(n+1) - 2} \quad (19)$$

$$\widehat{\%VR} = \left( 1 - \frac{(n-1)\hat{\sigma}^2}{\hat{p}_{.1}(1-\hat{p}_{.1})} \right) * 100 \quad (20)$$

y la estimación del porcentaje de reducción de la varianza es:

## 2.6 Método de Fleishman

Para generar datos no normales, utilizaremos el procedimiento propuesto por (Fleishman, 1978) conocido con el nombre de “Método Poder de Fleishman”, el cual utiliza una transformación de una función polinomial. Específicamente, toma la suma de una combinación lineal de un variable aleatoria normal, su cuadrado y su cubo como se muestra en la Ecuación N° 22. Donde Z es una variable con distribución desconocido y con parámetros ( $\mu = 0$ ;  $\sigma^2 = 1$ ;  $\gamma_1$ ;  $\gamma_2$ ), X es una variable aleatoria normalmente distribuida con media 0 y varianza 1  $X \sim N(0,1)$ . 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 ( $\gamma_2$ ). Refiriéndose a este contexto “la curtosis y la asimetría son los indicadores más importantes de la extensión a lo que la anormalidad afecta las inferencias usuales hechas en el análisis de varianza” (Scheffe, 1959).

$$Z = a + bX + cX^2 + dX^3 \quad (22)$$

Los coeficientes de asimetría y curtosis determinan el grado de contaminación de la normal. Distintos niveles de contaminación fueron propuestos por (Blanca et al., 2013, p. 365-371) los cuales se muestran en la Tabla N° 8-3.

**Tabla 8-3:** Coeficientes de Fleishman

Niveles de Contaminación	Asimetría ( $\gamma_1$ )	Kurtosis ( $\gamma_2$ )	Coeficientes de Flesihman (a,b,c,d)
Sin contaminación	0	0	(0,1,0,0)
Contaminación Baja	0.25	0.70	(-0.037,0.933,0.037,0.021)
Contaminación Moderado	0.75	1	(-0.119,0.956,0.119,0.0098)
Contaminación Alto	1.30	2	(-0.249,0.984,0.249,-0.016)
Contaminación Severo	2	6	(-0.314,0.826,0.314,0.023)

Fuente: Blanca et al., 2013

Realizado por: Tania Sánchez, 2019

## CAPITULO III

### 3 MARCO METODOLOGICO

Ya que el objetivo general de esta investigación es estimar la TIEP global de las pruebas de comparación de medias cuando sus supuestos son previamente probados, entonces primero se determinó los mejores pre - test para probar el supuesto de normalidad y homocedasticidad, luego creamos un algoritmo mediante un proceso de simulación estocástica que nos permita determinar la TIEP global en una prueba de comparación de medias. Para alejamientos de los supuestos, se propuso niveles de no normalidad leve moderada, alta, severa y niveles de heterocedasticidad dados por la razón  $\left(\frac{\sigma_1}{\sigma_2}\right) = 1, 1.2, 1.6, 2, 2.5$  y 3 (donde 1 representa perfecta homocedasticidad y 3 la heterocedasticidad más fuerte). Se usó tamaños muestrales balanceados  $n_1, n_2 = (5, 5) (10, 10)$  y desbalanceados  $n_1, n_2 = (5, 10) (10, 5)$ . Finalmente, para reducir la variabilidad de la estimación de la TIEP global se usó la técnica de reducción de varianzas denominada “Variables de Control”, la cual queda estimada como la proporción de rechazos de la hipótesis nula de igualdad de medias, cuando teóricamente esta es cierta.

#### 3.1 Obtención del test más potente para probar normalidad.

De las 7 opciones de pruebas utilizadas en la mayoría de libros y paquetes estadísticos: a) Kolmogorov Smirnov KS, b) Liliefors LL, c) Shapiro Wilk SH, d) Anderson Darling AD, e) Jarke Bera JB, f) Cramer-von Mises CVM y g) Pearson P, buscamos la prueba más potente para pretestear normalidad, para esto se usaron muestras no normales y se estimó la potencia mediante una función de simulación que estima esta probabilidad como la proporción de veces que se rechaza  $H_0$  (La muestra sigue una distribución normal) cuando esta es falsa.

La función usada se llama “*power.norm*” (Ver anexo D), donde los argumentos que se usaron para obtener los resultados de la Potencia fueron:  $m = 100000$  réplicas de simulación,  $n = 5, 10, 15, 20, 30$  tamaños muestrales, *coef.fleishman* alejamientos de la normalidad en diferentes niveles de intensidad (leve, moderada, alta y severa) y *semilla* = 127.

```
power.norm(m = 100000, n, alpha, coef.fleishman, semilla = 127)
```

Para determinar los diferentes niveles de contaminación de la normalidad, se utilizó distintas combinaciones de asimetría ( $\gamma_1$ ) y curtosis ( $\gamma_2$ ) obtenidas a partir de la Tabla N°10-4, que da valores para las contaminaciones según lo estudiado por (Blanca et al., 2013) y los coeficientes de fleishman (a, b, c, d) fueron calculados mediante la función “*fleishman.coef*” implementada en la librería BinNonNor (Inan y Demirtas, 2015).

Con respecto al nivel de significancia “alpha “, se consideró un  $\alpha = 0.05$ , esto debido a que al utilizar  $\alpha = 0.05, 0,10$  los patrones de la potencia fueron similares (Yap y Sim, 2011, p. 2153).

Una vez obtenidas las muestras, la función “*power.norm*” implementa internamente una función propia llamada “*rechazo.val*” (Ver anexo D), que determina vectores con los rechazos de la hipótesis nula de normalidad para diferentes test especificados en la Tabla N° 9-3. La proporción de estos rechazos (razón entre el total de rechazos y el número de muestras generadas) la usamos como el mejor estimador de la Potencia buscada.

**Tabla 9-3 :** Funciones utilizadas para las pruebas de normalidad

<b>Prueba</b>	<b>Función en R</b>
Kolmogorov Smirnov	ks.test,
Liliefors	lillie.test,
Shapiro Wilk	shapiro.test,
Anderson Darling	ad.test,
Jarke Bera	ajb.norm.test,
Cramer-von Mises	pearson.test
Pearson	cvm.test

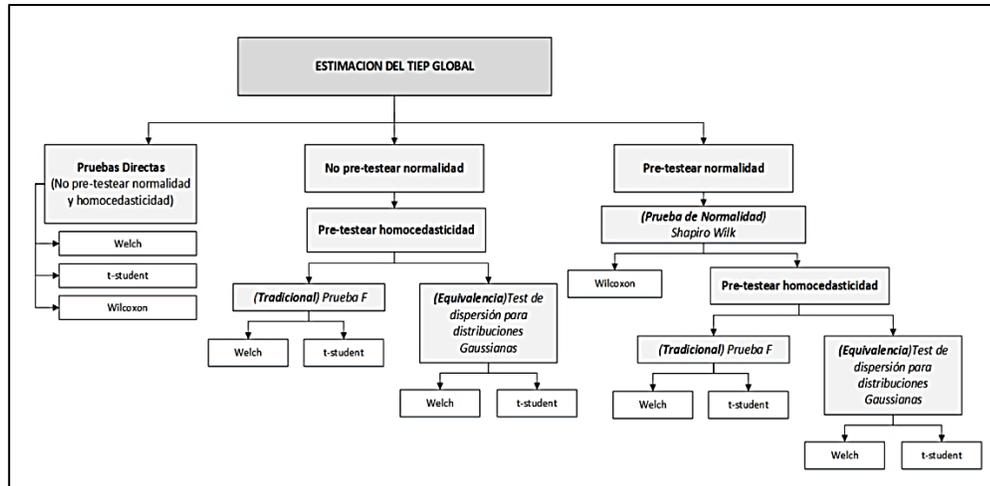
**Fuente:** Team R

**Realizado por:** Tania Sánchez, 2019

### **3.2 Proceso de simulación para la estimación del TIEP Global en pruebas de comparación de dos medias.**

Se utilizó la prueba Shapiro Wilks obtenida como resultado de la función “*Power.norm*”, ya que es considerado como la prueba más potente para pre-testear el supuesto de normalidad. En el caso de homocedasticidad se usó la prueba tradicional F (Fisher-Snedecor) como el test previo para contrastar homocedasticidad, dado que se basa en el mismo estadístico Q de la prueba de equivalencia de dispersión para dos distribuciones Gaussianas, esto hace que la comparación del pre-test tradicional con el pre-test de irrelevancia tenga mayor sentido. Además en el proceso de pre-testeo, la prueba F parece mostrarse más robusta que otras (por ejemplo Levene), para distintos niveles de contaminación de la normalidad y la homocedasticidad (Flores, Ocaña y Sánchez, 2018, pp. 10-17).

Una vez determinadas los pre-test adecuados para probar cada uno de los supuestos, se establecieron siete procedimientos comparables como se muestra en la Figura N°1-3. En los tres primeros se aplicó directamente la prueba t-Student (DT), Welch (DW) y Wilcoxon (DWIL) sin realizar ninguna prueba previa para determinar normalidad y homocedasticidad, en el cuarto y quinto procedimiento se realizó solamente el pre-test de igualdad de varianzas mediante las pruebas F (FTW) y de equivalencia de dispersión (ETW ) antes de decidir entre la prueba t-Student o Welch, y en los dos últimos se pre-testeó el supuesto de normalidad con la prueba Shapiro Wilk antes de decidir entre la prueba Wilcoxon o pre-testear el supuesto de igualdad de varianzas mediante las pruebas F (SFTW) y de equivalencia de dispersión (SETW) antes para decidir entre la prueba t-Student o Welch.



**Figura 1-3: Procedimientos para la estimación del TIEP Global**  
Realizado por: Tania Sánchez, 2019

Para estimar la TIEP en cada uno de los procesos descritos, construimos la función “*TIEP.norm*” (Anexo F), la cual mediante un algoritmo de simulación estima la TIEP como la proporción de rechazos de  $H_0: \mu_1 - \mu_2 = 0$  en una prueba de comparación de medias, garantizando el cumplimiento teórico de esta hipótesis.

```

TIEP.norm (m = 100000, semilla = 127, n1 = 5, n2 = 5, diffmean = 0,
ratioSigmas = 1, nominalAlpha = 0.05, omegaratio1 = 0.75^2,
omegaratio2 = 1/omegaratio1, alphaPretest = 0.05,
coef.fleishman){

```

Los argumentos de la función “*TIEP.norm*” que se usaron para obtener los resultados de la TIEP fueron:  $m = 100000$  réplicas de simulación, tamaños muestrales  $n_1, n_2 = (5, 5) (10, 10)$  tanto para casos balanceados como para desbalanceado  $n_1, n_2 = (5, 10) (10, 5)$ . Bajo las siguientes condiciones: todos los procedimientos con muestras provenientes de una distribución normal con media cero  $diffmean = 0$ , distintos niveles de heterogeneidad de 1, 1.2, 1.6, 2, 2.5 y 3, distintos niveles de contaminación de la normalidad *coef.fleishman* (leve, moderada, alta y severa) y con límites de irrelevancia para las pruebas de equivalencia *omegaratio1* y *omegaratio2*.

Los niveles de heterocedasticidad son determinados a partir de la razón entre varianzas  $\frac{\sigma_1}{\sigma_2}$ , ya que la distancia en valor absoluto de un par de varianzas no afecta la estimación de la TIEP en un proceso de pre-testeo, en su lugar lo que realmente interesa es la distancia de la razón entre ellas. Los niveles de contaminación de la normal fueron generados con el sistema de Flesihman a partir de la función “*fleishman.coef*” implementada en la librería BinNonNor cuyos parámetros fueron distintas combinaciones de la simetría y curtosis según corresponda la intensidad requerida como se observa en la Tabla N° 8-3 propuesta por (Blanca et al., 2013).

Los límites de equivalencia o irrelevancia  $\omega_1^2, \omega_2^2$  “*omegaratio1*” y “*omegaratio2*” para plantear las hipótesis de una prueba de equivalencia de dispersión de dos distribuciones Gaussianas fueron calculadas a partir de la función “*limvarRatios*” propuesto por (Flores Muñoz, 2017, p. 38) los resultados se muestra en la Tabla N° 10-3.

Con respecto al nivel de significancia, se consideró un “*nominalalpha*” de  $\alpha = 0.05$ , esto debido a que utilizar  $\alpha = 0.01, 0.05$  y  $0,10$  no influye en la estimación TIEP global (Flores Muñoz, 2017, p. 17).

**Tabla 10-3:** Límites de irrelevancia

Tamaño de muestra ( $n_1, n_2$ )	Límite de irrelevancia ( $\omega_1^2, \omega_2^2$ )
(5,5)	(0.2218,4.5079)
(10,10)	(0.1054,9.84843)
(5,10)	(0.7471,1.2726)
(10,5)	(0.7842,1.34425)

**Fuente:** Flores Muñoz, 2017

**Realizado por:** Tania Sánchez, 2019

Básicamente, la función “*TIEP.norm*” (Ver Anexo F), 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 y homocedasticidad indicados en secciones anteriores. A partir de estas muestras se aplican los diferentes procedimientos establecidos en la Figura N° 1-3, 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, del cual obtenemos una mejor precisión cuando aplicamos el método de reducción de varianzas “Variables de Control”.

### 3.3 Funciones previas usadas en TIEP.norm

La función “*TIEP.norm*” implementa internamente las siguientes funciones previas (Ver en el Anexo E):

- La función “*cont*” construye una tabla de contingencia 2\*2 de frecuencias relativas a partir de dos variables.
  
- La función “*estimControl*” calcula la estimación de una probabilidad mediante la técnica de variables de control para reducir varianzas, dada la tabla de frecuencias relativas y una probabilidad nominal (esperanza de una variable de control dicotómica)
  
- La función “*disp.eq*”, “*t.val*” y “*wel.val*” calcula los límites de la región crítica que establece la aceptación o rechazo de la hipótesis nula  $H_0$  de equivalencia de dispersión, las pruebas t-Student y Welch respectivamente.
  
- La función “*val.in*” devuelve TRUE si un valor se encuentra dentro de un intervalo, y FALSE en caso contrario.
  
- La función “*p.val*” devuelve el p valor de la prueba Wilcoxon
  
- La función “*rechazo\_pval*” devuelve TRUE si el p valor de una prueba estadística es mayor que el nivel de significancia y FALSE en caso contraria.
  
- La función “*rechazos\_sin\_pretest*” devuelve TRUE, cuando el valor del estadístico de la prueba para contrastar la igualdad de medias wilcoxon, t-Student y welch (cuando no se pre-testeo los supuestos) es mayor que estadístico tabulado,
  
- La función “*rechazo\_pretest\_hom*” devuelve TRUE, cuando el valor del estadístico de la prueba para contrastar la igualdad de medias t-Student y welch (cuando no se pre-testeo normalidad, pero sí homogeneidad) es mayor que estadístico tabulado,

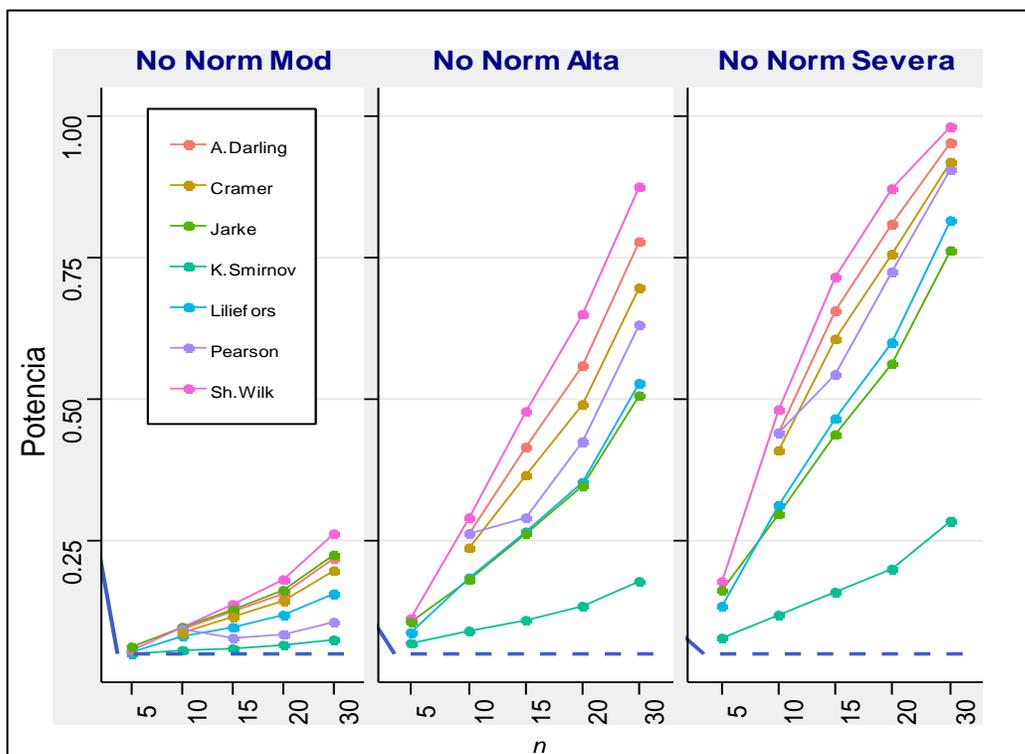
- La función “*rechazo\_pretest\_norm\_hom*” devuelve TRUE, cuando el valor del estadístico de la prueba para contrastar la igualdad de medias wilcoxon, t-Student y welch (cuando se pre-testeo ambos supuestos) es mayor que estadístico tabulado.

## CAPITULO IV

### 4 MARCO DE RESULTADOS, DISCUSIÓN Y ANÁLISIS DE RESULTADOS

#### 4.1 Resultados Preliminares

##### 4.1.1 Potencia de las pruebas de normalidad



**Figura 2-4: Estimación de la potencia en varios de niveles de contaminación normal**

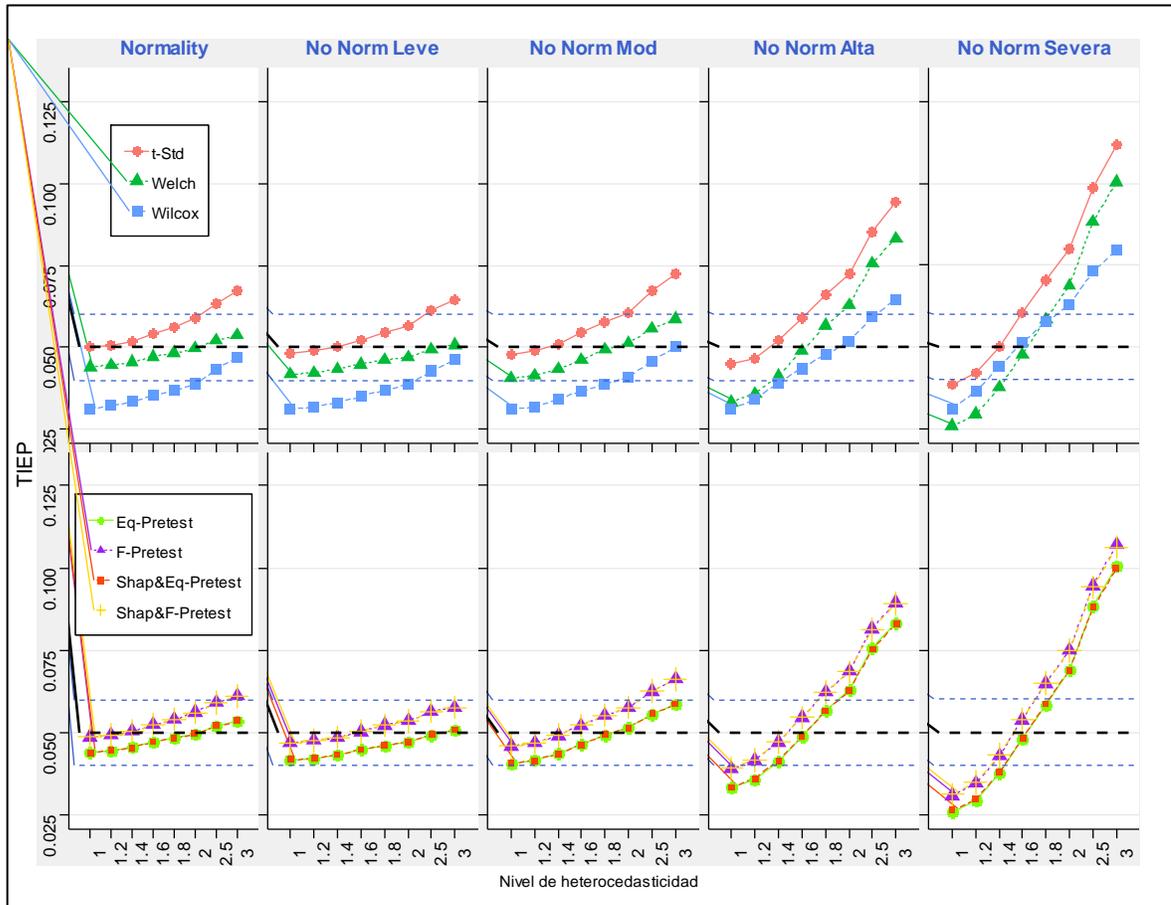
Realizado por: Tania Sánchez, 2019

Para los distintos niveles de contaminación de la normal moderada, alta y severa mostrados en la Figura N° 2-4, se puede observar que la potencia simulada va creciendo conforme aumenta el tamaño de la muestra, este comportamiento es el mismo para los 7 test utilizados, estos resultados concuerdan también con las observaciones realizadas por (Farrell y Rogers-Stewart, 2006, p. 811).

De acuerdo a lo observado, se concluye que, Shapiro-Wilk es la prueba con mayor potencia para todos los niveles de no normalidad y tamaños de muestra, mientras que la prueba de Kolmogorov-

Smirnov es la menos poderosa, de hecho, existen múltiples investigaciones (D'agostino et al., 1990; Mendes y Pala, 2003; Öztuna et al., 2006; Yazici y Yolacan, 2007; Yap y Sim, 2011) concuerdan con esta conclusión e incluso al respecto Tukey menciona "La construcción de Kolmogorov – Smirnov es lógicamente correcto, pero prácticamente inútil" (Tukey, 1993, p. 31).

## 4.2 Resultados Finales



**Figura 3-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales  $n_1 = 5, n_2 = 5$  con una significancia  $\alpha = 0.05$ .**

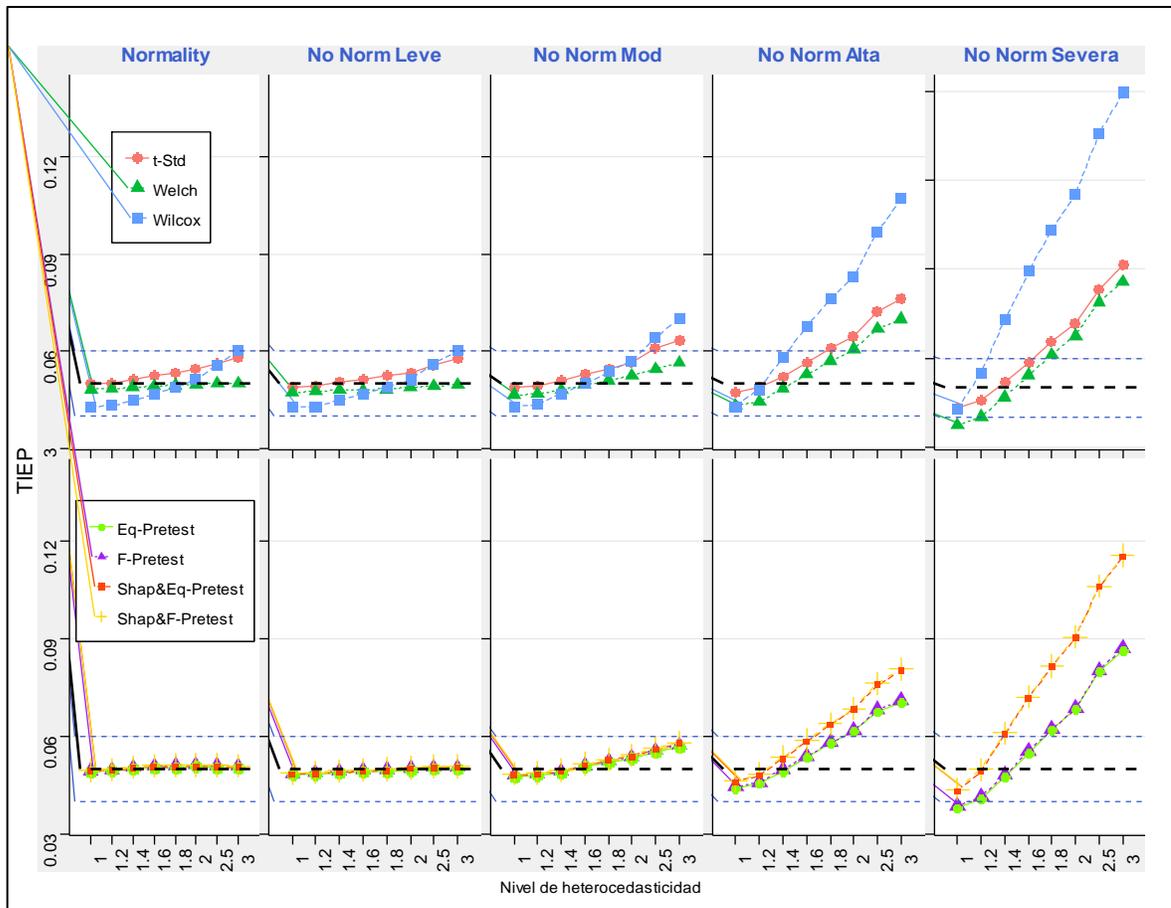
Realizado por: Tania Sánchez, 2019

La Figura N° 3-4 muestra la estimación de la TIEP global cuando dos medias poblacionales son comparadas en tamaños muestrales pequeños  $n_1=5$  y  $n_2=5$ .

La parte superior del gráfico muestra la estimación para los procedimientos directos (sin ningún pre-test). En estos casos, se observa que el test de Welch permanece robusto (dentro de los límites de Cochran), para cualquier nivel de heterocedasticidad, pero solo hasta una contaminación

moderada de la normalidad, para contaminaciones superiores a esta, parece ser que este test solo es robusto en determinados puntos de heterocedasticidad. Por otra parte, las pruebas t-Student y Wilcoxon muestran una TIEP totalmente descontrolada en la mayoría de los casos representados.

La parte inferior del gráfico muestra la estimación para los procedimientos en que se incluyen los pre-test. En estos casos, se puede observar que los cuatro procedimientos parecen mantenerse robustos, al menos hasta una contaminación moderada de la normalidad, independientemente del grado de heterocedasticidad, sin embargo, los procedimientos que más se acercan al nivel de significancia son aquellos que involucran en el proceso de pre-testeo a la prueba de equivalencia (ETW, SETW). Finalmente, para contaminaciones Alta y Severa de la normalidad, no podemos definir ninguna técnica como robusta, puesto que en la mayoría de escenarios la TIEP se encuentra fuera de los límites de robustez propuestos por Cochran.



**Figura 4-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales  $n_1 = 10, n_2 = 10$  con una significancia  $\alpha = 0.05$ .**

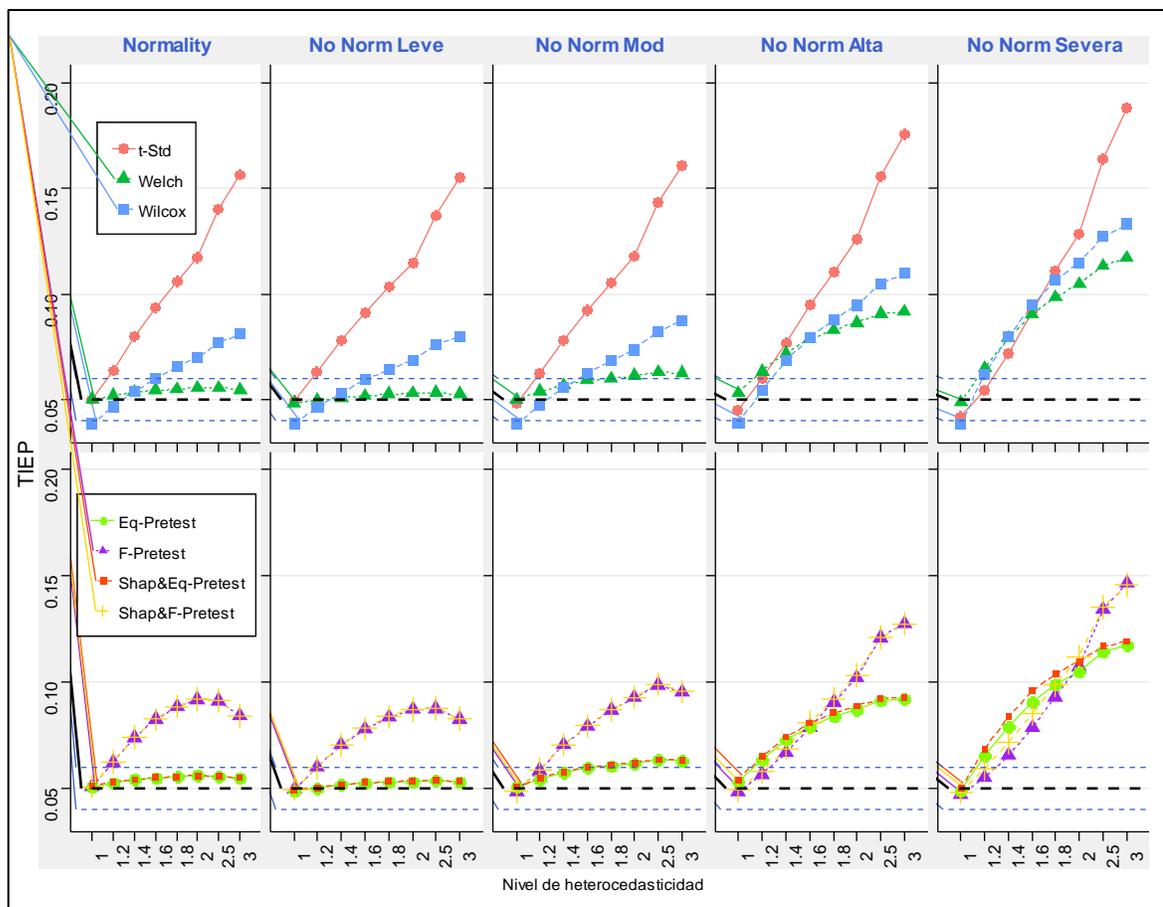
Realizado por: Tania Sánchez, 2019

La Figura N° 4-4 muestra la estimación de la TIEP global cuando dos medias poblacionales son comparadas en tamaños muestrales grandes  $n_1 = 10, n_2 = 10$ .

La parte superior del gráfico muestra la estimación para los procedimientos directos (sin ningún pre-test). En estos casos, se observa que el test de t-Student, Welch y Wilcoxon permanecen robusto (dentro de los límites de Cochran), para cualquier nivel de heterocedasticidad, pero solo hasta una contaminación moderada de la normalidad. Por otra parte, para contaminaciones alta y moderada, de la normalidad estos tests solo son robustos en determinados puntos de heterocedasticidad.

La parte inferior del gráfico muestra la estimación para los procedimientos en que se incluyen los pre-test. En estos casos, se puede observar que los cuatro 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 grado de heterocedasticidad al menos hasta una contaminación moderada de la normalidad. Nuevamente, para contaminaciones Alta y Severa de la normalidad, no existe ninguna técnica como robusta, puesto que en la mayoría de escenarios la TIEP se encuentra fuera de los límites de robustez propuestos por Cochran y se aleja más a medida que aumenta el nivel de heterocedasticidad.



**Figura 5-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales  $n_1 = 5$ ,  $n_2 = 10$  con una significancia  $\alpha = 0.05$ .**

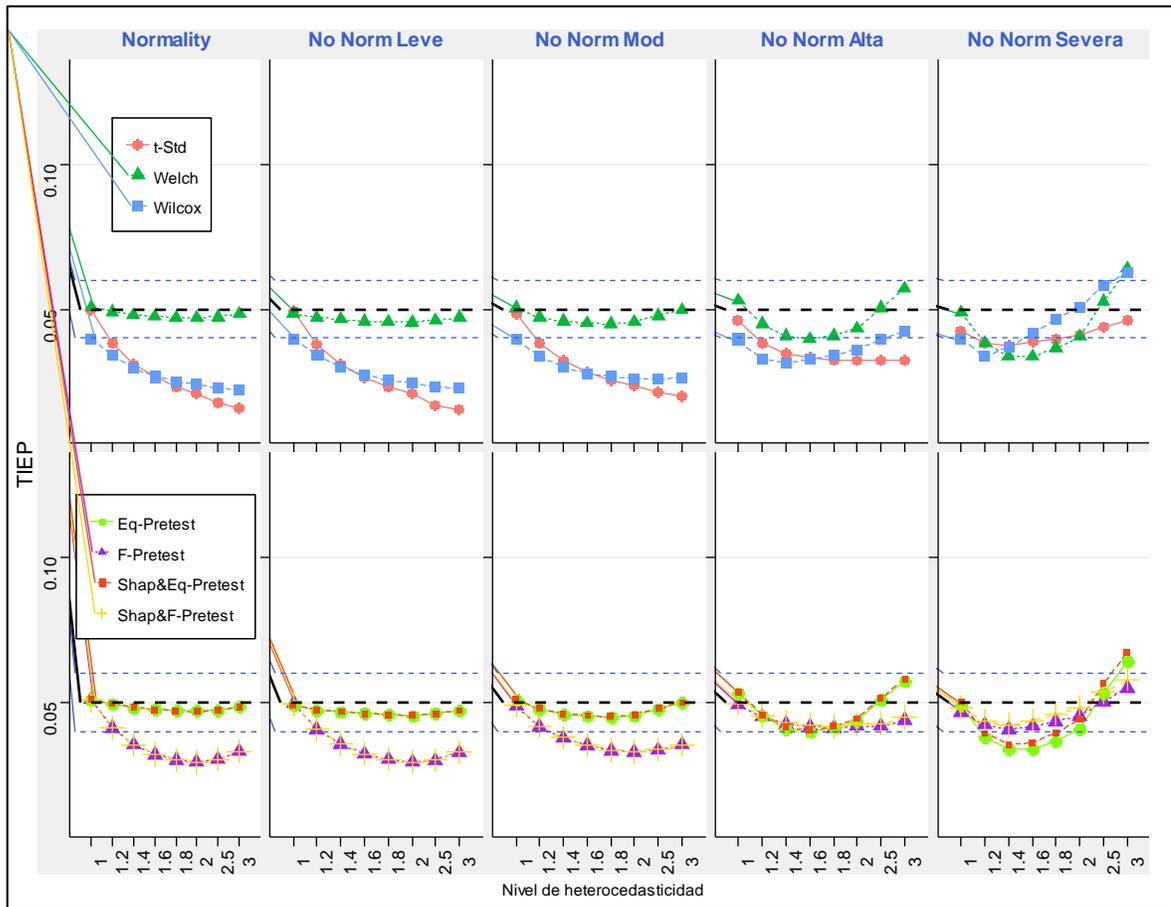
Realizado por: Tania Sánchez, 2019

La Figura N° 5-4 muestra la estimación de la TIEP global cuando dos medias poblacionales son comparadas en tamaños muestrales desbalanceados  $n_1 = 5$ ,  $n_2 = 10$ .

La parte superior del gráfico muestra la estimación para los procedimientos directos (sin ningún pre-test). En estos casos, se observa que la TIEP estimada, independiente del nivel de heterocedasticidad controla bastante bien alrededor del nivel de significancia cuando se utiliza

directamente el test de Welch, esto únicamente para niveles de contaminación de la normalidad cero, leve y moderado, donde la TIEP se encuentra dentro los límites de robustez propuestos por Cochran para cualquier grado de heterocedasticidad. A partir de una contaminación alta el uso de este test deja de ser robusto en determinados puntos de heterocedasticidad. Por parte las pruebas t-Student y Wilcoxon muestran una TIEP totalmente descontrolada en la mayoría de los casos representados.

La parte inferior del gráfico muestra la estimación para los procedimientos en que se incluyen los pre-test. En estos casos, se puede observar que los procedimientos que involucran en el proceso de pre-testeo a la prueba de equivalencia (ETW, SETW) parecen mantenerse robustos, al menos hasta una contaminación moderada de la normalidad, independientemente del grado de heterocedasticidad. Al contrario, los procedimientos que involucran el proceso de pre-testeo a la prueba tradicional F (FTW, SFTW) producen un descontrol mayor a la TIEP por encima del nivel de significancia en la mayoría de los casos representados. Finalmente, para contaminaciones Alta y Severa de la normalidad, no podemos definir ninguna técnica como robusta.



**Figura 6-4: Estimación de la TIEP global de una prueba de comparación de medias cuando se pre-testean sus supuestos, usando tamaños muestrales  $n_1 = 10$ ,  $n_2 = 5$  con una significancia  $\alpha = 0.05$**

Realizado por: Tania Sánchez, 2019

La Figura N° 6-4 muestra la estimación de la TIEP global cuando dos medias poblacionales son comparadas en tamaños muestrales pequeños  $n_1 = 10$ ,  $n_2 = 5$ .

La parte superior del gráfico muestra la estimación para los procedimientos directos (sin ningún pre-test). En estos casos, se observa que el test de Welch permanece robusto (dentro de los límites de Cochran), para cualquier nivel de heterocedasticidad, pero únicamente para niveles de contaminación de la normalidad cero, leve y moderado. A partir de contaminación alta de normalidad parece ser que este test solo es robusto en determinados puntos de heterocedasticidad. Lo dicho hasta aquí supone que hasta esta instancia Welch sigue siendo el mejor modelo.

La parte inferior del gráfico muestra la estimación para los procedimientos en que se incluyen los pre-test. En estos casos, se puede observar que los procedimientos que involucran en el proceso de pre-testeo a la prueba de equivalencia (ETW, SETW) siguen manteniéndose robustos al menos hasta una contaminación moderada de la normalidad, independientemente del grado de heterocedasticidad en comparación a los procedimientos que involucran en el proceso de pre-testeo a la prueba de tradicional F (FTW, SFTW), estas mantienen totalmente descontroladas a la TIEP por debajo del nivel de significancia en la mayoría de los casos representados. Finalmente, para contaminaciones Alta y Severa de la normalidad, no podemos definir ninguna técnica como robusta. Además, se puede observar que la TIEP global para muestras desbalanceadas es mayor cuando se compararon tamaños balanceados.

### 4.3 Discusión

Estudios iniciales mencionados en la introducción demostraron que en un escenario de perfecta normalidad realizar directamente una t-Student contrae graves alteraciones en la TIEP, estos resultados concuerdan con la presente investigación, donde aplicar directamente t-Student resultó ser el peor de todos los procedimientos para todos los escenarios de contaminación normal dado que, la alteración de TIEP comparada con un escenario de perfecta normalidad resulta estar más alejada del valor nominal  $\alpha = 0,05$  a medida que aumenta grado de heteroscedasticidad especialmente en tamaños desbalanceados y tamaños balanceados pequeños. No se recomienda en lo absolutos el uso directo de la prueba t-Student, hacerlo conlleva una elevada probabilidad de cometer un error.

En la mayoría de libros y cursos estadísticos recomiendan que al rechazar la hipótesis nula de normalidad, usar una prueba no paramétrica (Wilcoxon) es lo correcto, contrario a lo esperado los resultados mostraron que, aplicar directamente Wilcoxon para muestras alejadas de la normal, resultó ser el peor método, ya que se observó fuertes alteraciones de la TIEP especialmente en casos cuando existe un alto grado de heterocedasticidad tanta para muestras especialmente para muestras pequeñas y desbalanceadas. Bajo este criterio podemos definir a la prueba de Wilcoxon como un test poco útil para comparar medias de muestras provenientes de poblaciones donde no se puede asegurar el cumplimiento de la normalidad. Cabe recalcar que, en este escenario, el test de Welch, se muestra como una prueba mucho más robusta y digna de ser reemplazada por este procedimiento no paramétrico.

A diferencia de estudios pasados, en el presente trabajo se añadió alejamientos de la normalidad leve, moderada alta y severa, además se pre-testeo el supuesto de normalidad usando la prueba tradicional más potente Shapiro Wilk. Se demostró que los valores de la estimación de la TIEP son casi idénticos entre los procedimientos en que se probó únicamente el supuesto de homocedasticidad con los pre-test F y de equivalencia con aquellos que fueron anteceditas a la aceptación una prueba de normalidad utilizando el pret - test Shapiro Wilks en pruebas de comparación de medias, lo cual demostró que la prueba Shapiro Wilks no estuvo funcionando correctamente de tal manera que no detectó normalidad y no corrigió la estimación de la TIEP global, resultando ser inútil. Este resultado apoya la idea propuesta por (Box, 1979) el cual menciona que no tiene sentido usar modelos que sirven para contratar una hipótesis de normalidad

perfecta como se ha descrito tradicionalmente (dado que no existe). En este contexto el procedimiento que involucran en el proceso de pre-testeo a la prueba de equivalencia para verificar el supuesto de homocedasticidad sin ser antecedida a la aceptación una prueba de normalidad en pruebas de comparación de medias resultó ser el mejor procedimiento dado que controla la TIEP muy cerca del nivel nominal  $\alpha = 0.05$  en todos los tamaños, niveles de heterocedasticidad y en escenarios de perfecta normalidad, contaminación leve y severa; sin embargo las alteraciones de la TIEP empiezan hacerse notorias a partir de una contaminación alta y severa conforme se aumente el tamaño de muestra y grado de heterocedaasticidad. Estos resultados coinciden con los obtenidos por (Flores y Ocaña, 2018, pp. 64-70), los cuales mencionan que, en un escenario de normalidad perfecta el mejor pre-test para la verificación previa de homogeneidad es el test de equivalencia que al plantear una hipótesis con una zona de indiferencia adecuada la TIEP global se mantiene controlada en las pruebas de comparación de medias, también por lo obtenidos previamente por (Flores et al., 2018, pp. 10-14) enfocándose específicamente en el uso del pre - test de equivalencia donde exista desviaciones de la normalidad, la TIEP empieza a alejarse demasiado del nivel de significancia.

Mientras que en el caso donde se aplicó directamente Welch la estimación de la TIEP también fue similar al procedimiento en que se probó únicamente el supuesto de homocedasticidad con el pre-test de equivalencia, donde resultó ser los mejores métodos en un escenario de perfecta normalidad, contaminación leve y moderada. Resultados similares fueron los obtenidos por (Zimmerman, 2004; Rasch et al., 2011; Flores y Ocaña, 2018) los cuales demostraron que bajo el cumplimiento de normalidad perfecta es preferible usar directamente la prueba Welch.

Finalmente, aplicar directamente Welch sin necesidad de pre-testear los supuestos o únicamente probar el supuesto de homocedasticidad con el pre - test de equivalencia en pruebas de comparación de medias, según el criterio de Cochran son los mejores modelos ante escenarios donde se asegura normalidad y la contaminación normal leve y moderada. Al igual que en el estudio realizado por (Flores et al., 2018, p. 9), los resultados demuestran una vez más que la prueba de equivalencia sigue funcionando correctamente debido a que detecta heterocedasticidad en todos los casos y permite aplicar correctamente Welch . Por otro lado para contaminación alta y severa no se pudo decir que alguno de los 7 método fueron robustas, ya que ninguno demuestra controlar la TIEP alrededor de  $\alpha = 0.05$  y que al final en algún punto de heterocedasticidad caen fuera del intervalo de Cochran (0.04 – 0.06).

## CONCLUSIONES

- Los resultados de este estudio indicaron que, la prueba más potente ante escenarios de contaminación normal es Shapiro Wilks a diferencia de la de Kolmogorov Smirnov, Lilliefors, Anderson Darling, Jarke Bera y Cramer. Al ser esta la prueba más potente, se la utilizó para pre-testear el supuesto de normalidad.
- Cuando se utiliza directamente las pruebas para comparar medias de dos muestras en escenarios de no normalidad, la técnica más robusta es aplicar Welch directamente, la cual controla de mejor manera la TIEP alrededor del nivel de significancia, incluso mejor que la prueba no paramétrica de Wilcoxon, la cual a pesar que está diseñada para comparar medias en muestras no normales, provoca una fuerte alteración en la probabilidad global de cometer un error tipo I.
- Cuando se realiza el proceso de pre-testeo en escenarios cuando se asegura normalidad o a lo mucho una desviación moderada de la misma, la TIEP que quedan controladas son aquellas en cuyo procedimiento se incluye el pre-test de equivalencia, debido a que ésta supera la dificultad lógica del pre-test tradicional F, el cual presentó una TIEP con una afectación muy elevada y nada controlada.
- La evidencia surgida del estudio concluye que ninguno de los procedimientos planteados es confiable para una prueba de diferencia de medias poblacional en casos donde la contaminación normal sea alta o severa, debido que son muy deficientes en cuanto al control del error de tipo I.

## RECOMENDACIONES

De este estudio surgen una serie de planteamientos inconclusos, los cuales a nuestro parecer son importantes de investigar.

- La mejor prueba para pre-testear el supuesto de normalidad es la de Shapiro Wilks, sin embargo, tomando en cuenta el argumento de Box (no existe un conjunto de datos perfectamente normales por lo tanto no tiene sentido usar modelos que sirven para contratar una hipótesis de normalidad perfecta), se sugiere realizar un estudio donde se use un pre-test de equivalencia para probar este supuesto y saber si existe alguna mejora en la TIEP global en las pruebas de comparación de medias.
- Se observa definitivamente que pre-testear bajo el enfoque de equivalencia controla la TIEP de mejor manera que los procedimientos que incluyen pre-test bajo el enfoque tradicional. Nos parece importante desarrollar límites de irrelevancia para pre testear normalidad, similares a los límites de irrelevancia que existe para probar homocedasticidad.
- Recomendamos que para muestras no normales no se utilice prueba de Wilcoxon dado que tenemos una alta probabilidad de cometer un error de tipo I en su lugar el uso del test de Welch es el más adecuado.
- Pensamos que sería interesante agregar a este estudio la prueba de comparación de medias bajo el enfoque de equivalencia tanto si se pre-testean sus supuestos como si se aplica directamente.
- Finalmente se sugiere realizar una investigación generalizada, que consista en expandir lo estudiado a la comparación de dos o más medias poblacionales, usando pruebas previas para contrastar los supuestos de normalidad y homocedasticidad, y usando técnicas directas sin pre-test.

## BIBLIOGRAFÍA

**Althouse, L.A.W.B. Ferro. J.M.** "Detecting Departures from Normality: A Monte Carlo Simulation of a New Omnibus Test Based on Moments". *The Annual Meeting of the American Educational Research Association* [en línea], 1998, (California), pp. 1-31.[Consulta: 22 mayo 2018]. Disponible en: <https://eric.ed.gov/?id=ED422385>.

**Altman, D.G. y Bland, J.M.** "Statistics notes: Absence of evidence is not evidence of absence". *BMJ* [en línea], 1995, 311(7003), pp. 485. [Consulta: 25 enero 2019]. ISSN 0959-8138. DOI 10.1136/bmj.311.7003.485. Disponible en: <http://www.bmj.com/cgi/doi/10.1136/bmj.311.7003.485>.

**Blanca, M.J., et.al.** "Skewness and Kurtosis in Real Data Samples". *Methodology* [en línea], 2013, 9(2), pp. 78-84. [Consulta: 27 enero 2019]. ISSN 1614-1881. DOI 10.1027/1614-2241/a000057. Disponible en: <https://econtent.hogrefe.com/doi/10.1027/1614-2241/a000057>.

**Box, G.E.P.** "Robustness in the Strategy of Scientific Model Building". *Robustness in Statistics* [en línea], 1979, S.I.: Elsevier, pp. 201-236. [Consulta: 25 enero 2019]. Disponible en: <https://linkinghub.elsevier.com/retrieve/pii/B9780124381506500182>.

**Cochran, W.G.** "The  $\chi^2$  correction for continuity". *Iowa State College Journal of Science*, vol. 16, n°. 1 (1942), pp. 421-436.

**D'agostino, R.B., Belanger, A. y D'agostino, R.B.** "A Suggestion for Using Powerful and Informative Tests of Normality". *The American Statistician* [en línea], 1990, 44(4), pp. 316-321. [Consulta: 2 octubre 2018]. ISSN 0003-1305. DOI 10.1080/00031305.1990.10475751. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/00031305.1990.10475751>.

**Dhar, S.K.** A Review of: "*Medical Statistics at a Glance* , Third Edition, by A. Petrie and C. Sabin". *Journal of Biopharmaceutical Statistics* [en línea], 2010, 20(6), pp. 1223-1224. [Consulta: 26 enero 2019]. ISSN 1054-3406. DOI 10.1080/10543406.2010.514465. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/10543406.2010.514465>.

**Dufour, J., et.al.** Simulation- based finite sample normality tests in linear regressions. *The Econometrics Journal* [en línea], 1998, 1(1), pp. C154-C173. [Consulta: 19 agosto 2018]. ISSN 1368-4221. DOI 10.1111/1368-423X.11009. Disponible en: <https://academic.oup.com/ectj/article/1/1/C154-C173/5071718>.

**Easterling, R.G. y Anderson, H.E.** "The effect of preliminary normality goodness of fit tests on subsequent inference". *Journal of Statistical Computation and Simulation* [en línea], 1978, 8(1), pp. 1-11. [Consulta: 21 enero 2018]. ISSN 0094-9655. DOI 10.1080/00949657808810243. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/00949657808810243>.

**Fagerland, M.W.** "t-tests, non-parametric tests, and large studies—a paradox of statistical practice". *BMC Medical Research Methodology* [en línea], 2012, 12(1), pp. 78. [Consulta: 2 marzo 2018]. ISSN 1471-2288. DOI 10.1186/1471-2288-12-78. Disponible en: <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-78>.

**Farrell, P.J. y Rogers-Stewart, K.** "Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test". *Journal of Statistical Computation and Simulation* [en línea], 2006, 76(9), pp. 803-816. [Consulta: 7 noviembre 2018]. ISSN 0094-9655. DOI 10.1080/10629360500109023. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/10629360500109023>.

**Fleishman, A.I.** "A method for simulating non-normal distributions". *Psychometrika* [en línea], 1978, 43(4), pp. 521-532. [Consulta: 26 febrero 2018]. ISSN 0033-3123. DOI 10.1007/BF02293811. Disponible en: <http://link.springer.com/10.1007/BF02293811>.

**Flores Muñoz, P.**, 2017. Un pre test de irrelevancia de la diferencia de varianzas en la comparación de medias (tesis). (Maestría). Universitat Politècnica de Catalunya. [Consulta: 3 enero 2018]. Disponible en: <https://upcommons.upc.edu/handle/2117/105989>.

**Flores, P. y Ocaña, J.** "Heteroscedasticity irrelevance when testing means difference". *Sort* [en línea], 2018, 42(1), pp. 59-72. [Consulta: 16 enero 2018]. ISSN 20138830. DOI 10.2436/20.8080.02.69. Disponible en: <https://www.raco.cat/index.php/SORT/article/view/338210/429072>.

**Flores, P., Ocaña, J. y Sánchez, T.** "VERIFICACIÓN DE SUPUESTOS EN LAS PRUEBAS DE COMPARACIÓN DE MEDIAS. UNA REVISIÓN". *Ciencia Digital* [en línea], 2018, 2,(4.1), pp. 5-22. ISSN 2602-8085. DOI 10.33262/cienciadigital.v2i4.1..187. Disponible en: <http://cienciadigital.org/revistacienciadigital2/index.php/CienciaDigital/article/view/187>.

**Greene, H.W.** *Econometric Analysis*, New Jersey: Prentice-Hall. 2000.

**Hsu, P.L.** Contribution to the theory of "Student's" t-test as applied to the problem of two samples. *Statistical Research Memoirs* [en línea], 1938. [Consulta: 3 febrero 2018]. Disponible en: <https://psycnet.apa.org/record/1939-02268-001>.

**Inan, G. y Demirtas, H.** *Data generation with binary and continuous non-normal components*, R package *BinNonNor* [en línea]. 2015. [Consulta: 3 febrero 2018]. Disponible en: <https://cran.r-project.org/web/packages/BinNonNor/index.html>.

**Kim, J. y Cribbie, R.A.** "The variance homogeneity assumption and the traditional ANOVA: Exploring a better gatekeeper". [en línea], 2017. [Consulta: 25 enero 2019]. Disponible en: <https://yorkspace.library.yorku.ca/xmlui/handle/10315/33150>.

**Mara, C., Anderson, J.M. y Cribbie, R.A.** *Testing for equivalence of group variances (Unpublished doctoral dissertation)*. 2014. York University: Canada.

**Mendes, M. y Pala, A.** "Type I Error Rate and Power of Three Normality Tests". *Information Technology Journal* [en línea], 2003, 2(2), pp. 135-139. ISSN 18125638. DOI 10.3923/itj.2003.135.139. Disponible en: <http://www.scialert.net/abstract/?doi=itj.2003.135.139>.

**Moder, K., Rasch, D. y Kubinger, K.D.** Don't use the two-sample t-test anymore. *Proceedings of the "6th St. Petersburg Workshop on Simulation"*, Edited by SM Ermakov, VB Melas and AN Pepelyshev, pp. 258-264.

**Moser, B.K. y Stevens, G.R.** "Homogeneity of Variance in the Two-Sample Means Test". *The American Statistician* [en línea], 1992, 46(1), pp. 19-21. [Consulta: 21 agosto 2018]. ISSN 0003-1305. DOI 10.1080/00031305.1992.10475839. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475839>.

**Ocaña, J. y Vegas, E.** "Variance reduction for Bernoulli response variables in simulation". *Computational Statistics & Data Analysis* [en línea], 1995, 19(6), pp. 631-640. [Consulta: 27 enero 2019]. ISSN 0167-9473. DOI 10.1016/0167-9473(94)00023-C. Disponible en: <https://www.sciencedirect.com/science/article/pii/016794739400023C>.

**Overall, J.E., Atlas, R.S. y Gibson, J.M.** "Tests That are Robust against Variance Heterogeneity in  $k \times 2$  Designs with Unequal Cell Frequencies". *Psychological Reports* [en línea], 1995, 76(3), pp. 1011-1017. [Consulta: 25 enero 2019]. ISSN 0033-2941. DOI 10.2466/pr0.1995.76.3.1011. Disponible en: <http://journals.sagepub.com/doi/10.2466/pr0.1995.76.3.1011>.

**Öztuna, D., Elhan, A.H. y Tüccar, E.** "Investigation of four different normality tests in terms of type 1 error rate and power under different distributions". *Turkish Journal of Medical Sciences*, 2006, 36(3), pp. 171-176. ISSN 13000144.

**Park, H.M.** "Univariate Analysis and Normality Test Using SAS, Stata, and SPSS". [en línea],

2015. [Consulta: 27 mayo 2019]. Disponible en: <https://scholarworks.iu.edu/dspace/handle/2022/19742>.

**Quinn, G.P. Gerald P. y Keough, M.J.**, 2002. *Experimental design and data analysis for biologists* [en línea]. S.l.: Cambridge University Press. [Consulta: 2 febrero 2018]. ISBN 9780521009768. Disponible en: [https://books.google.es/books?hl=es&lr=&id=VtU3-y7LaLYC&oi=fnd&pg=PP17&dq=+Quinn,+G.,+%26+Keough,+M.+\(2002\).+Experimental+Design+and+Data+Analysis+for+Biologists.+Cambridge+University+Press.&ots=cAsr2wnmiz&sig=EkDYDPqxLBqXaX8Lzq\\_r2NnKj7E#v=onepage&q=Qui](https://books.google.es/books?hl=es&lr=&id=VtU3-y7LaLYC&oi=fnd&pg=PP17&dq=+Quinn,+G.,+%26+Keough,+M.+(2002).+Experimental+Design+and+Data+Analysis+for+Biologists.+Cambridge+University+Press.&ots=cAsr2wnmiz&sig=EkDYDPqxLBqXaX8Lzq_r2NnKj7E#v=onepage&q=Qui).

**Rasch, D., Kubinger, K.D. y Moder, K.** "The two-sample t test: pre-testing its assumptions does not pay off". *Statistical Papers* [en línea], 2011, 52(1), pp. 219-231. [Consulta: 3 mayo 2018]. ISSN 0932-5026. DOI 10.1007/s00362-009-0224-x. Disponible en: <http://link.springer.com/10.1007/s00362-009-0224-x>.

**Rochon, J. y Kieser, M.** "A closer look at the effect of preliminary goodness-of-fit testing for normality for the one-sample t-test". *British Journal of Mathematical and Statistical Psychology* [en línea], 2011, 64(3), pp. 410-426. [Consulta: 9 junio 2018]. ISSN 00071102. DOI 10.1348/2044-8317.002003. Disponible en: <http://doi.wiley.com/10.1348/2044-8317.002003>.

**Rothery, P.** "The Use of Control Variates in Monte Carlo Estimation of Power". *Applied Statistics* [en línea], 1982, 31(2), pp. 125-129. [Consulta: 2 julio 2018]. ISSN 00359254. DOI 10.2307/2347974. Disponible en: <https://www.jstor.org/stable/10.2307/2347974?origin=crossref>.

**Royston, J.P.** "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples". *Applied Statistics* [en línea], 1982, 31(2), pp. 115-124. [Consulta: 27 enero 2019]. ISSN 00359254. DOI 10.2307/2347973. Disponible en: <https://www.jstor.org/stable/10.2307/2347973?origin=crossref>.

**Royston, P.** Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing* [en línea], 1992, 2(3), pp. 117-119. [Consulta: 27 enero 2019]. ISSN 0960-3174. DOI 10.1007/BF01891203. Disponible en: <http://link.springer.com/10.1007/BF01891203>.

**Ryan, T.P. y Wiley I.** *Modern engineering statistics* [en línea]. S.l.: Wiley-Interscience. 2007. [Consulta: 22 enero 2018]. Disponible en: [https://books.google.es/books?hl=es&lr=&id=aZn7XNphKcgC&oi=fnd&pg=PR5&dq=Ryan,+T.+\(2007\).+Modern+Engineering+Statistics.&ots=KNUp\\_qiGva&sig=BhXm9gmki6TJ\\_LbGzGozlclZ0](https://books.google.es/books?hl=es&lr=&id=aZn7XNphKcgC&oi=fnd&pg=PR5&dq=Ryan,+T.+(2007).+Modern+Engineering+Statistics.&ots=KNUp_qiGva&sig=BhXm9gmki6TJ_LbGzGozlclZ0)

hg#v=onpage&q=Ryan%2C T. (2007). Modern Engineering Statistics.&f=false.

Scheffe, H. *The analyses of variance*. Wiley. New York: s.n, 1959.

**Schucany, W.R. y Tony NG, H.K.** "Preliminary Goodness-of-Fit Tests for Normality do not Validate the One-Sample Student t". *Communications in Statistics - Theory and Methods* [en línea], 2006, 35(12), pp. 2275-2286. [Consulta: 11 junio 2019]. ISSN 0361-0926. DOI 10.1080/03610920600853308. Disponible en: <https://www.tandfonline.com/doi/full/10.1080/03610920600853308>.

**Sedlmeier, P. y Gigerenzer, G.** "Do studies of statistical power have an effect on the power of studies". *Psychological bulletin*, vol. 105, n°. 2 (1989), pp. 309.

**Seier, E.** "Comparison of tests of univariate normality Comparison of Tests for Univariate Normality". [en línea], 2002. [Consulta: 27 mayo 2018]. Disponible en: <https://www.researchgate.net/publication/228965974>.

**Shapiro, S.S. y Wilk, M.B.** An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* [en línea], 1965, 52(3/4), pp. 591-611. [Consulta: 28 mayo 2018]. ISSN 00063444. DOI 10.2307/2333709. Disponible en: <https://www.jstor.org/stable/2333709?origin=crossref>.

**Tukey, J.W.** Graphic Comparisons of Several Linked Aspects: Alternatives and Suggested Principles. *Journal of Computational and Graphical Statistics* [en línea], 1993, 2(1), pp. 1-33. [Consulta: 27 enero 2019]. ISSN 1061-8600. DOI 10.1080/10618600.1993.10474595. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/10618600.1993.10474595>.

**Wackerly, D.D., Mendenhall, W. y L., S.R.** *Estadística matemática con aplicaciones. 7e.* [en línea]. S.l.: Cengage Learning Editores S.A. de C.V. 2009. [Consulta: 26 diciembre 2018]. ISBN 607481399X. Disponible en: [https://books.google.com.ec/books?id=Ho1KAQAACAAJ&dq=Wackerly,+R.+D.,+Mendenhall,+W.,+%26+Sheaffer,+R.+\(2008\).+Mathematical+statistics+with+applications.+California.:+Thomson+Brooks.+Cole,+Belmont.&hl=es&sa=X&ved=0ahUKEwjO99KEh43gAhXNmuAKHb4JD XIQ6AEINDAB](https://books.google.com.ec/books?id=Ho1KAQAACAAJ&dq=Wackerly,+R.+D.,+Mendenhall,+W.,+%26+Sheaffer,+R.+(2008).+Mathematical+statistics+with+applications.+California.:+Thomson+Brooks.+Cole,+Belmont.&hl=es&sa=X&ved=0ahUKEwjO99KEh43gAhXNmuAKHb4JD XIQ6AEINDAB).

**Welch, B.L.** "The Generalization of 'Student's' Problem when Several Different Population Variances are Involved". *Biometrika* [en línea], 1947, 34(1/2), pp. 28-35. [Consulta: 14 enero 2018]. ISSN 00063444. DOI 10.2307/2332510. Disponible en: <https://www.jstor.org/stable/2332510?origin=crossref>.

**Wellek, S.** *Testing statistical hypotheses of equivalence and noninferiority* [en línea]. S.l.: CRC Press, 2010. [Consulta: 25 enero 2018]. ISBN 9781439808184. Disponible en: <https://www.crcpress.com/Testing-Statistical-Hypotheses-of-Equivalence-and-Noninferiority/Wellek/p/book/9781439808184>.

**Wilcoxon, F.** "Individual comparisons of grouped data by ranking methods". *Journal of economic entomology* [en línea], 1946, 39(6) , pp. 219. [Consulta: 10 enero 2018]. ISSN 00220493. DOI 10.1093/jee/39.2.269. Disponible en: <https://academic.oup.com/jee/article-abstract/39/2/269/2203647>.

**Yap, B.W. y Sim, C.H.** "Comparisons of various types of normality tests". *Journal of Statistical Computation and Simulation* [en línea], 2011, 81(12), pp. 2141-2155. [Consulta: 22 junio 2018]. ISSN 0094-9655. DOI 10.1080/00949655.2010.520163. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/00949655.2010.520163>.

**Yazici, B. y Yolacan, S.** "A comparison of various tests of normality". *Journal of Statistical Computation and Simulation* [en línea], 2007, 77(2), pp. 175-183. [Consulta: 11 enero 2018]. ISSN 0094-9655. DOI 10.1080/10629360600678310. Disponible en: <http://www.tandfonline.com/doi/abs/10.1080/10629360600678310>.

**Zimmerman, D.W.** "A note on preliminary tests of equality of variances". *British Journal of Mathematical and Statistical Psychology* [en línea], 2004, 57(1), pp. 173-181. [Consulta: 27 marzo 2018]. ISSN 00071102. DOI 10.1348/000711004849222. Disponible en: <http://doi.wiley.com/10.1348/000711004849222>.

## ANEXOS

**Anexo A :** Tablas de resultados de la Estimación de la Robustez de las Pruebas de Normalidad

Niveles de Contaminación	Tamaño (n)	Estimación de la Robustez usando $\alpha=0.05$						
		K.Smirnov	Liliefors	Sh.Wilk	A.Darling	Jarke	Pearson	Cramer
Sin Contaminación	5	0.04922	0.04480	0.04702	NA	0.04974	NA	NA
	10	0.04979	0.04959	0.04872	0.04958	0.04877	0.06638	0.04897
	15	0.04949	0.04874	0.04945	0.05056	0.04981	0.05191	0.05072
	20	0.04978	0.04945	0.05205	0.05157	0.05008	0.04985	0.05098
	30	0.04937	0.04837	0.05177	0.05038	0.04962	0.05267	0.05061

**Anexo B:** Tablas de resultados de Estimación de la Potencia de las Pruebas de Normalidad

Niveles de Contaminación	Tamaño (n)	Estimación de la Potencia usando $\alpha=0.05$						
		K.Smirnov	Liliefors	Sh.Wilk	A.Darling	Jarke	Pearson	Cramer
Moderada	5	0.05121	0.05559	0.05899	NA	0.06270	NA	NA
	10	0.05659	0.08267	0.09802	0.09453	0.09679	0.09363	0.08870
	15	0.06220	0.09917	0.13885	0.12575	0.12896	0.08110	0.11629
	20	0.06749	0.11989	0.18186	0.15715	0.16240	0.08685	0.14382
	30	0.07719	0.15780	0.26442	0.21922	0.22527	0.10659	0.19655
Alta	5	0.07136	0.09025	0.11378	NA	0.10775	NA	NA
	10	0.09084	0.18510	0.29215	0.26234	0.18227	0.26501	0.23705
	15	0.11214	0.26673	0.47732	0.41490	0.26478	0.29165	0.36778
	20	0.13591	0.35468	0.65053	0.55858	0.34756	0.42669	0.49219
	30	0.17830	0.52771	0.87371	0.77690	0.50617	0.63108	0.69546

Severa	5	0.07856	0.13442	0.17780	NA	0.16410	NA	NA
	10	0.11924	0.31213	0.48230	0.44205	0.29668	0.44110	0.40870
	15	0.15928	0.46493	0.71661	0.65596	0.43725	0.54406	0.60649
	20	0.20136	0.60117	0.86981	0.80899	0.56371	0.72381	0.75533
	30	0.28554	0.81471	0.97979	0.95163	0.76300	0.90423	0.91818

**Anexo C : Tabla de resultados de la estimación de la Probabilidad de Error de Tipo I en las Pruebas de Comparación de Medias**

Niveles de contaminación	Estimación de la TIEP usando $\alpha=0.05$								
	Tamaño (n <sub>1</sub> ,n <sub>2</sub> )	Heteroced	Directo			Pre-Test Homocedasticidad		Pre-Test Normalidad & Homocedasticidad	
			t-Sstudent	Welch	Wilcox	F	Equiv	Shap-F	Shap-Equiv
Sin Contaminación	(5,5)	1	0.05000000	0.04399268	0.03101155	0.04875991	0.04400285	0.04872890	0.04397183
		1.2	0.05046316	0.04458110	0.03206672	0.04934886	0.04460126	0.04928787	0.04455026
		1.4	0.05186442	0.04552188	0.03325807	0.05042589	0.04558218	0.05036490	0.04552119
		1.6	0.05422029	0.04718557	0.03525735	0.05254815	0.04723553	0.05247734	0.04716472
		1.8	0.05625580	0.04829095	0.03714448	0.05411677	0.04833109	0.05403597	0.04828061
		2	0.05883777	0.04974885	0.03880833	0.05616296	0.04977883	0.05607199	0.04970819
		2.5	0.06347476	0.05213291	0.04328799	0.05911475	0.05213291	0.05906462	0.05210294
		3	0.06743036	0.05367676	0.04682486	0.06107835	0.05367676	0.06101840	0.05362680
	(10,10)	1	0.05000000	0.04835709	0.04288367	0.04953484	0.04931710	0.04959530	0.04938747
		1.2	0.05028086	0.04860132	0.04316683	0.04984474	0.04968563	0.04985518	0.04970607
		1.4	0.05126780	0.04903153	0.04480127	0.05049399	0.05017642	0.05051433	0.05019675
		1.6	0.05246739	0.04935519	0.04692096	0.05101819	0.05062079	0.05103853	0.05064113
		1.8	0.05335279	0.04954379	0.04898747	0.05114871	0.05054087	0.05115904	0.05055120
		2	0.05438221	0.04984551	0.05118048	0.05127196	0.05066412	0.05128229	0.05067445

		2.5	0.05622919	0.05010531	0.05588798	0.05110315	0.05059471	0.05119309	0.05070456
		3	0.05835224	0.05010120	0.06007739	0.05084085	0.05039136	0.05098071	0.05053122
	<b>(5,10)</b>	1	0.05000000	0.05036834	0.03871816	0.05082359	0.05098165	0.05083331	0.05101164
		1.2	0.06407109	0.05228471	0.04679119	0.06236688	0.05283472	0.06238671	0.05288471
		1.4	0.07967374	0.05379011	0.05395926	0.07406578	0.05432007	0.07405562	0.05437006
		1.6	0.09339675	0.05477971	0.06029523	0.08250384	0.05515960	0.08251384	0.05522964
		1.8	0.10602071	0.05508094	0.06585971	0.08858382	0.05541084	0.08858394	0.05551093
		2	0.11739604	0.05569551	0.07006314	0.09188420	0.05603541	0.09187426	0.05612544
		2.5	0.14010456	0.05555309	0.07681059	0.09134035	0.05579308	0.09137040	0.05590310
		3	0.15656662	0.05486505	0.08107932	0.08406361	0.05497502	0.08410360	0.05509499
		<b>(10,5)</b>	1	0.05000000	0.05063643	0.03957156	0.05091262	0.05104459	0.05091197
	1.2		0.03837187	0.04918629	0.03406422	0.04115711	0.04946137	0.04114647	0.04948135
	1.4		0.03118510	0.04804866	0.02957914	0.03558558	0.04825167	0.03561555	0.04828164
	1.6		0.02667416	0.04739829	0.02662587	0.03214778	0.04760066	0.03215756	0.04756983
	1.8		0.02339028	0.04700755	0.02519613	0.03018378	0.04718887	0.03018357	0.04713785
	2		0.02096440	0.04695722	0.02401215	0.02947252	0.04703821	0.02946210	0.04699718
	2.5		0.01751060	0.04733669	0.02254559	0.03046195	0.04739706	0.03049192	0.04736623
	3		0.01576153	0.04854687	0.02244789	0.03321068	0.04854687	0.03325064	0.04851626
	<b>Contaminación Leve</b>		1	0.04820236	0.04182501	0.03101155	0.04687079	0.04186549	0.04684994
		1.2	0.04904981	0.04222208	0.03192631	0.04768122	0.04223208	0.04762006	0.04218091
		1.4	0.05020284	0.04332279	0.03302688	0.04853069	0.04334294	0.04850984	0.04332209
		1.6	0.05232856	0.04481130	0.03509749	0.05036233	0.04482129	0.05033132	0.04479028

	<b>(5,5)</b>	1.8	0.05463505	0.04615255	0.03687332	0.05225328	0.04616254	0.05220245	0.04613187
		2	0.05653968	0.04704798	0.03858095	0.05358129	0.04704798	0.05355062	0.04702748
		2.5	0.06125470	0.04933057	0.04278841	0.05650432	0.04935055	0.05645402	0.04932040
		3	0.06461803	0.05077450	0.04606430	0.05770545	0.05077450	0.05766532	0.05074435
	<b>(10,10)</b>	1	0.04903266	0.04732548	0.04288367	0.04852726	0.04830930	0.04863742	0.04841947
		1.2	0.04952673	0.04767829	0.04297586	0.04888190	0.04859381	0.04891225	0.04862415
		1.4	0.05042124	0.04803662	0.04498126	0.04937901	0.04889265	0.04944916	0.04896280
		1.6	0.05150152	0.04820171	0.04673858	0.04964502	0.04917780	0.04976509	0.04929787
		1.8	0.05237529	0.04833833	0.04876518	0.04987180	0.04929452	0.05003167	0.04946429
		2	0.05337599	0.04890695	0.05115776	0.05027369	0.04977548	0.05047337	0.04997515
		2.5	0.05569934	0.04940525	0.05600674	0.05046377	0.04996469	0.05078330	0.05028422
	3	0.05785142	0.04970261	0.06035451	0.05044193	0.05003236	0.05082149	0.05041192	
	<b>(5,10)</b>	1	0.04927792	0.04863341	0.03871816	0.04970968	0.04910573	0.04975944	0.04918576
		1.2	0.06306006	0.04988825	0.04658130	0.06000989	0.05034903	0.06006971	0.05042906
		1.4	0.07783363	0.05111647	0.05306912	0.07050971	0.05164655	0.07051959	0.05173663
		1.6	0.09116710	0.05212589	0.05971523	0.07811806	0.05253583	0.07817798	0.05267591
		1.8	0.10339145	0.05289842	0.06454020	0.08367905	0.05327831	0.08376902	0.05346837
		2	0.11476593	0.05310173	0.06887451	0.08705151	0.05343163	0.08715154	0.05365169
		2.5	0.13735523	0.05352989	0.07603093	0.08763968	0.05375983	0.08780969	0.05404986
		3	0.15477690	0.05293057	0.08020112	0.08256147	0.05310053	0.08270154	0.05334063
		1	0.04919273	0.04855478	0.03957156	0.04955433	0.04901289	0.04961426	0.04909366
		1.2	0.03792967	0.04707124	0.03411459	0.04084712	0.04747854	0.04093788	0.04753891
		1.4	0.03117468	0.04649688	0.02998688	0.03582919	0.04678130	0.03593016	0.04684146

	<b>(10,5)</b>	1.6	0.02650090	0.04588690	0.02727786	0.03225767	0.04613094	0.03231846	0.04613072
		1.8	0.02323936	0.04552870	0.02547164	0.03040977	0.04572108	0.03047035	0.04570024
		2	0.02078351	0.04550350	0.02457466	0.02952333	0.04557428	0.02958412	0.04557407
		2.5	0.01684271	0.04616277	0.02334194	0.03014627	0.04621293	0.03022640	0.04624226
		3	0.01554198	0.04724857	0.02282823	0.03289629	0.04728874	0.03298641	0.04730829
<b>Contaminación Moderada</b>	<b>(5,5)</b>	1	0.04792100	0.04056466	0.03101155	0.04595182	0.04058481	0.04598163	0.04065527
		1.2	0.04883278	0.04148169	0.03194249	0.04691703	0.04148169	0.04692667	0.04152165
		1.4	0.05109336	0.04344258	0.03406686	0.04913868	0.04347255	0.04915866	0.04351286
		1.6	0.05443290	0.04623137	0.03655174	0.05216467	0.04624136	0.05216467	0.04626152
		1.8	0.05763192	0.04929051	0.03865487	0.05527048	0.04931049	0.05527065	0.04932083
		2	0.06061372	0.05144146	0.04085895	0.05765672	0.05145145	0.05764690	0.05146179
		2.5	0.06716636	0.05568174	0.04581064	0.06260773	0.05571171	0.06258809	0.05571206
	3	0.07249166	0.05867348	0.05032980	0.06628490	0.05867348	0.06624528	0.05864385	
	<b>(10,10)</b>	1	0.04890910	0.04669905	0.04288367	0.04815248	0.04777401	0.04853150	0.04815302
		1.2	0.04944085	0.04692626	0.04364720	0.04843669	0.04803864	0.04882582	0.04842777
		1.4	0.05078463	0.04804165	0.04682682	0.04925614	0.04889779	0.04966550	0.04930715
		1.6	0.05291071	0.04975795	0.05031379	0.05111316	0.05078504	0.05155264	0.05122452
		1.8	0.05474078	0.05109981	0.05373039	0.05249580	0.05204718	0.05293550	0.05249689
		2	0.05670865	0.05244962	0.05688523	0.05371684	0.05326822	0.05426660	0.05382799
		2.5	0.06092162	0.05453651	0.06437512	0.05583538	0.05526626	0.05656535	0.05599623
3	0.06344370	0.05657594	0.07032744	0.05734549	0.05691580	0.05818552	0.05775583		
		1	0.04858628	0.05025100	0.03871816	0.04860422	0.05054218	0.04878434	0.05068219

	(5,10)	1.2	0.06267302	0.05409796	0.04766094	0.05861720	0.05467860	0.05881703	0.05485855
		1.4	0.07827973	0.05707500	0.05566798	0.07041271	0.05770511	0.07061265	0.05793521
		1.6	0.09245365	0.05937675	0.06275179	0.07935445	0.05987666	0.07951440	0.06008672
		1.8	0.10561745	0.06026786	0.06872615	0.08682649	0.06083776	0.08698633	0.06105775
		2	0.11810190	0.06160206	0.07347012	0.09269065	0.06200194	0.09287049	0.06225193
		2.5	0.14345932	0.06302199	0.08241592	0.09875064	0.06339189	0.09893070	0.06368198
		3	0.16060987	0.06280537	0.08747608	0.09537463	0.06299531	0.09554475	0.06336538
	(10,5)	1	0.04855456	0.05050187	0.03957156	0.04875623	0.05088726	0.04898792	0.05109897
		1.2	0.03836119	0.04721733	0.03382631	0.04176286	0.04763398	0.04197414	0.04780445
		1.4	0.03231856	0.04583373	0.03000042	0.03783870	0.04608712	0.03796985	0.04617745
		1.6	0.02822705	0.04509260	0.02761064	0.03536836	0.04528497	0.03548930	0.04537531
		1.8	0.02542191	0.04489016	0.02672796	0.03351026	0.04506170	0.03365118	0.04513162
		2	0.02355532	0.04555089	0.02591670	0.03298138	0.04565186	0.03309191	0.04569139
		2.5	0.02131808	0.04778452	0.02600763	0.03377081	0.04794586	0.03391131	0.04803555
3	0.01992293	0.04980654	0.02657900	0.03548850	0.04986691	0.03559860	0.04996680		
<b>Contaminación Alta</b>	(5,5)	1	0.04480191	0.03325944	0.03093260	0.03905950	0.03325944	0.03913961	0.03339037
		1.2	0.04667760	0.03584862	0.03428962	0.04157972	0.03588894	0.04165948	0.03599919
		1.4	0.05211298	0.04144160	0.03896882	0.04711680	0.04144160	0.04721689	0.04156202
		1.6	0.05888017	0.04894463	0.04359927	0.05458929	0.04894463	0.05455879	0.04898477
		1.8	0.06615352	0.05671778	0.04777184	0.06239478	0.05671778	0.06229365	0.05670709
		2	0.07236915	0.06287606	0.05174632	0.06868035	0.06289604	0.06846914	0.06279509
		2.5	0.08535705	0.07552373	0.05913714	0.08150613	0.07553372	0.08119501	0.07525257

		3	0.09431861	0.08329457	0.06452798	0.08948420	0.08329457	0.08910331	0.08295382
	<b>(10,10)</b>	1	0.04725173	0.04332713	0.04276965	0.04462134	0.04416304	0.04641170	0.04596330
		1.2	0.04874512	0.04452317	0.04834426	0.04598693	0.04549883	0.04851030	0.04806200
		1.4	0.05208666	0.04872281	0.05825005	0.04990912	0.04941091	0.05360475	0.05314645
		1.6	0.05648435	0.05296485	0.06771278	0.05393202	0.05357334	0.05878001	0.05843123
		1.8	0.06085039	0.05707324	0.07617063	0.05817016	0.05789174	0.06382968	0.06355126
		2	0.06454628	0.06074804	0.08319498	0.06176589	0.06145691	0.06849643	0.06820735
		2.5	0.07209056	0.06695527	0.09683795	0.06812353	0.06762445	0.07624483	0.07575565
		3	0.07640896	0.06995251	0.10710657	0.07109206	0.07057253	0.08067416	0.08020455
	<b>(5,10)</b>	1	0.04549030	0.05326037	0.03875101	0.04845534	0.05343112	0.04929767	0.05415298
		1.2	0.06010935	0.06361040	0.05470288	0.05665073	0.06382086	0.05817258	0.06505256
		1.4	0.07707716	0.07226137	0.06877694	0.06682486	0.07262195	0.06881567	0.07421333
		1.6	0.09476527	0.07873473	0.07916901	0.07899495	0.07912485	0.08097535	0.08081580
		1.8	0.11055998	0.08321837	0.08794017	0.09058354	0.08375856	0.09227374	0.08541946
		2	0.12584375	0.08656370	0.09455412	0.10190369	0.08726368	0.10315396	0.08882455
		2.5	0.15556096	0.09070578	0.10505268	0.12085510	0.09113572	0.12097575	0.09229682
	3	0.17572129	0.09188093	0.10950324	0.12748322	0.09209087	0.12730401	0.09304192	
	<b>(10,5)</b>	1	0.04637090	0.05289523	0.03947524	0.04909239	0.05300662	0.04972881	0.05351167
		1.2	0.03831986	0.04499236	0.03277500	0.04503821	0.04528656	0.04553134	0.04575929
		1.4	0.03473126	0.04082655	0.03164594	0.04265248	0.04100829	0.04316474	0.04153054
		1.6	0.03324311	0.03967306	0.03266448	0.04164294	0.03990497	0.04223447	0.04050627
		1.8	0.03254002	0.04091285	0.03431202	0.04120667	0.04121597	0.04186683	0.04187485
		2	0.03238097	0.04339563	0.03590371	0.04178393	0.04373678	0.04255398	0.04453615

		2.5	0.03231798	0.05045931	0.03972887	0.04193525	0.05070013	0.04291486	0.05156815	
		3	0.03224648)	0.05712008)	0.04252530)	0.04393199)	0.05730053)	0.04493073)	0.05812794)	
<b>Contaminación Severa</b>	<b>(5,5)</b>	1	0.03861729	0.02586320	0.03098221	0.03079343	0.02586320	0.03107700	0.02620759	
		1.2	0.04226776	0.02946279	0.03666361	0.03452389	0.02946279	0.03480590	0.02979562	
		1.4	0.05002520	0.03762455	0.04405391	0.04289264	0.03762455	0.04312383	0.03789622	
		1.6	0.06027398	0.04768256	0.05118171	0.05374391	0.04770254	0.05399474	0.04806431	
		1.8	0.07052710	0.05839603	0.05764736	0.06479397	0.05839603	0.06490422	0.05864668	
		2	0.08007994	0.06873825	0.06286199	0.07498147	0.06876840	0.07490085	0.06882834	
		2.5	0.09878510	0.08834304	0.07308300	0.09452816	0.08836302	0.09416743	0.08806241	
		3	0.11187180	0.10051931	0.07942043	0.10703519	0.10053929	0.10626429	0.09996857	
		<b>(10,10)</b>	1	0.04336399	0.03755223	0.04279345	0.03860642	0.03807884	0.04376251	0.04327463
	1.2		0.04577268	0.04029698	0.05488640	0.04152213	0.04090428	0.04983829	0.04926004	
	1.4		0.05191872	0.04695919	0.07305137	0.04811690	0.04769721	0.06107384	0.06070418	
	1.6		0.05876731	0.05431589	0.08946769	0.05543434	0.05497495	0.07218758	0.07190808	
	1.8		0.06547008	0.06106107	0.10334526	0.06220834	0.06186935	0.08160606	0.08128708	
	2		0.07167225	0.06760234	0.11500955	0.06858017	0.06827098	0.09055017	0.09029100	
	2.5		0.08335382	0.07900072	0.13592389	0.08021976	0.07993993	0.10615399	0.10591407	
	3		0.09134762	0.08582453	0.14963833	0.08702367	0.08650414	0.11562022	0.11518052	
			1	0.04220243	0.04882929	0.03863939	0.04701992	0.04892966	0.04852503	0.05021386
	1.2		0.05464170	0.06511494	0.06174796	0.05516017	0.06525530	0.05910528	0.06842892	
	1.4		0.07201482	0.07918831	0.08021797	0.06551512	0.07931845	0.07146880	0.08378090	
	1.6		0.09174761	0.09046237	0.09514807	0.07863070	0.09080261	0.08514285	0.09593453	

	<b>(5,10)</b>	1.8	0.11076392	0.09851756	0.10676094	0.09304386	0.09900765	0.09884519	0.10395905
		2	0.1283490	0.1046734	0.1149535	0.1071987	0.1052235	0.1118791	0.1097349
		2.5	0.1639851	0.1133763	0.1270506	0.1339943	0.1139761	0.1352146	0.1168973
		3	0.1881254	0.1170213	0.1330110	0.1465341	0.1173412	0.1457340	0.1192023
	<b>(10,5)</b>	1	0.04225965	0.04889383	0.03954130	0.04676743	0.04898545	0.04835702	0.05040179
		1.2	0.03812175	0.03820951	0.03363347	0.04242954	0.03830113	0.04356951	0.03952188
		1.4	0.03738743	0.03364252	0.03677140	0.04084769	0.03392502	0.04232408	0.03552386
		1.6	0.03867171	0.03370797	0.04171355	0.04182852	0.03394030	0.04379203	0.03617697
		1.8	0.03994352	0.03660632	0.04649874	0.04353502	0.03686712	0.04600865	0.03967319
		2	0.04086763	0.04063478	0.05055707	0.04528471	0.04093575	0.04819594	0.04428909
		2.5	0.04389706	0.05270305	0.05814643	0.05037617	0.05299425	0.05353714	0.05648487
		3	0.04596738	0.06400256	0.06289565	0.05490047	0.06412287	0.05796026	0.06720113

## FUNCIONES

### Anexo D: Función power.norm

La función “*power.norm*” implementa un algoritmo que determina la estimación de la POTENCIA de 7 pruebas para determinar normalidad. Los argumentos que necesita la función son: el número de simulaciones (m), el número de observaciones para cada muestra generada (n), el nivel de significancia  $\alpha$  (alpha), los coeficientes de fleishman para la generación distribuciones normales y no normales (coef.fleishman) y la semilla de simulación (semilla). La mayoría de estos argumentos tienen valores por defecto que son comunes y que se ejecutan en caso de que el usuario no los especifique.

```
# rechazo_pval

rechazo_pval <- function(pval.stadic, alpha){
  ifelse (pval.stadic < alpha, 1, 0)
}

library(nortest)
library(normtest)

power.norm <- function(m = 100000, n, alpha, coef.fleishman, semilla = 127){
  old.seed <- .Random.seed
  set.seed(semilla)
  rejections <- replicate(m,
    {
      x1Control <- rnorm(n)
      x1 <- (coef.fleishman[1] + (coef.fleishman[2]*x1Control) +
        (coef.fleishman[3]* x1Control^2) + (coef.fleishman[4]*x1Control^3))

      k.test.pval <- ks.test(x1, "pnorm")$p
      kl.test.pal <- lillie.test(x1)$p
      sh.test.pval <- shapiro.test(x1)$p
      Ad.test.pval <- ifelse(n > 7, (ad.test(x1))$p, 0)
      JB.test.pval <- (ajb.norm.test(x1))$p
      DA.test.pval <- ifelse(n > 7, (pearson.test(x1))$p, 0)
      CVM.test.pval <- ifelse(n > 7, (cvm.test(x1))$p, 0)

      c(rechazo_pval(pval.stadic = k.test.pval, alpha = alpha),
        rechazo_pval(pval.stadic = kl.test.pal, alpha = alpha),
        rechazo_pval(pval.stadic = sh.test.pval, alpha = alpha),
        rechazo_pval(pval.stadic = Ad.test.pval, alpha = alpha),
        rechazo_pval(pval.stadic = JB.test.pval, alpha = alpha),
        rechazo_pval(pval.stadic = DA.test.pval, alpha = alpha),
        rechazo_pval(pval.stadic = CVM.test.pval, alpha = alpha)))

      pow <- rowSums(rejections)/m
      names(pow) <- c("K.Smirnov", "Liliefors", "Sh.Wilk", "A.Darling", "Jarke",
        "Pearson", "Cramer")
    }
  )
}
```

```
return(pow)
}
```

...

## Anexo E: Funciones auxiliares para TIEP.norm

### # disp.eq

```
disp.eq <- function(alpha = 0.05, n1, n2, omegaratio1,
                    omegaratio2 = 1/omegaratio1, tol = 1.0e-10, itmax = 50){
  err <- -alpha
  c1 <- sqrt(omegaratio1 * omegaratio2)

  while (err < 0){
    c1 <- c1 - 0.05
    h <- alpha + pf(c1/omegaratio2, n1, n2)
    c2 <- qf(h, n1, n2) * omegaratio2
    err <- pf(c2/omegaratio1, n1, n2) - pf(c1/omegaratio1, n1, n2) - alpha
  }

  c1L <- c1
  c1R <- c1 + 0.05
  it <- 0

  while (abs(err) >= tol && it <= itmax){
    it <- it + 1
    c1 <- (c1L + c1R) / 2
    h <- alpha + pf(c1/omegaratio2, n1, n2)
    c2 <- qf(h, n1, n2) * omegaratio2
    err <- pf(c2/omegaratio1, n1, n2) - pf(c1/omegaratio1, n1, n2) - alpha
    if (err <= 0)
      c1R <- c1 else
      c1L <- c1
  }
  c(c1, c2)
}
```

### # t.val

```
t.val <- function(x, y, n1=length(x), n2=length(y)){
  sp <- sqrt((var(x)*(n1-1)+var(y)*(n2-1))/(n1+n2-2))
  statistic <- (mean(x) - mean(y))/( sp * sqrt((1/n1) + (1/n2)))
  return(abs(statistic))
}
```

### # wel.val

```
wel.val <- function(x, y, n1 = length(x), n2 = length(y)){
  statistic <- (mean(x) - mean(y))/sqrt((var(x)/n1) + (var(y)/n2))
  v <- (var(x)/n1+var(y)/n2)^2/ (((var(x)/n1)^2/(n1-1))+((var(y)/n2)^2/(n2-1)))
  return(abs(statistic))
}
```

### # val.in

```
val.in <- function(q, int){
  (q > min(int) & q < max(int))
}
```

```

# estimControlTIEP

estimControlTIEP <- function(freqs, probControl) {
  m.j <- colSums(freqs)
  m <- sum(freqs)
  ptilde <- freqs *
  matrix(c((1 - probControl) / m.j[1], probControl / m.j[2]), ncol = 2,
         nrow = 2, byrow = TRUE)

  ptilde[is.na(ptilde)] <- 0
  result <- c( sum(ptilde[2,]), sqrt(
  ptilde[1,1] * ptilde[2,1] / ((m + 1) * (1 - probControl) - 2) +
  ptilde[1,2] * ptilde[2,2] / ((m + 1) * probControl - 2)

  )
  )
  names(result) <- c("estim.control", "se")
  return(result)
}

# p.val del Pre-test Wilcoxon

p.val<- function(x, y, n1 =length(x),n2 = length(y),N = n1+n2){
  sum.rank.x <- sum(rank(c(x,y))[1:n1])
  U1 <- sum.rank.x - (n1*(n1+1)) / 2 ;U1
  sum.rank.y <- sum(rank(c(x,y))[(n1+1):(N)])
  U2 <- sum.rank.y - (n2*(n2+1)) / 2;U2
  W<-n1*n2-U2
  p<-pwilcox(min(U1,U2), n1, n2)*2
}

# rechazo_pval

rechazo_pval<- function(pval.stadic,alpha ){
  ifelse (pval.stadic <= alpha ,T,F )
}

# rechazos_sin_pretest

rechazos_sin_pretest <-function(stad, valcrit){
  ifelse(stad > valcrit ,T, F)}

# rechazo_pretest_hom

rechazo_pretest_hom<- function( stad_hom , valcrit_hom , t.stud, tc , stat.welch , wc ){
  ifelse(val.in(stad_hom, valcrit_hom),rechazos_sin_pretest(stad = t.stud, valcrit= tc),
  rechazos_sin_pretest(stad = stat.welch, valcrit= wc)) }

# rechazo_pretest_norm_hom

rechazo_pretest_norm_hom<- function(alphaPretest, p.shap_x1 , p.shap_x2 , stad_hom,
valcrit_hom , t.stud , tc, stat.welch , wc , p.wilcox){
  ifelse((p.shap_x1 > alphaPretest | p.shap_x2 > alphaPretest ),rechazo_pretest_hom(
  stad_hom, valcrit_hom , t.stud, tc , stat.welch , wc ),
  rechazo_pval(p.wilcox,alphaPretest))}

```

```
# Coeficientes de Fleishman
```

```
library(BinNonNor)  
coef <- fleishman.coef(n.NN = 1, skewness.vec = 0, kurtosis.vec = 0)  
coef <- as.vector(coef)
```

## Anexo F: TIEP.norm

La función “**TIEP.norm**” implementa un algoritmo que estima la TIEP en siete casos comparables, los tres primeros cuando se aplica directamente la prueba t-Student ,Welch y Wilcoxon sin realizar ninguna prueba previa para determinar normalidad y homocedasticidad ,el tercero y cuarto cuando se realiza solamente el pre-test de igualdad de varianzas mediante las pruebas F y de equivalencia de dispersión antes de decidir entre la prueba t-Student o Welch , los dos últimos cuando se pretestea el supuesto de normalidad con la prueba Shapiro Wilk antes de decidir entre la prueba Wilcoxon y pretestear el supuesto de homogeneidad para decidir entre t-Student y Welch . Los argumentos que contiene TIEP.norm son : el número de simulaciones (m), la semilla de simulación (semilla), número de observaciones para las dos muestras generadas ( $n_1, n_2$ ), el valor de la diferencia entre medias teóricas en la hipótesis nula (diffmean), ratio entre  $\sigma_1$  y  $\sigma_2$  (ratioSigmas) que determina el grado de heterocedasticidad, el nivel de significancia  $\alpha$  (nominalAlpha) , los límites de equivalencia  $\omega_1^2$  y  $\omega_2^2$ (omegaratio1, omegaratio2) que se determinan mediante la función “limvarRatios “ ya implementada en el Anexo E, el nivel de significancia  $\alpha$  (alphaPretest) para los pre-test y los coeficientes de fleishman para la generación distribuciones normales y no normales (coef.fleishman). La mayoría de estos argumentos tienen valores por defecto que son comunes y que se ejecutan en caso de que el usuario no los especifique.

```
TIEP.norm<- function(m = 100000, semilla = 127, n1 = 5, n2 = 5, diffmean = 0,  
                    ratioSigmas = 1, nominalAlpha = 0.05, omegaratio1 = 0.75^2,  
                    omegaratio2 = 1/omegaratio1, alphaPretest = 0.05,  
                    coef.fleishman){  
  
  old.seed <- .Random.seed  
  set.seed(semilla)  
  int <- disp.eq(alpha = alphaPretest, n1 = n1, n2 = n2,  
               omegaratio1 = omegaratio1 , omegaratio2 = omegaratio2 )  
  N <- n1 + n2  
  tc <- qt(nominalAlpha/2, n1 + n2 - 2, lower.tail = F)  
  fvar1 <- qf(alphaPretest/2, n1 - 1, n2 - 1, lower.tail = T)  
  fvar2 <- qf(alphaPretest/2, n1 - 1, n2 - 1, lower.tail = F)  
  fvar<- c(fvar1, fvar2)
```

```

rejections <- replicate(m,
                        {
x1Control <- rnorm(n1)
x2Control <- rnorm(n2)

x1 <- (coef.fleishman[1] + (coef.fleishman[2]*x1Control) +
(coef.fleishman[3]* x1Control^2) + (coef.fleishman[4]*x1Control^3))*ratioSigmas
x2 <- (coef.fleishman[1] + (coef.fleishman[2]*x2Control) +
(coef.fleishman[3]* x2Control^2) + (coef.fleishman[4]*x2Control^3))+ diffmean

t.control <- t.val(x = x1Control, y = x2Control)
t.stud <- t.val(x = x1, y = x2)
stat.welch <- wel.val(x = x1, y = x2)
Q <- var(x1)/var(x2)
v <- (((var(x1)/n1)+(var(x2)/n2))^2)/((((var(x1)/n1)^2)/(n1-1))+
(((var(x2)/n2)^2)/(n2-1)))

wc <- qt(nominalAlpha/2, v, lower.tail = F)
p.wilcox <- p.val(x1,x2)
p.shap_x1 <- shapiro.test(x1)$p.value
p.shap_x2 <- shapiro.test(x2)$p.value

c(rechazos_sin_pretest(t.control, tc), rechazos_sin_pretest(t.stud, tc),
  rechazos_sin_pretest(stat.welch, wc), rechazo_pval(p.wilcox, nominalAlpha),
  rechazo_pretest_hom(Q, fvar, t.stud, tc, stat.welch, wc),
  rechazo_pretest_hom(Q, int, t.stud, tc, stat.welch, wc),
  rechazo_pretest_norm_hom(alphaPretest, p.shap_x1, p.shap_x2, Q, fvar, t.stud,
    tc, stat.welch, wc, p.wilcox),
  rechazo_pretest_norm_hom(alphaPretest, p.shap_x1, p.shap_x2, Q, int, t.stud,
    tc, stat.welch, wc, p.wilcox)))

.Random.seed <- old.seed
z <- qnorm(0.975)
controlResult <- numeric(7)
names(controlResult) <- c("t-Std", "Welch", "Wilcox", "F-Pretest", "Eq-Pretest",
  "Shap&F-Pretest", "Shap&Eq-Pretest")
attr(controlResult, "precis") <- rep(NaN, 7)
for (itest in 1:7) {
  mij <- cont(rejections[1,], rejections[itest + 1,])
  aux.control <- estimControlTIEP(mij, nominalAlpha)
  controlResult[itest] <- aux.control[1]
  attr(controlResult, "precis")[itest] <- z * aux.control[2]
}
return(controlResult)
}
}

```