



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

EFEECTO DE LAS ALTERACIONES DE LA NORMALIDAD
SOBRE PRUEBAS DE COMPARACIÓN DE MEDIAS
PARAMÉTRICAS Y NO PARAMÉTRICAS

Trabajo de Integración Curricular

Tipo: Proyecto de Investigación

Presentado para optar al grado académico de:

INGENIERO EN ESTADÍSTICA

AUTOR: LUIS REINALDO CHÁVEZ YUQUI

DIRECTOR: ING. PABLO JAVIER FLORES MUÑOZ, MSc.

Riobamba – Ecuador

2022

© 2022, Luis Reinaldo Chávez Yuqui

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.

Yo, LUIS REINALDO CHAVEZ YUQUI, declaro que el presente Trabajo de Integración Curricular es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Integración Curricular; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 10 de noviembre de 2022



Luis Reinaldo Chávez Yuqui

060460202-9

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO
FACULTAD DE CIENCIAS
CARRERA ESTADÍSTICA

El Tribunal del Trabajo de Integración Curricular, certifica que: El trabajo de Integración Curricular; Tipo: Proyecto de Investigación, **EFFECTO DE LAS ALTERACIONES DE LA NORMALIDAD SOBRE PRUEBAS DE COMPARACIÓN DE MEDIAS PARAMÉTRICAS Y NO PARAMÉTRICAS**, realizado por el señor: **LUIS REINALDO CHÁVEZ YUQUI**, ha sido minuciosamente revisado por los Miembros del Tribunal del Trabajo de Integración Curricular. El mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal autoriza su presentación.

	FIRMA	FECHA
Dra. Jaqueline Elizabeth Balseca Castro, Mgs. PRESIDENTE DEL TRIBUNAL		2022-11-10
Ing. Pablo Javier Flores Muñoz, MSc. DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR		2022-11-10
Ing. Johanna Enith Aguilar Reyes, Mgs. MIEMBRO DEL TRIBUNAL		2022-11-10

DEDICATORIA

A mis padres, Luis y Martha, quienes con su ejemplo supieron inculcarme la disciplina, honestidad y sobre todo el trabajo arduo para la consecución de metas en mi vida. A mi hermana Alexandra por ser mi ejemplo a seguir, por su apoyo e inquebrantable confianza. Finalmente, a todo mi entorno familiar quienes siempre estuvieron pendientes y con añoranza de éxito a lo largo de mi vida estudiantil.

Luis

AGRADECIMIENTO

Al Ing. Pablo Flores quien gracias a sus enseñanzas fue el artífice de mi interés por la simulación estadística, también agradecerle por su dedicación, tiempo y guía en el desarrollo de este proyecto de investigación. A la Ing. Johanna Aguilar por su paciencia y entereza mostrada a lo largo de mi estancia en la universidad. A mis amigos Fausto y Verónica quienes a lo largo de estos 4 años supieron brindarme su apoyo incondicional. Finalmente, a la Escuela Superior Politécnica de Chimborazo por brindarme la oportunidad de culminar una carrera universitaria, de forma gratuita y con educación de calidad.

Luis

ÍNDICE DE CONTENIDOS

ÍNDICE DE TABLAS.....	ix
ÍNDICE DE ILUSTRACIONES.....	x
ÍNDICE DE ANEXOS	xi
RESUMEN.....	xii
SUMMARY.....	xiii
INTRODUCCIÓN	1

CAPÍTULO I

1.	PROBLEMA DE INVESTIGACIÓN.....	3
1.1.	Planteamiento del Problema.....	3
1.2.	Limitaciones y delimitaciones.....	3
1.3.	Problema General de Investigación.....	3
1.4.	Problemas específicos de Investigación	3
1.5.	Objetivos.....	4
1.5.1.	<i>Objetivo General</i>	4
1.5.2.	<i>Objetivos Específicos</i>	4
1.6.	Justificación.....	4
1.6.1.	<i>Justificación Teórica</i>	4
1.6.2.	<i>Justificación Metodológica</i>	5
1.6.3.	<i>Justificación Práctica</i>	5

CAPÍTULO II

2.	MARCO TEÓRICO.....	6
2.1.	Antecedentes de investigación	6
2.2.	Referencias teóricas.....	7
2.2.1.	<i>Población</i>	7
2.2.2.	<i>Muestra</i>	7
2.2.3.	<i>Muestra aleatoria</i>	7
2.2.4.	<i>Parámetro</i>	8
2.2.5.	<i>Estadístico</i>	8
2.2.6.	<i>Hipótesis estadística</i>	8

2.2.6.1.	<i>Hipótesis simple</i>	8
2.2.6.2.	<i>Hipótesis compuesta</i>	8
2.2.7.	<i>Pruebas de hipótesis</i>	8
2.2.7.1.	<i>Hipótesis nula</i>	9
2.2.7.2.	<i>Hipótesis alternativa</i>	9
2.2.7.3.	<i>Estadístico de prueba</i>	9
2.2.7.4.	<i>Región de rechazo</i>	9
2.2.7.5.	<i>Región de no rechazo</i>	9
2.2.7.6.	<i>Valor crítico</i>	10
2.2.8.	<i>Tipos de pruebas de hipótesis</i>	10
2.2.9.	<i>Errores en una prueba de hipótesis</i>	10
2.2.9.1.	Error Tipo I.....	11
2.2.9.2.	Error Tipo II.....	11
2.2.10.	<i>Potencia</i>	12
2.2.11.	<i>Nivel de significancia</i>	12
2.2.12.	<i>Regla de decisión</i>	12
2.2.13.	<i>Valor p</i>	12
2.2.14.	<i>Simulación de Monte Carlo</i>	13
2.2.15.	<i>Ley de los grandes números</i>	13
2.2.16.	<i>Pruebas de hipótesis para comparar 2 medias poblacionales</i>	14
2.2.16.1.	<i>Prueba t-Student</i>	14
2.2.16.2.	<i>Prueba de Wilcoxon</i>	14
2.2.17.	<i>Pruebas de hipótesis para comparar más de 2 medias poblacionales</i>	15
2.2.17.1.	<i>ANOVA unifactorial</i>	15
2.2.17.2.	<i>Prueba de Kruskal-Wallis</i>	15
2.2.18.	<i>Método de Fleishman</i>	16
2.2.19.	<i>Criterio de Cochran</i>	16

CAPÍTULO III

3.	MARCO METODOLÓGICO	17
3.1.	Enfoque de Investigación	17
3.2.	Nivel de Investigación	17
3.3.	Diseño de Investigación	17
3.3.1.	<i>Según la manipulación o no de la variable independiente</i>	17
3.3.2.	<i>Según las intervenciones en el trabajo de campo</i>	18
3.4.	Tipo de estudio	18

3.5.	Población y Planificación, selección y cálculo del tamaño de la muestra	18
3.5.1.	<i>Población de estudio</i>	18
3.5.2.	<i>Tamaño de muestra</i>	18
3.5.3.	<i>Método de muestreo</i>	19
3.6.	Métodos, técnicas e instrumentos de investigación.....	19
3.6.1.	<i>Obtención de muestras no normales</i>	20
3.6.2.	<i>Proceso de simulación para la estimación de la TIEP y Potencia en pruebas de comparación de medias</i>	21

CAPÍTULO IV

4.	MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS.....	23
4.1.	Estimación de la TIEP de las pruebas de comparación de medias	23
4.2.	Estimación de la Potencia de las pruebas de comparación de medias	27

CONCLUSIONES	30
---------------------------	-----------

RECOMENDACIONES.....	32
-----------------------------	-----------

GLOSARIO

BIBLIOGRAFÍA

ANEXOS

ÍNDICE DE TABLAS

Tabla 1-2:	Tipos de pruebas de hipótesis.....	10
Tabla 2-2:	Regiones de rechazo	10
Tabla 3-2:	Tipos de errores	11
Tabla 4-2:	Valores-p	13
Tabla 1-3:	Tamaños muestrales	19
Tabla 2-3:	Coefficientes de Fleishman.....	20
Tabla 3-3:	Funciones utilizadas para las pruebas de comparación de medias	21

ÍNDICE DE ILUSTRACIONES

Ilustración 1-4:	Estimación de la TIEP de las pruebas de comparación de 2 medias	23
Ilustración 2-4:	Estimación de la TIEP de las pruebas de comparación de 3 medias	24
Ilustración 3-4:	Estimación de la TIEP de las pruebas de comparación de 5 medias	25
Ilustración 4-4:	Estimación de la Potencia de las pruebas de comparación de 2 medias	27
Ilustración 5-4:	Estimación de la Potencia de las pruebas de comparación de 3 medias	28
Ilustración 6-4:	Estimación de la Potencia de las pruebas de comparación de 5 medias	29

ÍNDICE DE ANEXOS

ANEXO A: TABLA DE RESULTADOS DE LA ESTIMACIÓN DE LA TIEP Y POTENCIA
DE LAS PRUEBAS DE COMPARACIÓN DE MEDIAS

ANEXO B: FUNCIÓN TIEPPOW

ANEXO C: FUNCIÓN r_{nonorm}

RESUMEN

El objetivo de este trabajo fue determinar el efecto de las alteraciones de la normalidad en pruebas de comparación de medias, mediante la estimación de su Probabilidad de Error Tipo I (TIEP) y Potencia. Para ello se creó un algoritmo de simulación basado en Monte Carlo, el cual estimó a la TIEP como la proporción de rechazos de la hipótesis nula cuando esta es cierta, y a la Potencia como la proporción de rechazos de la hipótesis nula cuando esta es falsa. Para la obtención de las distintas contaminaciones medibles de la normalidad en las muestras, se empleó el método de Fleishman. Los resultados mostraron que cuando existe normalidad, la robustez de los test paramétricos son mejores que la de sus alternativas no paramétricas, lo cual teóricamente se esperaba. Ante alteraciones de la normalidad la TIEP de los test paramétricos se controlaron de mejor manera que las de sus alternativas no paramétricas todo esto hasta ciertos niveles de contaminación, dependiendo el nivel en el cual aún son robustas, del número de medias a comparar y de los tamaños muestrales. Pero en general la TIEP de los test paramétricos son muy variables, mientras que la TIEP de los test no paramétricos permanecen constantes. La Potencia de los test paramétricos fueron mejores que la de sus alternativas no paramétricas, esto también hasta cierto grado de contaminación de la normalidad, sin embargo, cuando los tamaños muestrales fueron grandes la Potencia tanto de los test paramétricos y no paramétricos presentaron valores similares, cercanos o iguales a uno. Este trabajo da una visión del uso correcto de las pruebas de comparación de medias ante distintos escenarios de no normalidad, tamaño muestral y número de tratamientos a compararse, sugerimos que estos resultados sean aplicados en análisis inferenciales, lo cual conllevará a una correcta toma de decisiones.

Palabras clave: <PRUEBA DE COMPARACIÓN DE MEDIAS>, <POTENCIA>, <PROBABILIDAD DE ERROR TIPO I (TIEP) >, <SIMULACIÓN>, <MONTE CARLO (ALGORITMO DE SIMULACIÓN)>.



2180-DBRA-UTP-2022

SUMMARY

The objective of this work was to determine the effect of alterations in normality in comparison tests of means, by estimating their Type I Error Probability (TIEP) and Power. For this, a simulation algorithm based on Monte Carlo was created, which estimated the TIEP as the proportion of rejections of the null hypothesis when it is true, and the Power as the proportion of rejections of the null hypothesis when it is false. To obtain the different measurable contaminations of normality in the samples, the Fleishman method was used. The results showed that when there is normality, the robustness of the parametric tests is better than their non-parametric alternatives, which was theoretically expected. In the event of alterations in normality, the TIEP of the parametric tests were better controlled than their non-parametric alternatives, all this up to certain levels of contamination, depending on the level at which they are still robust, the number of means to be compared and the sample sizes. In general, the TIEP of the parametric tests is highly variable, while the TIEP of the non-parametric tests remains constant. The Power of the parametric tests was better than that of their non-parametric alternatives, this also up to a certain degree of contamination of normality, however, when the sample sizes were large, the Power of both the parametric and non-parametric tests presented similar values, close to or equal to one. This work gives a vision of the correct use of means comparison tests in different scenarios of non-normality, sample size and number of treatments to be compared, it is suggested that these results should be applied in inferential analysis, which will lead to correct decision making.

Keywords: <MEANS COMPARISON TEST>, <POWER>, <TYPE I ERROR PROBABILITY (TIEP)>, <SIMULATION>, <MONTE CARLO (SIMULATION ALGORITHM)>.



Edgar Mesias Jaramillo Moyano
0603497397

INTRODUCCIÓN

Entre las técnicas estadísticas más utilizadas para probar la igualdad de dos medias poblacionales independientes, destacan el contraste paramétrico de la *t-Student* (Student, 1908, pp. 1-25) y el contraste no paramétrico de *Wilcoxon* (Wilcoxon, 1992, pp. 196-202). Para probar la igualdad de más de 2 medias los test más empleados son el test paramétrico del ANOVA unifactorial (Fisher, 1938, pp. 376-386) y el no paramétrico de *Kruskal-Wallis* (Kruskal & Wallis, 1952, pp. 583-621).

En las pruebas de la *t-Student* y la del ANOVA unifactorial un supuesto que se debe cumplir es el de la normalidad (las muestras tienen que distribuirse de forma aproximadamente normal), y en el caso de que no se cumpla este supuesto la teoría nos menciona que debemos aplicar la prueba de *Wilcoxon* (como alternativa a la *t-Student*) y la prueba de *Kruskal-Wallis* (como alternativa al ANOVA unifactorial).

Sin embargo ciertas investigaciones como la de (Stonehouse & Forrester, 1998, pp. 63-74), mostraron que, en la mayoría de los escenarios a excepción de ciertos casos puntuales, el test paramétrico de la *t-Student* presenta mayor robustez que su alternativa no paramétrica ante la violación de los supuestos de normalidad y homocedasticidad, y sugieren que el test de *Wilcoxon* no debería ser propiamente un análogo no paramétrico de la prueba t, como usualmente se lo hace.

(Blanca Mena, et al., 2017, pp. 552-557) en un estudio acerca de la violación del supuesto de normalidad en el ANOVA unifactorial, mostraron que esta prueba de hipótesis es robusta en términos de la estimación de Error Tipo I, en el 100% de los casos estudiados.

Si bien estos estudios arrojan resultados muy interesantes, y aportan a la cantidad muy reducida de investigaciones, referentes a la medición de la eficiencia de las pruebas de comparación de medias ante el no cumplimiento del supuesto de la normalidad en las muestras, casi ninguna de estas y otras publicaciones emplean contaminaciones medibles de la normalidad.

Es por esto que se plantea la presente investigación, en la cual, gracias a la creación de un algoritmo que tiene como base la simulación de Monte Carlo, permitirá estimar la Probabilidad de Error tipo I (TIEP por sus siglas en inglés) y la Potencia de las pruebas de comparación de medias cuando no se da el cumplimiento de la normalidad en las muestras. Suponiendo muestras homocedásticas para la obtención de las mismas, alejadas de una perfecta normalidad, se simularán muestras no normales, mediante el método de *Fleishman* (Fleishman, 1978, pp. 521-532). Posteriormente usando distintos tamaños muestrales y diferentes desviaciones medibles de la normalidad, se compararán los resultados de las estimaciones de la TIEP y de la Potencia, con el fin de verificar si las pruebas de hipótesis funcionan adecuadamente bajo los escenarios que teóricamente lo deberían hacer.

Todo lo que deseamos llevar a cabo en este trabajo, es posible, gracias a que en la actualidad contamos con computadoras y softwares estadísticos, capaces de ejecutar simulaciones a grandes

escalas, algo que en la antigüedad no se podía realizar y por ello no se podía hacer este tipo de investigaciones.

La importancia de este trabajo radica en el sentido de que permitirá a los analistas de datos, aplicar adecuadamente una prueba de hipótesis de comparación de medias, y de esta manera obtener conclusiones validas en sus investigaciones.

El primer capítulo empieza describiendo el planteamiento del problema, limitaciones y delimitaciones, problema general y problemas específicos de investigación, los objetivos y la justificación. En el segundo capítulo se presentan los antecedentes y los conceptos de los métodos de simulación y de las pruebas de hipótesis a utilizar. El tercer capítulo describe todo lo referente al marco metodológico. En el cuarto capítulo se analizan y discuten los principales resultados encontrados, y finalmente se muestran las conclusiones y recomendaciones de la investigación.

CAPÍTULO I

1. PROBLEMA DE INVESTIGACIÓN

1.1. Planteamiento del Problema

Teóricamente, una prueba de hipótesis está diseñada de tal modo que su TIEP (Probabilidad de rechazar H_0 cuando esta es verdadera) coincida con el nivel de significancia α y se espera que su Potencia (Probabilidad de rechazar H_0 cuando esta es falsa) esté muy cerca de 1, sin embargo, estudios han mostrado que, bajo ciertas condiciones, este escenario ideal no siempre es cierto (Cohen, 1962, pp. 145-153). Suponiendo muestras homocedásticas, nos preguntamos si la TIEP y Potencia de los test paramétricos y no paramétricos de comparación de medias, se alteran ante desviaciones o contaminaciones de la normalidad.

1.2. Limitaciones y delimitaciones

Una de las limitaciones del proyecto de investigación es la reducida cantidad de estudios previos recientes, que hagan referencia a la eficiencia de las pruebas de comparación de medias cuando se presenta la violación del supuesto de normalidad en las muestras, otra limitación importante es la imposibilidad de comprobar matemáticamente que en un contraste de hipótesis su TIEP se encuentre controlada y que su Potencia sea alta, en vez de esto, nosotros lograremos obtener valores estimados de estos parámetros gracias un proceso de simulación .

Sobre la delimitación de este estudio, nos enfocaremos únicamente en la violación del supuesto de la normalidad, por ende, en todos los casos trabajaremos con muestras homocedásticas, para los tamaños de las muestras, se definirán ciertos tamaños muestrales representativos, en cuanto a los contrastes de hipótesis a los cuales vamos a medir su eficiencia, están los test paramétricos de la *t-Student*, el ANOVA unifactorial, y los contrastes no paramétricos de *Wilcoxon* y de *Kruskal-Wallis*.

1.3. Problema General de Investigación

¿Cómo se afectan la Probabilidad de Error Tipo I y la Potencia de las pruebas de comparación de medias cuando se altera la normalidad en las muestras?

1.4. Problemas específicos de Investigación

- ¿Cómo obtener muestras no normales que sigan una distribución totalmente desconocida?

- ¿De qué manera un algoritmo de simulación, nos ayuda a obtener valores de la estimación de la TIEP y de la Potencia de las pruebas de comparación de medias, ante alteraciones medibles de la normalidad?
- ¿Cuáles de las pruebas de hipótesis estudiadas son más eficientes para contrastar diferencias significativas entre medias, cuando existen o no alteraciones de la normalidad en las muestras?

1.5. Objetivos

1.5.1. Objetivo General

- Determinar el efecto de las alteraciones de la normalidad sobre pruebas de comparación de medias paramétricas y no paramétricas, mediante la estimación de su Error Tipo I y Potencia, usando algoritmos de simulación

1.5.2. Objetivos Específicos

- Crear funciones propias que simulen muestras con alteraciones medibles de la normalidad con base a los coeficientes de Fleishman.
- Crear funciones que estimen a través de un proceso de simulación la Probabilidad de Error Tipo I y la Potencia de las pruebas de comparación de medias, cuando se usan muestras con alteraciones medibles de la normalidad.
- Estudiar la alteración de las pruebas de comparación de medias, usando distintos tamaños muestrales y diferentes alteraciones medibles de la normalidad en las muestras.

1.6. Justificación

1.6.1. Justificación Teórica

Este proyecto de investigación, se realiza con el propósito de aportar a la falta de estudios previos, acerca de la afectación de la TIEP y Potencia de las pruebas de comparación de medias, cuando no se da el cumplimiento de la normalidad en las muestras, puesto que gracias a los resultados que se obtengan en el presente estudio se podrá determinar bajo ciertos escenarios cuál de los 2 test el paramétrico o no paramétrico es más eficiente.

1.6.2. Justificación Metodológica

Teóricamente se espera que la TIEP de una prueba de hipótesis coincida con el nivel de significancia, y que su Potencia presente un valor cercano a 1, pero no existe una forma matemática de comprobar que esto realmente suceda, por eso al ser ambos parámetros desconocidos, mediante un proceso de simulación de Monte Carlo se puede estimar. A la TIEP como la proporción de veces que se rechaza una hipótesis de igualdad de medias que teóricamente es cierta. Mientras que a la Potencia como la proporción de veces que se rechaza una hipótesis de igualdad de medias que teóricamente es falsa.

Al realizar estas estimaciones, se generarán ciertos errores, los cuales disminuirán, gracias a la creación de algoritmos óptimos de simulación, logrando estimaciones de la TIEP y de la Potencia lo más perfectas posibles.

1.6.3. Justificación Práctica

La mayoría de investigadores, cuando no se cumple el supuesto de la normalidad en las muestras, recurren inmediatamente a un contraste no paramétrico, ignorando ante la falta de información, si este nuevo test elegido será o no más eficiente que el test paramétrico, por ende, la presente investigación permitirá ampliar los conocimientos acerca de si conviene o no, ante violaciones de la normalidad, el optar por un contraste no paramétrico.

Por ejemplo, enfocándose en un estudio médico realizado por (Zambrano, et al., 2015), se requería hacer un análisis de comparación de la efectividad del paracetamol vs el ibuprofeno en el control del dolor postquirúrgico, en alargamientos de corona ; el estudio demostró que ambos medicamentos tenían el mismo efecto supresor del dolor, pero que hubiese pasado, si como consecuencia de que en el test su TIEP está alejada del nivel de significancia o su Potencia es baja, se llegaba a determinar que un medicamento tenía mayor efecto supresor que otro, pues quizá se hubiese comprado mayores cantidades de dicho medicamento , y capaz ese producto costaba más que el otro, ocasionando gastos innecesarios a el hospital, en resumen, se hubiesen obtenido conclusiones no válidas, de ahí la necesidad de conocer si el test que estamos empleando es realmente el más eficiente o no.

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Antecedentes de investigación

La eficiencia de una prueba de hipótesis es fundamental, puesto que cuando mayor eficiencia se muestra en el contraste de hipótesis mejores serán las conclusiones obtenidas. Un trabajo interesante respecto a la eficiencia en particular de la prueba *t-Student* en la comparación de 2 medias, fue realizado por (Montilla & Kromrey, 2010, pp. 101-107) quienes estudiaron la robustez de las pruebas T en comparación de medias, ante la violación de los supuestos de normalidad y homocedasticidad.

En este trabajo del año 2010 solo se utilizaron contrastes paramétricos y se examinó la robustez de los mismos cuando los tamaños de las muestras son iguales/desiguales y bajo el efecto de violación de los supuestos de normalidad y homocedasticidad. Se obtuvo conclusiones interesantes acerca de la Probabilidad de Error tipo I, y una de ellas fue que, para muestras heterocedásticas y con contaminaciones de la Normalidad, la TIEP presenta serias afectaciones, excepto cuando se consideran tamaños de muestras grandes e iguales.

Otro autor (Sánchez Turcios, 2015, pp. 59-61) afirmó que en la prueba *t-Student*, aunque una de las muestras, no provenga de una población normal pero la otra sí, y la razón de la varianza más grande a la más pequeña sea menor a 2, la prueba resultaba aún adecuada.

En una investigación de (Stonehouse & Forrester, 1998, pp. 63-74), se indagó sobre la robustez de las pruebas t y U de *Mann-Whitney* (prueba de *Mann-Whitney-Wilcoxon*), ante la violación combinada de los supuestos, mostrando que la prueba t es suficientemente robusta para su uso en todos los casos probables, a excepción de cuando la asimetría es severa o cuando las varianzas de la población y los tamaños de las muestras difieren. En contra de lo que la teoría nos dice, los autores mencionan que la prueba de *Mann-Whitney-Wilcoxon* mostró una dramática falta de robustez en muchos casos, por lo cual recomiendan que este test no debería ser propiamente un análogo no paramétrico de la prueba t, como usualmente se lo hace.

Refiriéndose a la violación del supuesto de Normalidad en el ANOVA unifactorial, los autores (Mena, et al., 2017, pp. 552-557) demostraron que este test empleado para 3 grupos, es robusto en términos de la estimación de Error Tipo I, en el 100% de los casos estudiados, independientemente del grado de contaminación de la normalidad en las muestras, del tamaño, de si son balanceadas o desbalanceadas, de si la distribución es igual o desigual en los grupos.

En un estudio comparativo de la Potencia del ANOVA y del test de *Kruskal-Wallis* ante violaciones de la Normalidad, ejecutado por (Van Hecke, 2012, pp. 241-247), quien empleó un método de permutación como método de simulación para estimar la potencia de las pruebas. Obtuvo como

resultado que, en el caso de poblaciones asimétricas, en términos del análisis de su potencia estimada, la prueba no paramétrica de *Kruskal-Wallis* se comporta mejor que el ANOVA unifactorial.

Como ya hemos visto, algunas investigaciones manifiestan que, ante desviaciones de la normalidad, los test de la t-Student y el ANOVA unifactorial bajo ciertos escenarios, resultan ser más robustos que su alternativa no paramétrica, dichos estudios analizaron bien sea la TIEP o la Potencia de las pruebas, pero no ambas a la vez, y aún más lo que debemos destacar es que un número casi nulo de estas publicaciones, hicieron uso de algún método que les permitiese poseer contaminaciones medibles de la normalidad en las muestras, por eso creemos que para estudiar de mejor manera la eficiencia de estos contrastes de hipótesis, se requiere de la estimación de ambos parámetros (TIEP y Potencia), y del empleo del método de *Fleishman* para simular las muestras no normales con diferentes grados de contaminación.

2.2. Referencias teóricas

2.2.1. Población

Es un conjunto o colección completa de elementos, de los cuales nos interesa estudiar algunas de sus características. Cada medición que se hace sobre cada elemento de la población es un valor de una variable aleatoria X con función de probabilidad $f(x)$.

2.2.2. Muestra

Es cualquier subconjunto de una población. Se debe buscar que la muestra sea lo más representativa posible, para que la generalización sea confiable y se disminuya el sesgo.

2.2.3. Muestra aleatoria

Sea la población dada por la variable aleatoria X con función de probabilidad $f(x)$. Una muestra aleatoria es una colección de variables aleatorias X_1, \dots, X_n tales que:

- X_1, \dots, X_n son independientes entre sí.
- Cada variable aleatoria X_i , para $i = 1, 2, \dots, n$ tiene la misma distribución que X .

Es decir X_1, \dots, X_n son iid (independientes e idénticamente distribuidas).

2.2.4. Parámetro

Es un valor numérico que se le asigna a la población. Se lo ignora a causa de algunos factores como el tamaño poblacional, el acceso a los datos que componen la misma, etc. Usualmente se lo denota con la letra minúscula griega θ .

2.2.5. Estadístico

Es una función de una muestra aleatoria X_1, \dots, X_n que no depende de parámetros desconocidos. Los Estadísticos también son variables aleatorias, puesto que son funciones de las variables aleatorias observadas en una muestra. La distribución de probabilidad de un estadístico se llama distribución muestral. Lo denotaremos como $\hat{\theta}$.

$$\hat{\theta} = T(X_1, \dots, X_n)$$

2.2.6. Hipótesis estadística

Es una afirmación acerca de la distribución de una o más variables aleatorias. Se puede especificar una hipótesis, dando el tipo de distribución y el valor o valores del parámetro o los parámetros que la definen (Moya & Saravia, 1988, p. 701).

2.2.6.1. Hipótesis simple

Una hipótesis es simple cuando se precisa por completo la distribución de probabilidad en estudio.

2.2.6.2. Hipótesis compuesta

Una hipótesis es compuesta si no se especifica por completo la distribución de probabilidad en estudio.

2.2.7. Pruebas de hipótesis

Es una técnica estadística que se emplea para validar o no una afirmación, en base a la información adquirida de una muestra (Sanabria Brenes, 2011, p. 73). Los elementos más sustanciales de una prueba de hipótesis estadística son los siguientes: hipótesis nula, hipótesis alternativa, estadístico de prueba y la región de rechazo (Wackerly, et al., 2010).

2.2.7.1. *Hipótesis nula*

La hipótesis nula es la que se supone como cierta y se busca evidencias en contra de esta suposición, se la denota como H_0 . Es una afirmación que habitualmente hace referencia al valor poblacional de un parámetro o a la distribución de una variable aleatoria. Para que H_0 no sea rechazada no debe haber una fuerte evidencia en su contra.

2.2.7.2. *Hipótesis alternativa*

La hipótesis alternativa es lo que se quiere demostrar, es denotada como H_1 , se plantea en oposición a la hipótesis nula y es aceptada en caso de que H_0 sea rechazada. La hipótesis alternativa es la negación de la hipótesis nula, es decir incluye todo lo que H_0 excluye.

Como suponemos que H_0 es cierta, no buscamos evidencia de H_0 , sino de H_1 . Si no encontramos evidencia de H_1 , esto no quiere decir que H_0 sea cierta, H_0 nunca se considera probada, aunque puede ser rechazada por los datos de la muestra.

2.2.7.3. *Estadístico de prueba*

Es un resultado muestral que satisface las dos condiciones siguientes: primero proporcionar información empírica relevante sobre la afirmación propuesta en H_0 y segundo poseer una distribución muestral conocida (Molina & Lara, 2018).

2.2.7.4. *Región de rechazo*

También denominada región crítica, es el área de distribución muestral que atañe a los valores del estadístico de contraste que se localizan tan apartados de la afirmación establecida en H_0 , es decir que es muy poco probable que ocurran si H_0 es verdadera (Molina & Lara, 2018).

2.2.7.5. *Región de no rechazo*

Es el área de la distribución muestral que corresponde a los valores del estadístico de contraste próximos a la afirmación establecida en H_0 . Es decir, los valores del estadístico de contraste que nos conducen a no rechazar H_0 .

2.2.7.6. Valor crítico

Es el valor del estadístico de contraste, a partir del cual se rechaza H_0 . Se lo denota como $\hat{\theta}_c$. Separa la región de rechazo de la de no rechazo.

2.2.8. Tipos de pruebas de hipótesis

Hay 2 tipos principales de pruebas, cada una de las cuales es identificada por la forma en que se formulan H_0 y H_1 , son las pruebas bilaterales y las unilaterales.

Tabla 1-2: Tipos de pruebas de hipótesis

Prueba de hipótesis		
Bilateral o de 2 colas	Unilateral: cola a la derecha	Unilateral: cola a la izquierda
$H_0: \theta = \theta_0$	$H_0: \theta \leq \theta_0$	$H_0: \theta \geq \theta_0$
$H_1: \theta \neq \theta_0$	$H_1: \theta > \theta_0$	$H_1: \theta < \theta_0$

Realizado por : Chávez, L., 2022.

Cualquiera que sea el caso, en H_0 siempre debe aparecer un signo de igualdad, ya sea como una igualdad en sí (=) o como una desigualdad no estricta (\leq o \geq). En la Tabla N° 2-2 se muestran las regiones de rechazo para los contrastes mencionados anteriormente.

Tabla 2-2: Regiones de rechazo

Prueba de Hipótesis	Región de rechazo
$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$	$P(\hat{\theta} < \hat{\theta}_c \theta = \theta_0) + P(\hat{\theta} > \hat{\theta}_c \theta = \theta_0)$
$H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_0$	$P(\hat{\theta} > \hat{\theta}_c \theta = \theta_0)$
$H_0: \theta \geq \theta_0$ $H_1: \theta < \theta_0$	$P(\hat{\theta} < \hat{\theta}_c \theta = \theta_0)$

Realizado por : Chávez, L., 2022.

2.2.9. Errores en una prueba de hipótesis

Los dos tipos de errores que se pueden presentar se denominan error tipo I y error tipo II.

2.2.9.1. Error Tipo I

Se comete cuando se rechaza H_0 siendo ésta verdadera o cierta. A la probabilidad de cometer este error se le denota por la letra α y se la conoce como la Probabilidad de Error Tipo I (TIEP). (Rincón, 2019, p. 270)

$$\alpha = P(\text{"Rechazar } H_0 \text{ " | " } H_0 \text{ es verdadera"})$$

Por tanto $1 - \alpha$ es la probabilidad de no cometer el Error Tipo I, es decir es la probabilidad de tomar una decisión correcta cuando H_0 es cierta. Este valor α es la medida de la región o regiones de rechazo. (Sanabria Brenes, 2011). El error tipo I es mayor cuando el valor crítico $\hat{\theta}_c$ es cercano al valor nulo θ_0 .

2.2.9.2. Error Tipo II

El error tipo II se comete cuando no se rechaza H_0 siendo esta falsa, la probabilidad de este error se le denota por la letra β .

$$\beta = P(\text{"No rechazar } H_0 \text{ " | " } H_0 \text{ es falsa"})$$

A su vez $1 - \beta$ es la probabilidad de que no ocurra el error tipo II es decir es la probabilidad de tomar una decisión correcta cuando H_0 es falsa, y a esta probabilidad se la conoce como la Potencia de la Prueba (Sanabria Brenes, 2011, p. 84).

Tabla 3-2: Tipos de errores

		Naturaleza de H_0	
		H_0 cierta	H_0 falsa
Decisión sobre H_0	Rechazar H_0	Error Tipo I $P = \alpha$	Decisión correcta $P = 1 - \beta$
	No rechazar H_0	Decisión correcta $P = 1 - \alpha$	Error Tipo II $P = \beta$

Fuente: Molina & Lara., 2018.

Realizado por : Chávez, L., 2022.

Las probabilidades α y β no son complementarias, es decir, no obligatoriamente suman 1, puesto que los eventos condicionantes que aparecen en las probabilidades condicionales anteriores son diferentes. (Rincón, 2019). Es obvio que deseamos que los valores de estas probabilidades sean pequeñas, pero α y β suelen estar en contraposición, es decir usualmente cuando se disminuye α , β aumenta o viceversa, por lo tanto, se podrá controlar solo a uno de los errores, y este será al

error tipo I. Se controla este error por medio del nivel de significancia que lo definimos más adelante.

2.2.10. Potencia

Es la probabilidad de que no ocurra el error tipo II, es decir es igual a $1 - \beta$ como ya se lo había mencionado anteriormente.

$$\text{Potencia} = P(\text{"Rechazar } H_0 \text{"} | \text{"}H_0 \text{ es falsa"})$$

El concepto de potencia es empleado para medir la eficiencia de un test de hipótesis (Molina & Lara, 2018). Conforme más alejada se encuentra H_1 de H_0 menor será la probabilidad de cometer un error tipo II y, por lo tanto, la potencia poseerá valores más contiguos a 1.

2.2.11. Nivel de significancia

Es el valor máximo aceptable para α , la probabilidad de que ocurra el error tipo I (Sanabria Brenes, 2011).

$$\text{nivel de significancia} = \text{máx}_{\text{aceptable}} P(\text{"Error Tipo I"})$$

De manera general el nivel de significancia de una prueba es dada. Formalmente este valor no es la probabilidad α del error tipo I, sino una cota superior para α . Pero, si se considera α igual al nivel de significancia, entonces $\hat{\theta}_c$ será el valor crítico permitido que genere el máximo error tipo I aceptable y por lo tanto el más cercano a θ_0 . Además, bajo esta consideración, el nivel de significancia es la medida de la región de rechazo (Sanabria Brenes, 2011).

2.2.12. Regla de decisión

Es el enfoque que se utiliza para decidir si la hipótesis nula planteada debe o no ser rechazada en favor de la hipótesis alternativa. Esta regla nos señala que se ha de rechazar H_0 en favor de H_1 cuando el valor del estadístico de contraste caiga dentro de la región de rechazo, caso contrario, no se rechazará H_0 .

2.2.13. Valor p

El p-valor de un contraste de hipótesis es la probabilidad de observar un valor del estadístico de contraste tanto o más extremo que el que se ha observado suponiendo que H_0 es cierta. Por tanto, p-valores pequeños nos llevarán a rechazar la hipótesis nula. Específicamente, si el p-valor obtenido es más pequeño que el nivel de significancia, rechazaremos H_0 , caso contrario, no

podremos rechazar H_0 . En resumen, cuanto más bajo es el p-valor, más ridículo me sentiría defendiendo mi pensamiento inicial, dadas las evidencias observadas. En la Tabla N° 4-2 se muestra la redacción del p-valor asociado a cada tipo de prueba de hipótesis.

Tabla 4-2: Valores-p

Prueba de hipótesis	p-valores
$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$	Si $\hat{\theta}$ tiene una distribución simétrica respecto a θ_0 entonces: $p - valor = \begin{cases} 2P(\hat{\theta} < \hat{\theta} \theta = \theta_0) & \text{si } \hat{\theta} < \theta_0 \\ 2P(\hat{\theta} > \hat{\theta} \theta = \theta_0) & \text{si } \hat{\theta} > \theta_0 \end{cases}$
$H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_0$	$p - valor = P(\hat{\theta} \geq \hat{\theta} \theta = \theta_0)$
$H_0: \theta \geq \theta_0$ $H_1: \theta < \theta_0$	$p - valor = P(\hat{\theta} \leq \hat{\theta} \theta = \theta_0)$

Fuente: Sanabria Brenes, 2011.

Realizado por : Chávez, L., 2022.

2.2.14. Simulación de Monte Carlo

Es un método estadístico destinado a la solución de problemas matemáticos complejos, por medio de la generación de variables aleatorias, es fundamentalmente un muestreo experimental cuyo propósito es estimar las distribuciones de las variables de salida que depende de variables probabilísticas de entrada (Azofeifa, 2004, pp. 97-109).

2.2.15. Ley de los grandes números

Sea X_1, \dots, X_n una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas con media finita μ . Entonces, cuando $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mu$$

en donde la convergencia se verifica en el sentido casi seguro (ley fuerte) y también en probabilidad (ley débil) (Poisson, 1837, p. 7).

2.2.16. Pruebas de hipótesis para comparar 2 medias poblacionales

2.2.16.1. Prueba t-Student

Sea X_1, \dots, X_n una muestra aleatoria de tamaño n_1 extraída de la población $N(\mu_1, \sigma_1)$, y Y_1, \dots, Y_n una muestra aleatoria de tamaño n_2 extraída de la población $N(\mu_2, \sigma_2)$, ambas muestras independientes, con varianzas desconocidas pero iguales $\sigma_1^2 = \sigma_2^2$.

La Hipótesis nula a plantear es la siguiente; $H_0: \mu_1 - \mu_2 = 0$, mientras que la alternativa es: $H_1: \mu_1 - \mu_2 \neq 0$

El estadístico de contraste es:

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Donde

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

Y el p-valor se lo calcula como:

$$p - \text{valor} = P(|T_{\text{calculada}}| \geq t_{n_1+n_2-2, 1-\alpha/2})$$

Si el $p - \text{valor} < \alpha$ se rechaza H_0 en favor de H_1 .

2.2.16.2. Prueba de Wilcoxon

El test de *Mann-Whitney-Wilcoxon* (WMW), también denominado como *Wilcoxon rank-sum test* o *u-test*, es un contraste de hipótesis no paramétrico que surge como alternativa a la prueba t-Student cuando no existe Normalidad en las muestras a comparar (Amat, 2017). El estadístico de contraste es:

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

Siendo

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

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

Donde R_1, R_2 son las sumatorias de los rangos de la muestra 1 y 2 respectivamente.

La distribución del estadístico U para muestras lo suficientemente grandes (más de 20 observaciones) se aproxima bastante bien a la normal, dicha aproximación viene dada por la siguiente expresión:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1)$$

2.2.17. Pruebas de hipótesis para comparar más de 2 medias poblacionales

2.2.17.1. ANOVA unifactorial

El análisis de varianza (ANOVA) de una vía es empleada a fin de establecer si existen diferencias estadísticamente significativas entre las medias de tres o más grupos independientes (Ferrero, 2017). Es una generalización de los t-test independientes para más de dos muestras. Se lo utiliza cuando los datos no están pareados y se requiere indagar si existen diferencias significativas entre las medias de una variable aleatoria continua en los distintos niveles de otra variable cualitativa o factor (Amat, 2016). El estadístico de contraste es:

$$F_{ratio} = \frac{\text{Cuadrados Medios del Factor}}{\text{Cuadrados Medios del Error}} = \frac{\text{Intervarianza}}{\text{Intravarianza}} \sim F_{k-1, N-k}$$

Donde

k : número de niveles del factor.

N : número total de observaciones.

Intervarianza: varianza entre las medias de los distintos niveles.

Intravarianza: varianza dentro de los niveles, conocida como varianza residual o de error.

Como el estadístico F_{ratio} sigue una distribución F con $k - 1$ y $N - k$ grados de libertad, es plausible determinar la probabilidad de obtener valores iguales o más extremos que los observados.

2.2.17.2. Prueba de Kruskal-Wallis

El test de *Kruskal-Wallis* es la alternativa no paramétrica al test ANOVA de una vía para muestras no apareadas. Es una generalización del test de *Mann-Whitney* para más de dos grupos. Es un test que utiliza rangos para constatar la hipótesis de que k muestras han sido extraídas de una misma población (Amat, 2016).

En el test de *Kruskal-Wallis* se contrasta si las diferentes muestras están equidistribuidas y que por lo tanto pertenecen a una misma distribución. El estadístico de contraste es:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Donde

N : es el número total de observaciones entre todos los grupos.

k : número de grupos.

R_i : suma de rangos en el grupo i .

n_i : número de observaciones en el grupo i

Dicho estadístico H sigue aproximadamente una distribución χ^2 con $(k - 1)$ grados de libertad. La aproximación es mejor conforme mayor es el número de muestras y el tamaño de las mismas.

2.2.18. Método de Fleishman

Se utiliza el procedimiento propuesto por (Fleishman, 1978, pp. 521-532) para generar muestras no normales cuya distribución sea desconocida. Partiendo de una normal estándar Z ($Z \sim N(\mu = 0, \sigma = 1)$), el autor demostró que una transformación lineal de esta variable Z de la forma:

$$Y = a + bZ + cZ^2 + dZ^3$$

Sigue una distribución totalmente desconocida de parámetros $(\mu = 0, \sigma = 1, \gamma_1, \gamma_2)$, donde γ_1, γ_2 representan respectivamente los coeficientes poblacionales de asimetría y curtosis, los cuales a su vez definen los grados de contaminación de la normal.

Este método calcula los coeficientes a, b, c y d mediante una transformación polinómica que implica diferentes valores del tercer y cuarto momento central, es decir la asimetría (γ_1) y curtosis (γ_2). El primer y segundo momento se fijan arbitrariamente en 0 y 1, respectivamente.

2.2.19. Criterio de Cochran

El criterio de Cochran señala que un modelo de prueba de hipótesis puede ser considerado útil o preciso si la TIEP presenta una desviación respecto al nivel de significancia de máximo del 20% de su valor, es decir la TIEP de la prueba tiene que ubicarse dentro del intervalo $[\alpha \pm 0.2 \alpha]$ (Cochran, 1942, pp. 421-436).

CAPÍTULO III

3. MARCO METODOLÓGICO

3.1. Enfoque de Investigación

La presente investigación presenta un enfoque cuantitativo, puesto que tiene como objeto de estudio a un fenómeno cuantificable, como lo es la afectación de la TIEP y la Potencia de las pruebas de comparación de medias cuando se altera la normalidad en las muestras, además para obtener y evaluar la información nos basamos de cantidades numéricas, buscando la misma, ser lo más objetiva posible, centrándose en los datos y no en subjetividades.

3.2. Nivel de Investigación

El nivel de investigación es explicativo, pues la intención de este estudio es brindar mayor información acerca del efecto de las alteraciones de la normalidad sobre pruebas de comparación de medias. Como requerimos indagar de una manera más detallada a este fenómeno, se estudia el comportamiento de la TIEP y la Potencia de estos contrastes, bajo distintos escenarios.

Al ser ambos parámetros desconocidos se los estimó a través de un algoritmo de simulación sobre la base de Monte Carlo; para la obtención de las muestras no normales, se empleó el método de *Fleishman*, finalmente usando distintos tamaños muestrales y diferentes desviaciones medibles de la normalidad, se compararon los resultados de las estimaciones de la TIEP y de la Potencia. En el caso de TIEP, se usó el criterio de Cochran para definir si la prueba de hipótesis es útil o no. Todo lo mencionado anteriormente se lo realizó con el fin de verificar si las pruebas de hipótesis funcionan adecuadamente y, en el caso de no hacerlo, explicar las razones de la baja de eficiencia del contraste. No se pudo estudiar todos los escenarios, por eso se habló de casos específicos.

3.3. Diseño de Investigación

3.3.1. *Según la manipulación o no de la variable independiente*

Según la manipulación de la variable independiente (variable con datos normales o con datos no normales), este diseño de investigación es de índole experimental, puesto que los datos se generaron mediante algoritmos de simulación. En este caso, se simularon muestras con contaminaciones medibles de la normalidad, empleando el método de *Fleishman* y, por ende, se modificó de acuerdo con lo que se necesitaba investigar. Como se estudió la TIEP y la Potencia

bajo distintos escenarios, dependiendo del escenario a indagar, se modificó la muestra de acuerdo con el nivel de contaminación de normalidad que se quería tener.

3.3.2. *Según las intervenciones en el trabajo de campo*

El presente proyecto según el período temporal es transversal, pues los datos se generaron gracias a un algoritmo de simulación y por ende, al trabajar con simulaciones, las mismas se las realizó en un momento específico del tiempo, es decir el estudio se lo conduce en un tiempo determinado y no en intervalos de tiempo.

3.4. Tipo de estudio

Es un estudio del tipo documental, puesto que para obtener la información acerca de la afectación de la TIEP y la Potencia de las pruebas de comparación de medias cuando se altera la normalidad en las muestras, se recurrió a libros, artículos académicos y otras fuentes documentales. No se obtuvo datos de la realidad, más bien se generaron mediante un algoritmo de simulación y se los modificó de acuerdo con lo que se buscó estudiar, es decir, se pudo controlar los escenarios en los cuales se produjeron los datos.

3.5. Población y Planificación, selección y cálculo del tamaño de la muestra

3.5.1. *Población de estudio*

La población de estudio, en este caso, hace referencia a poblaciones estadísticas teóricas, la una proveniente de una distribución normal, y la otra proveniente de una distribución desconocida (con contaminación de la normalidad).

3.5.2. *Tamaño de muestra*

Por la naturaleza de la investigación, se simularon muestras balanceadas y desbalanceadas de distintos tamaños, sin excluir posibles nuevos tamaños muestrales que surgieron según las necesidades del estudio. Los tamaños muestrales que hemos empleado se muestran en la Tabla N° 1-3.

Tabla 1-3: Tamaños muestrales

Muestras	Balanceadas					Desbalanceadas				
	n_1		n_2			n_1		n_2		
Dos	3		7			5		5		
	5		15			10		10		
	10		20			15		15		
	20		30			25		25		
Tres	n_1	n_2		n_3		n_1	n_2		n_3	
	3	5		7		5	5		5	
	5	10		15		10	10		10	
	10	15		20		15	15		15	
	15	20		40		25	25		25	
Cinco	n_1	n_2	n_3	n_4	n_5	n_1	n_2	n_3	n_4	n_5
	3	4	5	6	7	5	5	5	5	5
	3	5	7	10	25	10	10	10	10	10
	5	10	15	20	25	15	15	15	15	15
	5	10	20	40	50	25	25	25	25	25

Realizado por : Chávez,L.,2022.

3.5.3. Método de muestreo

Se simularon muestras aleatorias, unas provenientes de una población teórica con distribución normal estándar, y otras que provenían de una población con distribución totalmente desconocida (contaminación de la normalidad).

3.6. Métodos, técnicas e instrumentos de investigación

Puesto que se estimó la TIEP y la Potencia de las pruebas de comparación de medias, ante alteraciones de la normalidad, se creó un algoritmo de simulación sobre la base de Monte Carlo, el cual estima a la TIEP como la proporción de rechazos de H_0 cuando H_0 es cierta, y a la Potencia como la proporción de rechazos de H_0 cuando H_0 es falsa. Para el alejamiento de la Normalidad en las muestras, se utilizó el método de *Fleishman*, obteniendo diferentes niveles de contaminación de la Normalidad, que van desde una contaminación nula hasta una contaminación exagerada. En todos los caso,s se trabajó con muestras teóricamente homocedásticas. En cuanto a los tamaños muestrales, se emplearon tanto muestras balanceadas como desbalanceadas de diferentes tamaños, los cuales se observan en la Tabla N° 1-3.

3.6.1. Obtención de muestras no normales

Para simular las muestras con contaminaciones medibles de la normalidad, como ya lo hemos mencionado anteriormente emplearemos el método de *Fleishman*, sabemos que este procedimiento calcula los coeficientes a, b, c, d , mediante una transformación polinómica que implica diferentes valores del tercer y cuarto momento central, es decir la asimetría (γ_1) y curtosis (γ_2), lo cual permite la deformación medible de la normalidad.

Se definió diferentes niveles de contaminación de la normalidad de acuerdo con distintas combinaciones de la simetría y curtosis estudiadas en investigaciones previas (Blanca, et al., 2013).

Los coeficientes de *Fleishman* para estos niveles de no normalidad fueron calculados usando la función “*fleishman.coef*” existente en el *software* estadístico R, del paquete “BinNonNor” (Inan, et al., 2021). Los valores de asimetría, curtosis y coeficientes de *Fleishman* calculados con sus respectivos niveles de alejamiento de la normalidad, se muestran en la Tabla N° 2-3.

Los niveles de contaminación que van de 0 a 6 representan los alejamientos de la normalidad en diferentes intensidades, en el cual 0 representa una contaminación nula mientras que 6 una contaminación exagerada.

Tabla 2-3: Coeficientes de Fleishman

Nivel de contaminación	Asimetría (γ_1)	Curtosis (γ_2)	Coefficientes de Fleishman (a, b, c, d)
0 (nula)	0	0	(0,1,0,0)
1 (leve)	0.25	0.7	(-0.037, 0.933, 0.037, 0.021)
2 (moderada)	0.75	1	(-0.119, 0.956, 0.119, 0.0098)
3 (alta)	1.30	2	(-0.249, 0.984, 0.249, -0.016)
4 (severa)	2	6	(-0.314, 0.826, 0.314, 0.023)
5 (extrema)	3	15	(-0.346, 0.588, 0.346, 0.086)
6 (exagerada)	4	30	(-0.335, 0.308, 0.335, 0.160)

Fuente: Blanca, et al., 2013.

Realizado por : Chávez, L., 2022.

Una vez que obtenidos los coeficientes de *Fleishman*, se crearon las muestras no normales a partir de una función creada en R, llamada “*rnonorm*” (Ver Anexo C) la misma se muestra a continuación con sus respectivos argumentos.

rnonorm(n, f.coef)

Donde:

n: tamaño de la muestra no normal.

f.coef: coeficientes de *Fleishman* que nos permiten obtener la muestra no normal.

3.6.2. Proceso de simulación para la estimación de la TIEP y Potencia en pruebas de comparación de medias

La estimación de la TIEP y Potencia de las pruebas de comparación de medias, se lo realizó mediante la construcción de una función en R denominada “TIEPPOW”, la cual a través de un algoritmo de simulación sobre la base de Monte Carlo, estima a la TIEP como la proporción de rechazos de H_0 cuando H_0 es cierta ($\mu_1 = \mu_2 = \dots = \mu_k$), y a la Potencia como la proporción de rechazos de H_0 cuando H_0 es falsa ($\mu_i \neq \mu_j$ para algún $i \neq j$), garantizando en el caso de la TIEP el cumplimiento teórico de H_0 , y para el caso de la Potencia el no cumplimiento teórico de H_0 . El algoritmo de simulación consta de los siguientes pasos:

1. Se crearon muestras teóricamente homocedásticas, de distintos tamaños, tanto balanceadas como desbalanceadas, las cuales deben garantizar que H_0 sea teóricamente cierta en el caso de la estimación de la TIEP, o no en el caso de la Potencia.
2. Las muestras no normales son obtenidas gracias a la función “rnorm” que se crearon con anterioridad, la cual a partir de los coeficientes de Fleishman produce las contaminaciones de la normalidad en las muestras.
3. Se aplicaron sobre estas muestras, los test de hipótesis de comparación de medias descritos en el capítulo II, y especificados en la Tabla N° 3-3.
4. Se verificó si el p-value que arroja cada test es menor al nivel de significancia (rechazo de H_0), con la finalidad de guardar en un vector, un valor lógico que indique “TRUE” si se da la situación anterior o “FALSE” en caso contrario.
5. Se repitió el proceso 100000 veces y se determinó la proporción de rechazos de H_0 , definida como la razón entre el total de rechazos de la hipótesis nula y el número de réplicas de simulación.

Tabla 3-3: Funciones utilizadas para las pruebas de comparación de medias

Prueba de hipótesis	Función en R
t-Student	t.test
Mann-Whitney-Wilcoxon (WMW)	wilcox.test
ANOVA unifactorial	oneway.test
Kruskal-Wallis	kruskal.test

Fuente: Team R.

Realizado por : Chávez, L., 2022.

La función “TIEPPOW” (Ver Anexo B) y sus argumentos se muestran a continuación:

TIEPPOW (nSim, n, f.coef, diffMean, alpha)

Donde

nSim: número de réplicas de simulación

n: vector que contiene en sus componentes los tamaños muestrales.

f. coef: coeficientes de *Fleishman* que nos permiten obtener las muestras con contaminaciones de la Normalidad.

diffMean: diferencia en las medias teóricas.

alpha: nivel de significancia empleado en los test de hipótesis.

En esta investigación, los argumentos de la función “TIEPPOW” que se emplearon para obtener las estimaciones de la TIEP y de la Potencia fueron: $nSim = 100000$ réplicas de simulación, tamaños muestrales mostrados en la Tabla N° 1-3, coeficientes de *Fleishman* observados en la Tabla N° 2-3, $diffmean = 0$ usado para la estimación de la TIEP, y $diffmean = 2$ para la estimación de la Potencia. En cuanto al nivel de significancia, el valor α usado en cada una de las pruebas fue de 0.05.

CAPÍTULO IV

4. MARCO DE ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

4.1. Estimación de la TIEP de las pruebas de comparación de medias

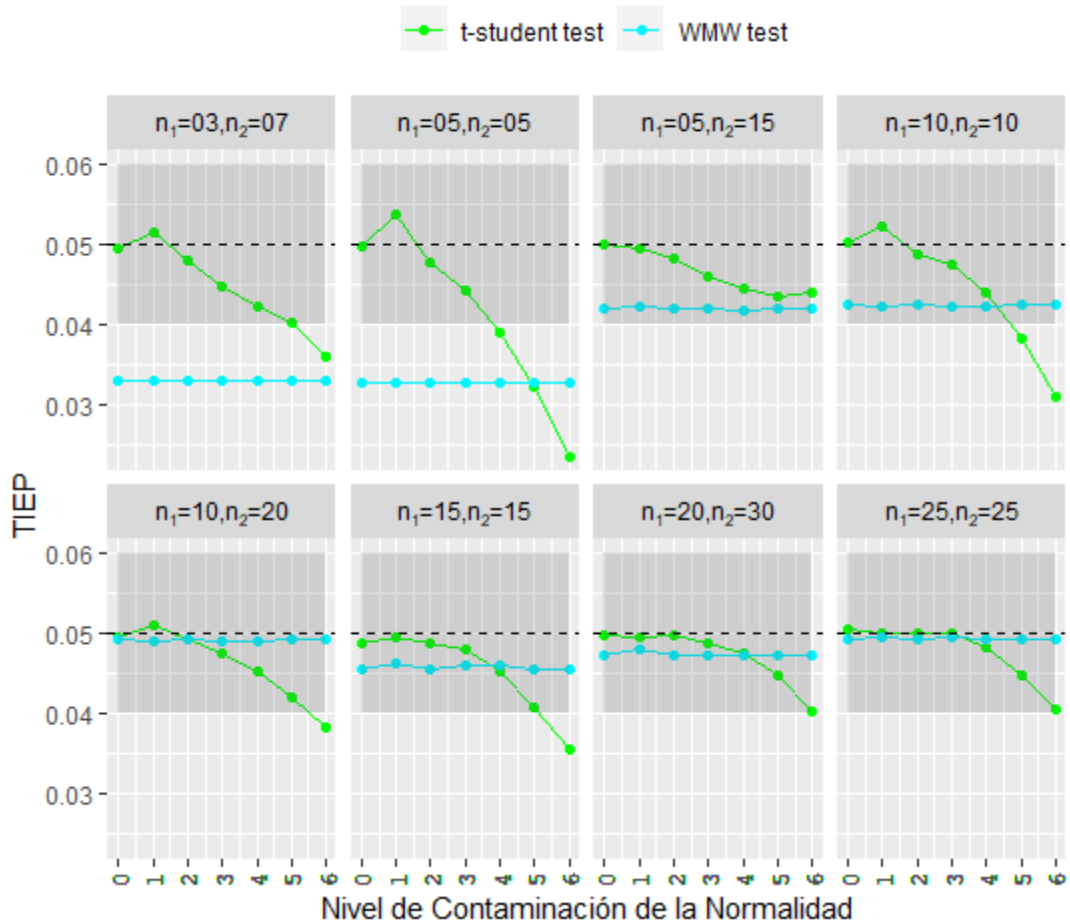


Ilustración 1-4: Estimación de la TIEP de las pruebas de comparación de 2 medias, elaborado en *RStudio*

Realizado por : Chávez, L., 2022.

Para distintos niveles de contaminación de la normalidad que van de cero a seis, la ilustración N° 1-4, muestra la estimación de la TIEP cuando 2 medias poblacionales son comparadas, empleando muestras balanceadas y desbalanceadas.

Se observa que, cuando existe normalidad (nivel de contaminación igual a cero), la TIEP de la *t-Student* es visiblemente mejor que la del test de *Mann–Whitney–Wilcoxon*, puesto que, en todos los casos, la TIEP del test paramétrico se encuentra dentro del criterio de Cochran (regiones sombreadas en la gráfica) y sobre todo muy cercana al nivel de significancia $\alpha = 0.05$. También

debemos aclarar que conforme los tamaños muestrales crecen la TIEP del test de *Wilcoxon* se asemeja mucho a la de la *t-Student*.

Al producirse contaminaciones en la normalidad, también la TIEP de la *t-Student* se controla mejor que la de *Mann-Whitney-Wilcoxon*, pero a partir de un nivel de contaminación igual a cuatro, la TIEP del test no paramétrico empieza a comportarse de mejor manera que la de la *t-Student*. Para muestras grandes que poseen contaminaciones leves de la normalidad, las TIEPs de ambos test respetan el criterio de Cochran y por ende son aceptables, pero aun estando dentro del criterio de Cochran la mejor TIEP es la del test paramétrico, sin embargo al ir aumentando el nivel de contaminación, va disminuyendo la efectividad del test paramétrico de la *t-Student*, en cambio la del test de *Mann-Whitney-Wilcoxon* se mantiene constante en todos los niveles de contaminación sean leves o moderados. En muestras pequeñas la TIEP de la *t-Student* produce mejores valores casi siempre.

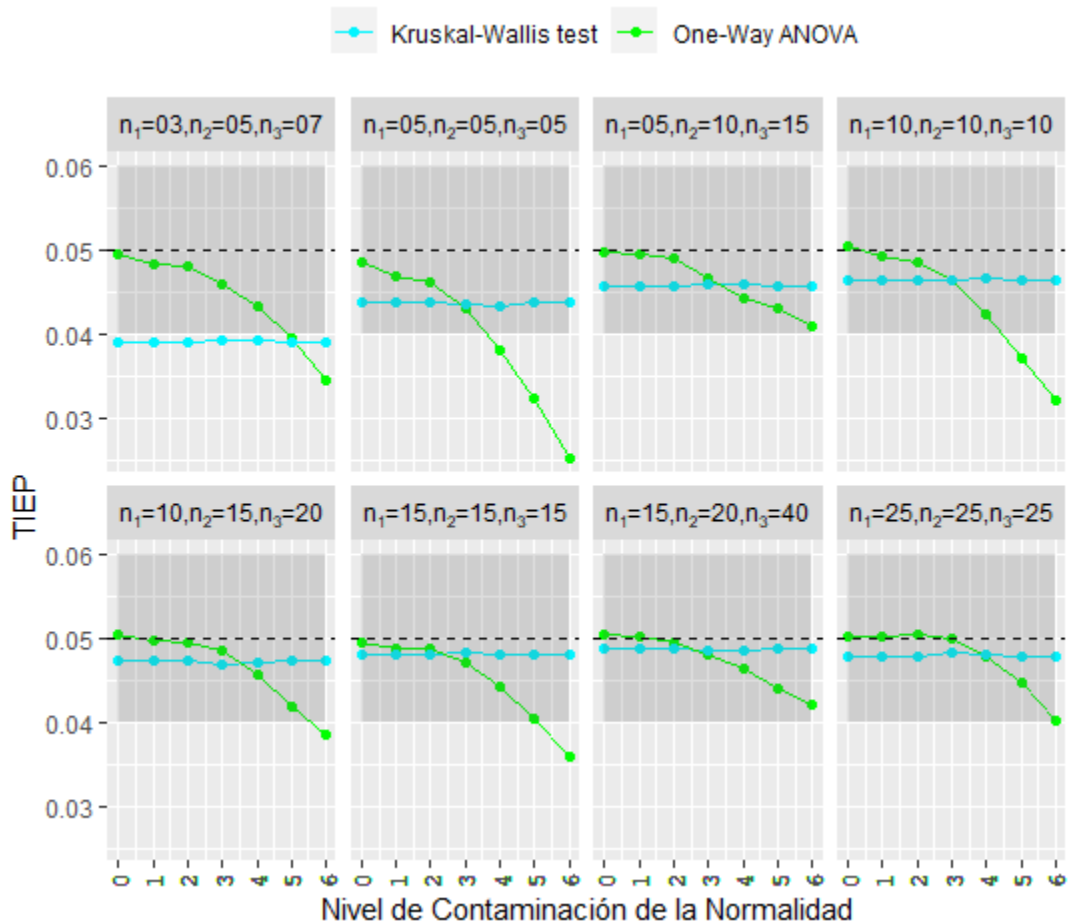


Ilustración 2-4: Estimación de la TIEP de las pruebas de comparación de 3 medias, elaborado en *RStudio*

Realizado por : Chávez, L., 2022.

Ante distintos niveles de contaminación de la normalidad, la ilustración N° 2-4 muestra la estimación de la TIEP cuando 3 medias poblacionales son comparadas utilizando muestras balanceadas y desbalanceadas.

Para muestras normales la TIEP del ANOVA unifactorial se controla bastante bien muy cercana del nivel de significancia. Conforme se aumentan los tamaños en las muestras, la TIEP del test de Kruskal-Wallis se acerca en gran medida a la del ANOVA unifactorial.

Cuando se da las contaminaciones de la normalidad en las muestras, la TIEP del test paramétrico se controla de mejor manera que la de su alternativa no paramétrica, pero todo esto se produce hasta un nivel de contaminación igual a dos, a partir de un nivel tres de contaminación la TIEP del test de *Kruskal-Wallis* muestra un mejor y constante comportamiento que la del ANOVA unifactorial. En muestras balanceadas, tanto pequeñas como grandes y a medida que la contaminación aumenta, la TIEP del test de *Kruskal-Wallis* va mejorando en su valor y se va acercando al nivel de significancia, en contrariedad la TIEP del ANOVA unifactorial empieza a salir del criterio de Cochran.

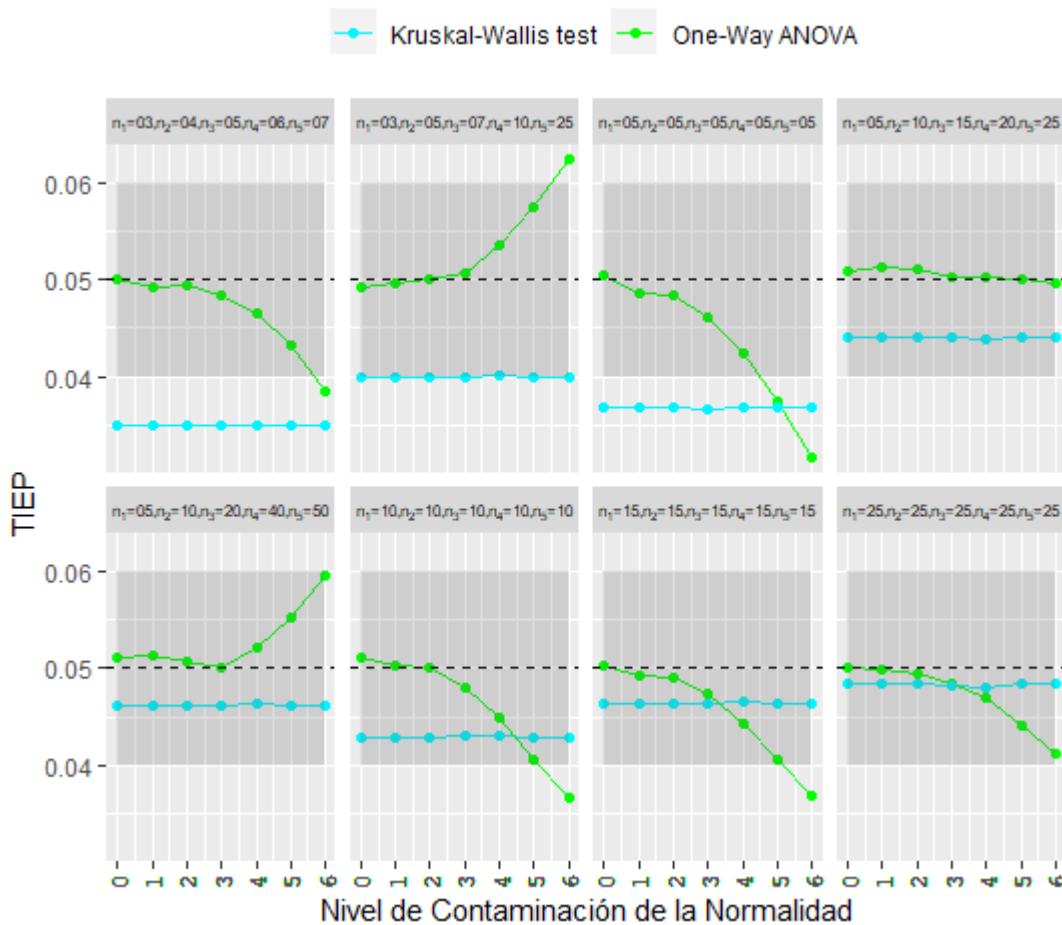


Ilustración 3-4: Estimación de la TIEP de las pruebas de comparación de 5 medias, elaborado en *RStudio*

Realizado por : Chávez, L., 2022.

Para distintos niveles de contaminación de la normalidad, la ilustración N° 3-4 muestra la estimación de la TIEP de las pruebas de comparación de 5 medias poblacionales, utilizando muestras tanto balanceadas como desbalanceadas.

Cuando las muestras son normales, la TIEP del ANOVA unifactorial es mejor que la producida por el test de *Kruskal-Wallis*, puesto que en todos los casos la estimación de este parámetro se encuentra dentro del criterio de Cochran y está muy cercana al nivel de significancia $\alpha=0.05$.

Ante alteraciones de la normalidad, cuando las muestras son pequeñas y desbalanceadas, la TIEP del test paramétrico muestra un mejor comportamiento que su alternativa no paramétrica, esto se da hasta un nivel de contaminación igual a cinco. En el caso balanceado cuando las muestras son grandes y a medida que la contaminación aumenta, la TIEP del test de *Kruskal-Wallis* va mejorando en su valor, en oposición la TIEP del ANOVA unifactorial empieza a salir de la región que representa al criterio de Cochran.

Se observa que en general para las pruebas de comparación de 5 medias, la TIEP del test paramétrico es muy variante puesto que en algunos casos posee un comportamiento creciente y en otros decreciente, mientras que la TIEP del test no paramétrico permanece constante en todos los niveles de contaminación de la normalidad (de 0 a 6).

4.2. Estimación de la Potencia de las pruebas de comparación de medias

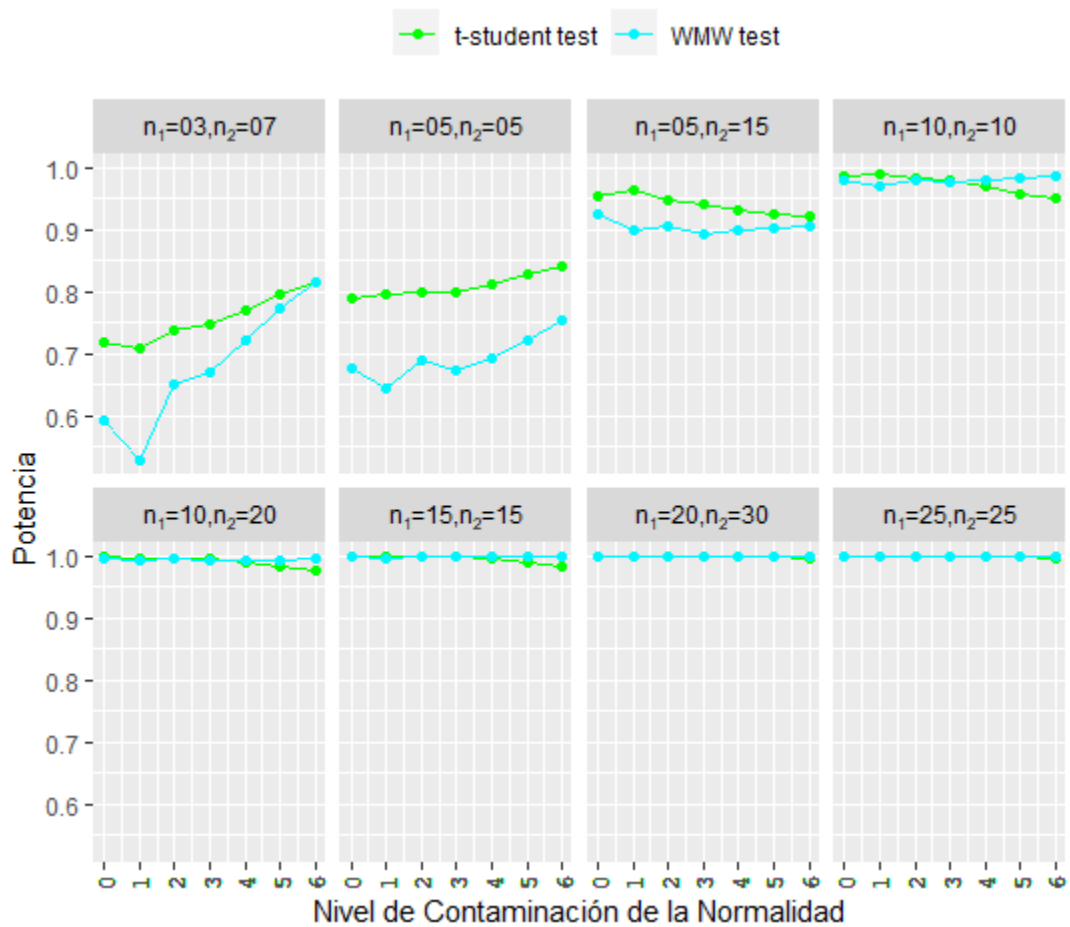


Ilustración 4-4: Estimación de la Potencia de las pruebas de comparación de 2 medias, elaborado en *RStudio*

Realizado por : Chávez, L., 2022.

Para distintos niveles de contaminación de la normalidad que van de cero a seis, la ilustración N° 4-4, muestra la estimación de la Potencia de las pruebas de comparación de 2 medias poblacionales, utilizando muestras tanto balanceadas como desbalanceadas.

Cuando existe normalidad en las muestras, la Potencia de la *t-Student* es mejor que la del test de *Mann–Whitney–Wilcoxon*, pero conforme los tamaños muestrales crecen, las potencias de ambos test se controlan de manera muy similar y sobre todo óptimamente (cerca a 1).

Ante alteraciones de la normalidad, cuando las muestras son pequeñas, la Potencia del test paramétrico es mejor que la de su alternativa no paramétrica, exceptuando cuando los tamaños muestrales son $n_1 = 10, n_2 = 10$, caso en el cual a partir de una contaminación igual a cuatro la Potencia de la prueba de *Wilcoxon* es mejor que la de la *t-Student*. En muestras grandes las potencias de ambos test presentan valores óptimos y muy similares independientemente del nivel de contaminación, y de si son muestras balanceadas o desbalanceadas.

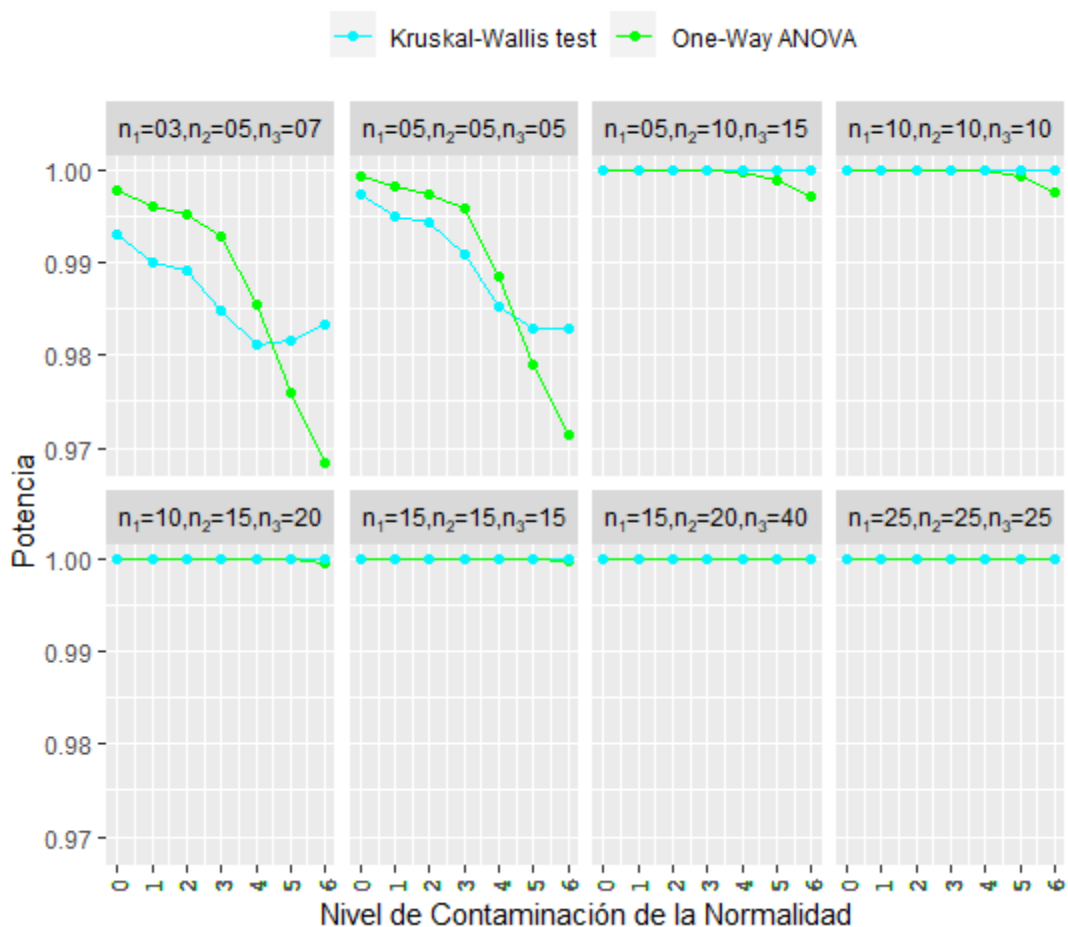


Ilustración 5-4: Estimación de la Potencia de las pruebas de comparación de 3 medias, elaborado en *RStudio*

Realizado por : Chávez, L., 2022.

Ante distintos niveles de contaminación de la normalidad la ilustración N° 5-4 muestra la estimación de la Potencia cuando 3 medias poblacionales son comparadas utilizando muestras balanceadas y desbalanceadas.

Cuando existe normalidad en las muestras, la Potencia del test paramétrico es ligeramente mejor (diferencias de unas milésimas) que la del test no paramétrico, recalando que conforme los tamaños muestrales crecen, las potencias de ambos test muestran un buen y muy parecido control cercano a uno.

Al producirse contaminaciones de la normalidad en las muestras, la Potencia de ambos contrastes de hipótesis poseen un buen y muy parecido comportamiento en la mayoría de los casos, exceptuando cuando los tamaños muestrales son $n_1 = 03, n_2 = 05, n_3 = 07$, y $n_1 = 05, n_2 = 05, n_3 = 05$, situaciones en las cuales la Potencia del ANOVA unifactorial es levemente mejor que la producida por el test de *Kruskal-Wallis*, esto hasta un nivel de contaminación igual a cuatro.

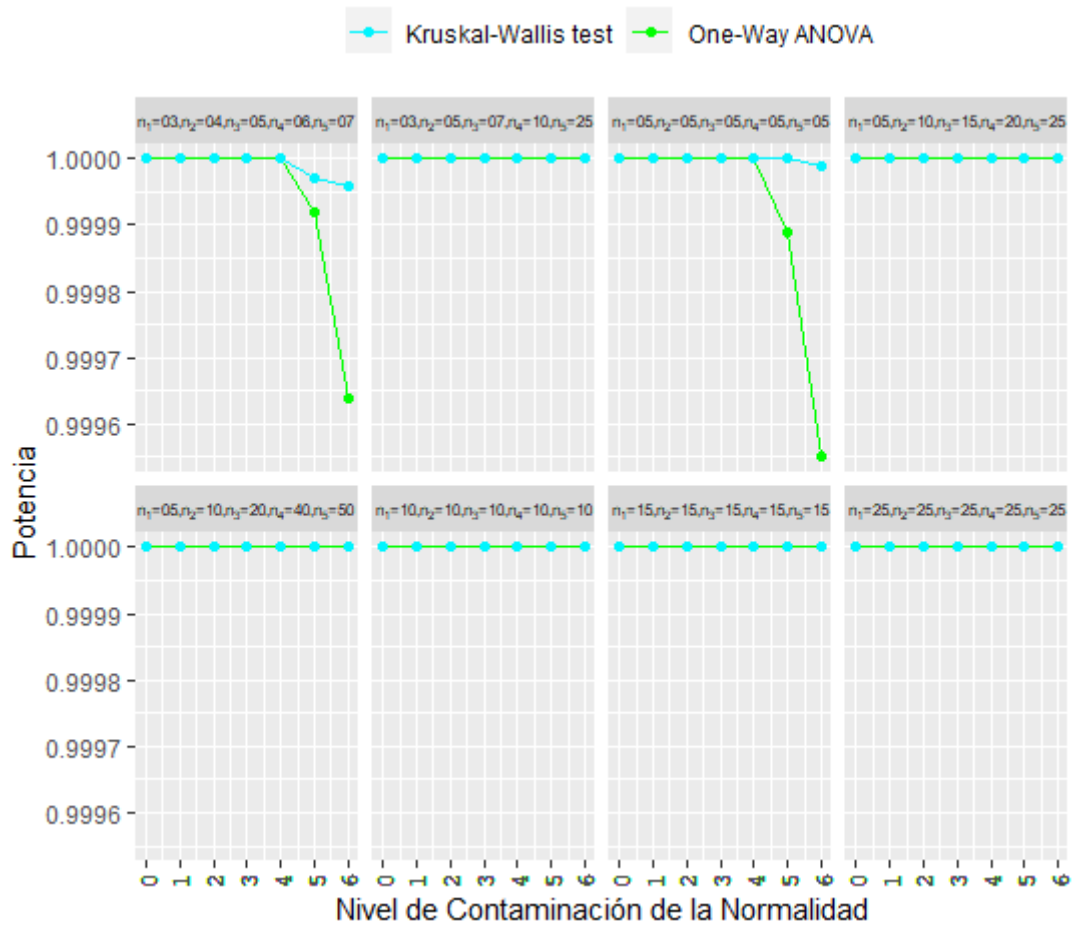


Ilustración 6-4: Estimación de la Potencia de las pruebas de comparación de 5 medias, elaborado en *RStudio*

Realizado por : Chávez, L., 2022.

Para distintos niveles de contaminación de la normalidad, la ilustración N° 6-4 muestra la estimación de la Potencia de las pruebas de comparación de 5 medias poblacionales, utilizando muestras tanto balanceadas como desbalanceadas.

En la mayoría de los casos sean muestras normales o no, la potencia de ambos contrastes es la mejor, es decir toman un valor igual a uno. Hay ligeras excepciones en las cuales, a partir de un nivel de contaminación igual a cinco, la Potencia del test de *Kruskal Wallis* es insignificamente mejor (diferencias irrisorias) que la del ANOVA unifactorial.

CONCLUSIONES

- En el *software R*, se creó la función denominada “rnorm”, mediante lo cual se pudo simular muestras con alteraciones medibles de la normalidad, sobre la base del método de *Fleishman*, dicho método nos permitió de acuerdo a distintas combinaciones de la simetría y curtosis, obtener diferentes niveles de contaminación de la normalidad, que va desde un nivel 0 (contaminación nula) hasta un nivel 6 (contaminación exagerada), y a su vez esta función facilitó la generación de las muestras no normales necesarias para la detección de la TIEP y la Potencia de las pruebas de comparación de medias ante contaminaciones medibles de la normalidad.
- Para la estimación de la TIEP y la Potencia de las pruebas de comparación de medias ante alteraciones medibles de la normalidad, se creó en R, la función llamada “TIEPPOW” la cual a través de un algoritmo de simulación sobre la base de Monte Carlo, estimó a la TIEP como la proporción de rechazos de H_0 cuando H_0 es cierta, y a la Potencia como la proporción de rechazos de H_0 cuando H_0 es falsa, garantizando en el caso de la TIEP el cumplimiento teórico de H_0 , y para el caso de la Potencia el no cumplimiento teórico de H_0 . Cabe aclarar que en todos los casos se trabajaron con muestras homocedásticas y para la contaminación de la normalidad en las mismas, se hizo uso de la función “rnorm” descrita con anterioridad.
- En las pruebas de comparación de dos medias, la prueba de la *t-Student* es robusta ante desviaciones de la normalidad, puesto que controla de buena manera la TIEP (se encuentra alrededor del nivel de significancia) y posee una Potencia alta (cercana a 1), en cambio contrario a lo que se pensaría el test no paramétrico de *Mann-Whitney-Wilcoxon* no controla de buena manera el Error de Tipo I, este patrón deja de funcionar cuando existen grandes contaminaciones de la normalidad (a partir de un nivel igual a cuatro).
- Lo que sucede en las pruebas de comparación de 2 medias no necesariamente ocurre cuando se aumenta el número de medias a comparar, es así que, en las pruebas de comparación de 3 medias, la prueba paramétrica es robusta ante desviaciones de la normalidad siempre y cuando el nivel de contaminación no sea demasiado grande (hasta un nivel igual a 2), puesto que hasta este nivel de contaminación esta prueba controla de mejor manera la TIEP y posee una Potencia alta; a partir de un nivel 3 la prueba no paramétrica empieza a ser la mejor opción, porque si bien el test paramétrico posee una potencia alta en estos niveles ya no controla de buena manera su TIEP.
- Cuando se comparó 5 medias poblacionales en muestras pequeñas y desbalanceadas, el test paramétrico permanece robusto ante desviaciones de la Normalidad hasta un nivel de contaminación (igual a 5), mientras que en muestras grandes y balanceadas y a medida que la contaminación aumenta, el test no paramétrico muestra mayor robustez.

- Los resultados de este estudio indicaron que la TIEP de los test paramétricos son muy variables en función de los componentes externos (tamaño muestral y nivel de contaminación) puesto que, en algunos casos poseen un comportamiento creciente y en otros decreciente, mientras que la TIEP de los test no paramétricos permanecen constantes, es decir en los test no paramétricos su TIEP se comporta siempre de la misma manera sea de forma adecuada (dentro del criterio de Cochran) o inadecuada (fuera del criterio de Cochran), en cambio la TIEP del test paramétrico tiene un comportamiento incierto.
- Conforme se aumentan los tamaños muestrales, la Potencia tanto de los test paramétricos y no paramétricos empiezan a presentar un comportamiento constante de gran similitud, y cercanos o iguales a 1, es decir poseen una Potencia adecuada, sin importar si existe una leve o exagerada contaminación de la normalidad en las muestras, de si son balanceadas o desbalanceadas. Esto se da por la ley de los grandes números, gracias a la cual, a medida que aumentamos los tamaños muestrales, la Potencia empieza a ser constante y toma un valor cercano a 1.
- En síntesis, la TIEP de las pruebas paramétricas de comparación de medias presentan un buen comportamiento hasta ciertos niveles de contaminación de la normalidad, dependiendo el nivel en el cual aún son robustas, del número de medias a comparar y de los tamaños muestrales. En cuanto a la Potencia, la producida por los test paramétricos es mejor que la de sus alternativas no paramétricas, esto también hasta cierto grado de contaminación de la normalidad, pero cuando los tamaños muestrales son grandes la Potencia tanto de los test paramétricos y no paramétricos presentan valores similares y cercanos o iguales a 1

RECOMENDACIONES

- Se sugiere para la realización de estos tipos de estudios -en los cuales se deben emplear algoritmos de simulación- la creación de funciones propias que nos permitan llevar a cabo este fin, dado que dichos algoritmos no se encuentran implementados en ningún *software*.
- Se recomienda que, en las pruebas de comparación de dos medias, para muestras que poseen hasta un nivel de contaminación de normalidad igual a tres (contaminación alta), se utilice el test de la *t-Student* dado que, hasta este nivel la prueba paramétrica controla de buena manera su TIEP y posee una Potencia alta, a partir de un nivel de contaminación cuatro (contaminación severa) la mejor opción es el test no paramétrico de *Mann-Whitney-Wilcoxon*.
- Al comparar 3 medias poblacionales, en muestras con un nivel a lo mucho de contaminación dos (contaminación moderada), se sugiere emplear el ANOVA unifactorial, mientras que a partir de un nivel de contaminación tres, el uso de la prueba no paramétrica de *Kruskal-Wallis* será lo más adecuado.
- Finalmente, en las pruebas de comparación de cinco medias, se aconseja que, cuando las muestras sean pequeñas, desbalanceadas y que presenten hasta un nivel de contaminación cinco (contaminación extrema) utilizar el ANOVA unifactorial; para muestras grandes, balanceadas, y con los últimos niveles de contaminación, la mejor alternativa será la utilización del test no paramétrico.

GLOSARIO

Algoritmo: es un conjunto de instrucciones definidas, ordenadas y delimitadas para resolver un problema en específico (Maluenda , 2021).

Coefficiente de asimetría: es una medida de la asimetría (falta de simetría) de la distribución de un conjunto de datos con respecto a la media. Un valor positivo de este coeficiente significa que la distribución se encuentra sesgada hacia la izquierda. Un resultado negativo implica que la distribución se sesga hacia la derecha (Rincón , 2017, pp. 69-71).

Coefficiente de curtosis: es un coeficiente el cual indica que tan apuntada o achatada se encuentra cierta distribución respecto a una normal. Un valor positivo indica que los datos están muy concentrados hacia la media (distribución leptocúrtica). Un resultado negativo muestra que los datos están muy dispersos (distribución platicúrtica). La distribución normal exige que la curtosis sea igual a 0(distribución mesocúrtica) (Rincón , 2017, pp. 72-74).

Distribución Normal Estándar: Decimos que la variable aleatoria X tiene una distribución normal estándar si tiene una distribución normal con parámetros $\mu = 0$ y $\sigma = 1$ (Rincón, 2014, p. 271).

Muestras heterocedásticas: Son aquellas muestras que tienen respectivamente poblaciones de varianzas diferentes (Corder & Foreman, 2014, p. 1).

Muestras homocedásticas: Son aquellas muestras que tienen respectivamente poblaciones de varianzas aproximadamente iguales (Corder & Foreman, 2014, p. 1).

Potencia: es la probabilidad de que no ocurra el error tipo II es decir es la probabilidad de rechazar la hipótesis nula cuando esta es falsa (Sanabria Brenes, 2011, p. 84).

Probabilidad de Error tipo I: es la probabilidad de rechazar la hipótesis nula cuando esta es cierta (Sanabria Brenes, 2011, p. 84).

Pruebas paramétricas: son aquellas pruebas que asumen distribuciones estadísticas de la variable de interés en la población de estudio, por tanto, deben cumplir ciertos supuestos como lo son que cada muestra deba ajustarse a una distribución aproximadamente normal, ser homocedásticas, y sus observaciones independientes (Corder & Foreman, 2014, p. 2).

Pruebas no paramétricas: se las utiliza cuando no se cumplen los supuestos paramétricos, es decir su aplicación no depende de la distribución de la variable de interés en la población de estudio (Corder & Foreman, 2014, p. 2).

Simulación: es el desarrollo de un modelo lógico-matemático de un sistema, de manera que se obtenga una imitación de un proceso del sistema a través del tiempo (Azofeifa, 2004).

Transformación lineal: Sean V y W espacios vectoriales reales. Una transformación lineal T de V en W es una función que asigna a cada vector $v \in V$ un vector único $Tv \in W$ y que satisface, para cada u y v en V y cada escalar α : $T(u + v) = Tu + Tv$ y $T(\alpha v) = \alpha Tv$ (Grossman & Flores, 2012, p. 481).

BIBLIOGRAFÍA

AMAT RODRIGO, Joaquin. *ANOVA análisis de varianza para comparar múltiples medias* [blog]. [Consulta: 31 enero 2022]. Disponible en: https://www.cienciadedatos.net/documentos/19_anova.html

AMAT RODRIGO, Joaquin. *Test Kruskal-Wallis* [blog]. [Consulta: 30 enero 2022]. Disponible en: https://www.cienciadedatos.net/documentos/20_Kruskal-Wallis_test.html

AMAT RODRIGO, Joaquin. *Test de Wilcoxon Mann Whitney como alternativa al t-test* [blog]. [Consulta: 13 enero 2022]. Disponible en: https://www.cienciadedatos.net/documentos/17_mann-whitney_u_test

AZOFEIFA, Carlos. “Aplicación de la Simulación Monte Carlo en el cálculo del riesgo usando Excel”. *Tecnología en Marcha*, vol. 17, n° 1 (2004), pp. 97-109.

BLANCA MENA, María J.; et al. “Non-normal data: is ANOVA still a valid option”. *Psicothema*, vol. 29, n° 4 (2017), pp. 552-557.

BLANCA MENA, María J.; et al. “Skewness and Kurtosis in Real Data Samples”. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, vol. 9, n° 2 (2013), pp.78-84.

COCHRAN, William G. “The x² correction for continuity”. *Iowa State College Journal of Science*, vol. 16 (1942), pp.421-436.

COHEN, Jacob. “The statistical power of abnormal-social psychological research: A review”. *The Journal of Abnormal and Social Psychology*, vol. 65, n° 3 (1962), pp.145-153.

CORDER, Gregory; & FOREMAN, Dale. *NONPARAMETRIC STATISTICS. A Step-by-Step Approach*. 2ª ed. New Jersey: John Wiley & Sons, Inc., 2014, p.1.

FERRERO, Rosana. *¿Qué es el ANOVA de una vía?* [blog]. [Consulta: 31 enero 2022]. Disponible en: <https://www.maximaformacion.es/blog-dat/que-es-el-anova-de-una-via/>

FISHER, Ronald A. “The Statistical Utilization of Multiple Measurements”. *Annals of Eugenics*, vol. 8, n° 4 (1938), pp. 376-386.

FLEISHMAN, Allen. “A method for simulating non-normal distributions”. *Psychometrika*, vol. 43, n° 4 (1978), pp. 521-532.

GROSSMAN, Stanley; & FLORES, José. *Álgebra lineal*. 7ª ed. México, D.F.: McGRAW-HILL/INTERAMERICANA EDITORES, S.A, 2012, p.481.

INAN, Gul; et al. *BinNonNor: Data Generation with Binary and Continuous Non-Normal Components*. GPL-2 | GPL-3 2021

KRUSKAL, William H. ; & WALLIS, Wilson A. “Use of Ranks in One-Criterion Variance Analysis”. *Journal of the American Statistical Association*, vol. 47, n° 260 (1952), pp. 583-621.

MALUENDA, Raquel. *Profile Software Services* [blog]. [Consulta: 3 julio 2022]. Disponible en: <https://profile.es/blog/que-es-un-algoritmo-informatico/>

MOLINA, David; & LARA, Ana. *Estadística Universidad de Granada* [blog]. [Consulta: 12 enero 2022]. Disponible en: <https://wpd.ugr.es/~bioestad/guia-de-r/practica-6/>

MONTILLA, Josefa M.; & KROMREY, Jeffrey. “Robustez de las pruebas T en comparación de medias, ante violación de supuestos de normalidad y homocedasticidad”. *Ciencia e Ingeniería*, vol. 31, n° 2 (2010), pp. 101-107.

MOYA, Rufino; & SARA VIA, Gregorio. *Probabilidad e Inferencia Estadística*. 2ª ed. Lima: San Marcos, 1988, p.701.

PÉREZ LÓPEZ, César. *Muestreo estadístico. Conceptos y problemas resueltos*. 1ª ed. Madrid: PEARSON EDUCACIÓN, S.A., 2005.

POISSON, Simon D. “Recherches sur la probabilité des jugements en matière criminelle et en matière civile . s.l’”. *Bachelier*, (1837).

RINCÓN, Luis. *Estadística Descriptiva*. 1ª ed. México, D.F.: Facultad de Ciencias, UNAM, 2017.

RINCÓN, Luis. *Introducción a la Probabilidad.* 1ª ed. México, D.F.: Facultad de Ciencias, UNAM, 2014.

RINCÓN, Luis. *Una Introducción a la Estadística Inferencial.* 1ª ed. México, D.F.: Facultad de Ciencias, UNAM, 2019.

SANABRIA BRENES, Giovanni. *Comprendiendo la Estadística Inferencial.* 1ª ed. Cartago: Editorial Tecnológica de Costa Rica, 2011.

SÁNCHEZ TURCIOS, Reinaldo A. “t-Student: Usos y abusos”. *Revista mexicana de cardiología*, vol. 26, n° 1 (2015), pp. 59-61.

STONEHOUSE, John; & FORRESTER, Guy. “Robustness of the t and U tests under combined assumption violations”. *Journal of Applied Statistics*, vol. 25, n° 1 (1998), pp. 63-74.

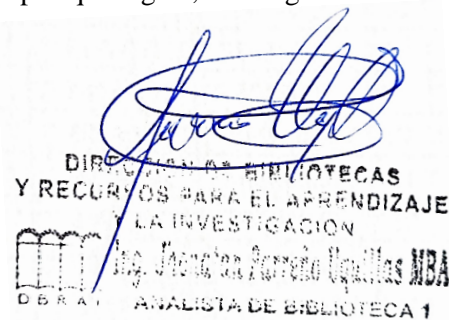
STUDENT. “The Probable Error of a Mean”. *Biometrika*, vol. 6, n° 1 (1908), pp. 1-25.

VAN HECKE, Tanja. “Power study of anova versus Kruskal-Wallis test”. *Journal of Statistics and Management Systems*, vol. 15, n° 3 (2012), pp. 241-247.

WACKERLY, Dennis; et al. *Estadística matemática con aplicaciones.* 7ª ed. México, D.F.: Grupo Editorial Iberoamérica, 2010.

WILCOXON, Frank. “Individual comparisons by ranking methods”. *Breakthroughs in statistics. s.l.:Springer*, (1992), pp. 196-202.

ZAMBRANO, María; et al. Estudio comparativo in vivo de la efectividad del paracetamol vs el ibuprofeno en el control del dolor postquirúrgico, en alargamientos de corona.2015.



ANEXOS

ANEXO A: TABLA DE RESULTADOS DE LA ESTIMACIÓN DE LA TIEP Y POTENCIA DE LAS PRUEBAS DE COMPARACIÓN DE MEDIAS

Nivel de Contaminación	n_1	n_2	TIEP t-student test	TIEP WMW test	POWER t-student test	POWER WMW test
0	3	7	0.04952	0.03295	0.71805	0.59193
	5	15	0.04993	0.04203	0.95525	0.92455
	10	20	0.04957	0.04914	0.99861	0.99781
	20	30	0.04987	0.04737	1.00000	1.00000
	5	5	0.04963	0.03267	0.79134	0.67557
	10	10	0.05035	0.04248	0.98811	0.98081
	15	15	0.04878	0.04560	0.99963	0.99891
	25	25	0.05052	0.04932	1.00000	0.99999
1	3	7	0.05143	0.03307	0.70847	0.52979
	5	15	0.04938	0.04218	0.96446	0.90057
	10	20	0.05104	0.04892	0.99786	0.99482
	20	30	0.04961	0.04800	0.99978	0.99999
	5	5	0.05370	0.03283	0.79735	0.64403
	10	10	0.05228	0.04234	0.99123	0.97100
	15	15	0.04950	0.04616	0.99838	0.99799
	25	25	0.04990	0.04962	0.99973	1.00000
2	3	7	0.04808	0.03295	0.73667	0.65026
	5	15	0.04831	0.04203	0.94728	0.90534
	10	20	0.04919	0.04914	0.99741	0.99581
	20	30	0.04966	0.04737	0.99997	0.99999
	5	5	0.04784	0.03267	0.79912	0.68940
	10	10	0.04884	0.04248	0.98375	0.98076
	15	15	0.04864	0.04560	0.99909	0.99899
	25	25	0.05007	0.04932	1.00000	0.99999
3	3	7	0.04471	0.03308	0.74759	0.67181
	5	15	0.04606	0.04197	0.94125	0.89413
	10	20	0.04749	0.04911	0.99624	0.99291
	20	30	0.04883	0.04719	0.99995	0.99999
	5	5	0.04423	0.03285	0.79905	0.67511
	10	10	0.04760	0.04223	0.98049	0.97636
	15	15	0.04806	0.04596	0.99852	0.99857
	25	25	0.05010	0.04948	0.99999	0.99999
	3	7	0.04221	0.03305	0.77026	0.72109
	5	15	0.04460	0.04170	0.93306	0.89813

4	10	20	0.04535	0.04910	0.99092	0.99333
	20	30	0.04758	0.04721	0.99986	0.99997
	5	5	0.03887	0.03277	0.81182	0.69301
	10	10	0.04392	0.04231	0.96912	0.97863
	15	15	0.04516	0.04593	0.99536	0.99873
	25	25	0.04835	0.04917	0.99991	1.00000
5	3	7	0.04021	0.03295	0.79518	0.77478
	5	15	0.04358	0.04203	0.92661	0.90395
	10	20	0.04192	0.04914	0.98360	0.99440
	20	30	0.04468	0.04737	0.99893	0.99998
	5	5	0.03234	0.03267	0.82775	0.72234
	10	10	0.03830	0.04248	0.95800	0.98244
	15	15	0.04087	0.04560	0.98927	0.99910
	25	25	0.04482	0.04932	0.99924	1.00000
6	3	7	0.03611	0.03295	0.81460	0.81414
	5	15	0.04398	0.04203	0.92349	0.90598
	10	20	0.03829	0.04914	0.97625	0.99538
	20	30	0.04038	0.04737	0.99705	1.00000
	5	5	0.02362	0.03267	0.84250	0.75423
	10	10	0.03098	0.04248	0.94972	0.98690
	15	15	0.03548	0.04560	0.98242	0.99942
	25	25	0.04049	0.04932	0.99783	1.00000

Nivel de Contaminación	n_1	n_2	n_3	TIEP One-Way ANOVA	TIEP Kruskal-Wallis test	POWER One-Way ANOVA	POWER Kruskal-Wallis test
0	3	5	7	0.04948	0.03901	0.99790	0.99307
	5	10	15	0.04977	0.04577	1.00000	1.00000
	10	15	20	0.05043	0.04738	1.00000	1.00000
	15	20	40	0.05044	0.04873	1.00000	1.00000
	5	5	5	0.04856	0.04379	0.99931	0.99745
	10	10	10	0.05052	0.04626	1.00000	1.00000
	15	15	15	0.04952	0.04814	1.00000	1.00000
	25	25	25	0.05027	0.04788	1.00000	1.00000
1	3	5	7	0.04829	0.03901	0.99602	0.99012
	5	10	15	0.04939	0.04577	1.00000	1.00000
	10	15	20	0.04981	0.04738	1.00000	1.00000
	15	20	40	0.05018	0.04873	1.00000	1.00000
	5	5	5	0.04681	0.04379	0.99815	0.99506
	10	10	10	0.04921	0.04626	1.00000	1.00000

	15	15	15	0.04875	0.04814	1.00000	1.00000
	25	25	25	0.05028	0.04788	1.00000	1.00000
2	3	5	7	0.04800	0.03901	0.99511	0.98907
	5	10	15	0.04900	0.04577	1.00000	1.00000
	10	15	20	0.04944	0.04738	1.00000	1.00000
	15	20	40	0.04952	0.04873	1.00000	1.00000
	5	5	5	0.04619	0.04379	0.99739	0.99425
	10	10	10	0.04852	0.04626	1.00000	1.00000
	15	15	15	0.04884	0.04814	1.00000	1.00000
	25	25	25	0.05037	0.04788	1.00000	1.00000
3	3	5	7	0.04594	0.03912	0.99278	0.98484
	5	10	15	0.04653	0.04594	1.00000	0.99999
	10	15	20	0.04853	0.04694	1.00000	1.00000
	15	20	40	0.04814	0.04856	1.00000	1.00000
	5	5	5	0.04306	0.04347	0.99586	0.99086
	10	10	10	0.04639	0.04632	1.00000	1.00000
	15	15	15	0.04720	0.04827	1.00000	1.00000
	25	25	25	0.05003	0.04821	1.00000	1.00000
4	3	5	7	0.04320	0.03916	0.98553	0.98108
	5	10	15	0.04420	0.04580	0.99983	0.99995
	10	15	20	0.04574	0.04707	1.00000	1.00000
	15	20	40	0.04641	0.04864	1.00000	1.00000
	5	5	5	0.03804	0.04327	0.98848	0.98521
	10	10	10	0.04220	0.04655	0.99998	0.99999
	15	15	15	0.04420	0.04800	1.00000	1.00000
	25	25	25	0.04784	0.04803	1.00000	1.00000
5	3	5	7	0.03941	0.03901	0.97604	0.98152
	5	10	15	0.04302	0.04577	0.99889	0.99990
	10	15	20	0.04198	0.04738	0.99994	1.00000
	15	20	40	0.04405	0.04873	1.00000	1.00000
	5	5	5	0.03226	0.04379	0.97901	0.98281
	10	10	10	0.03703	0.04626	0.99939	0.99999
	15	15	15	0.04041	0.04814	0.99998	1.00000
	25	25	25	0.04482	0.04788	1.00000	1.00000
6	3	5	7	0.03450	0.03901	0.96847	0.98325
	5	10	15	0.04099	0.04577	0.99715	0.99989
	10	15	20	0.03852	0.04738	0.99960	1.00000
	15	20	40	0.04214	0.04873	1.00000	1.00000
	5	5	5	0.02525	0.04379	0.97139	0.98278
	10	10	10	0.03197	0.04626	0.99764	0.99997
	15	15	15	0.03592	0.04814	0.99980	1.00000

	25	25	25	0.04014	0.04788	1.00000	1.00000		
Nivel de Contaminación	n_1	n_2	n_3	n_4	n_5	TIEP One-Way ANOVA	TIEP Kruskal-Wallis test	POWE R One-Way ANOVA	POWE R Kruskal Wallis test
0	3	4	5	6	7	0.05006	0.03508	1.00000	1.00000
	3	5	7	10	25	0.04914	0.03989	1.00000	1.00000
	5	10	15	20	25	0.05094	0.04401	1.00000	1.00000
	5	10	20	40	50	0.05108	0.04605	1.00000	1.00000
	5	5	5	5	5	0.05038	0.03687	1.00000	1.00000
	10	10	10	10	10	0.05115	0.04281	1.00000	1.00000
	15	15	15	15	15	0.05016	0.04631	1.00000	1.00000
	25	25	25	25	25	0.05015	0.04850	1.00000	1.00000
1	3	4	5	6	7	0.04929	0.03508	1.00000	1.00000
	3	5	7	10	25	0.04961	0.03989	1.00000	1.00000
	5	10	15	20	25	0.05117	0.04401	1.00000	1.00000
	5	10	20	40	50	0.05138	0.04605	1.00000	1.00000
	5	5	5	5	5	0.04868	0.03687	1.00000	1.00000
	10	10	10	10	10	0.05032	0.04281	1.00000	1.00000
	15	15	15	15	15	0.04926	0.04631	1.00000	1.00000
	25	25	25	25	25	0.04991	0.04850	1.00000	1.00000
2	3	4	5	6	7	0.04936	0.03508	1.00000	1.00000
	3	5	7	10	25	0.05001	0.03989	1.00000	1.00000
	5	10	15	20	25	0.05104	0.04401	1.00000	1.00000
	5	10	20	40	50	0.05069	0.04605	1.00000	1.00000
	5	5	5	5	5	0.04837	0.03687	1.00000	1.00000
	10	10	10	10	10	0.05014	0.04281	1.00000	1.00000
	15	15	15	15	15	0.04899	0.04631	1.00000	1.00000
	25	25	25	25	25	0.04941	0.04850	1.00000	1.00000
3	3	4	5	6	7	0.04828	0.03500	1.00000	1.00000
	3	5	7	10	25	0.05067	0.03989	1.00000	1.00000
	5	10	15	20	25	0.05019	0.04405	1.00000	1.00000
	5	10	20	40	50	0.04996	0.04607	1.00000	1.00000
	5	5	5	5	5	0.04619	0.03659	1.00000	1.00000
	10	10	10	10	10	0.04795	0.04310	1.00000	1.00000
	15	15	15	15	15	0.04740	0.04635	1.00000	1.00000
	25	25	25	25	25	0.04835	0.04815	1.00000	1.00000
	3	4	5	6	7	0.04649	0.03509	1.00000	1.00000
	3	5	7	10	25	0.05356	0.04012	1.00000	1.00000
	5	10	15	20	25	0.05028	0.04390	1.00000	1.00000
	5	10	20	40	50	0.05219	0.04643	1.00000	1.00000

4	5	5	5	5	5	0.04245	0.03682	1.00000	1.00000
	10	10	10	10	10	0.04484	0.04308	1.00000	1.00000
	15	15	15	15	15	0.04421	0.04660	1.00000	1.00000
	25	25	25	25	25	0.04688	0.04806	1.00000	1.00000
5	3	4	5	6	7	0.04332	0.03508	0.99992	0.99997
	3	5	7	10	25	0.05740	0.03989	1.00000	1.00000
	5	10	15	20	25	0.04993	0.04401	1.00000	1.00000
	5	10	20	40	50	0.05510	0.04605	1.00000	1.00000
	5	5	5	5	5	0.03753	0.03687	0.99989	1.00000
	10	10	10	10	10	0.04063	0.04281	1.00000	1.00000
	15	15	15	15	15	0.04059	0.04631	1.00000	1.00000
	25	25	25	25	25	0.04413	0.04850	1.00000	1.00000
6	3	4	5	6	7	0.03858	0.03508	0.99964	0.99996
	3	5	7	10	25	0.06232	0.03989	1.00000	1.00000
	5	10	15	20	25	0.04962	0.04401	1.00000	1.00000
	5	10	20	40	50	0.05960	0.04605	1.00000	1.00000
	5	5	5	5	5	0.03162	0.03687	0.99955	0.99999
	10	10	10	10	10	0.03676	0.04281	1.00000	1.00000
	15	15	15	15	15	0.03688	0.04631	1.00000	1.00000
	25	25	25	25	25	0.04123	0.04850	1.00000	1.00000

ANEXO B: FUNCIÓN TIEPPOW

```
TIEPPOW<- function(nSim,n, f.coef, diffMean , alpha = 0.05){
  npatron<-n
  df<-diffMean
  nt<-length(npatron)
  nobs<-sum(npatron)
  facts <- factor(rep(1:nt, npatron))
  vec<-if(df==0){
    rep(0,nt)
  }else{
    seq(0,(df*nt)-df,df)
  }
  mu <- rep(vec,npatron)

  pvals <- replicate(nSim, {
    muestra <- rnonorm(nobs, f.coef) + mu
    if(nt==2){
      c('t-test' = t.test(muestra ~ facts, var.equal = T)$p.value,
        'wilcoxon-test' = wilcox.test(muestra ~ facts)$p.value)
    }else{
      c('one-way anova' = oneway.test(muestra ~ facts, var.equal = T)$
p.value,
        'kruskal-test' = kruskal.test(muestra ~ facts)$p.value)
    }
  })
  rowSums(pvals < alpha) / nSim
}
```

ANEXO C: FUNCIÓN rnonorm

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



epoch

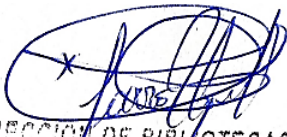
Dirección de Bibliotecas y
Recursos del Aprendizaje

UNIDAD DE PROCESOS TÉCNICOS Y ANÁLISIS BIBLIOGRÁFICO Y
DOCUMENTAL

REVISIÓN DE NORMAS TÉCNICAS, RESUMEN Y BIBLIOGRAFÍA

Fecha de entrega: 09 / 12 / 2022

INFORMACIÓN DEL AUTOR/A (S)
Nombres – Apellidos: Luis Reinaldo Chávez Yuqui
INFORMACIÓN INSTITUCIONAL
Facultad: Ciencias
Carrera: Estadística
Título a optar: Ingeniero en Estadística
f. Analista de Biblioteca responsable: Ing. Leonardo Medina Ñuste MSc.


DIRECCIÓN DE BIBLIOTECAS
Y RECURSOS PARA EL APRENDIZAJE
Y LA INVESTIGACIÓN
Ing. Leonardo Medina Ñuste MBA
DBRA1 ANALISTA DE BIBLIOTECA 1

2180-DBRA-UTP-2022