RELEVANCIA DE LOS TESTS ESTADÍSTICOS t Y F EN COMPARACIÓN DE MEDIAS PARA MUESTRAS INDEPENDIENTES

RELEVANCE OF STATISTICAL TESTS: t AND F, TO COMPARE MEANS FOR INDEPENDENT SAMPLES

Montilla, Josefa María
Área de Estadística.
Departamento de Ciencias Económicas Administrativas y Contables.
Universidad de Los Andes, Trujillo-Venezuela.
josefam76@cantv.net

Resumen

Los estadísticos t y F, también denominados métodos paramétricos constituyen un set de procedimientos matemáticos que permiten determinar el nivel de significancia estadística de una observada diferencia entre las medias poblacionales independientes. Estos procedimientos dependen básicamente de una serie de supuestos básicos como: normalidad de la distribución poblacional, homogeneidad de varianzas poblacionales e independencia de las observaciones. Sin embargo, frecuentemente nos encontramos con una serie de datos que no satisfacen esos supuestos, lo cual indica que existe la necesidad de identificar cuáles tests estadísticos son robustos a la violación de los supuestos fundamentales de los tests. Por consiguiente, esta investigación explora y describe fundamentalmente los estadísticos t de Student y F de Fisher-Snedecor, examinando sus supuestos básicos, el grado de robustez de los mismos a violaciones de los supuestos básicos de normalidad y homogeneidad de varianzas. Al mismo tiempo, describe algunos procedimientos designados con el nombre de no-paramétricos y robustos respecto a la violación de los supuestos, los cuales son considerados como alternativas en el procedimiento de comparación de medias, tales como el t' de Welch y t* de Yuen; F'-Folded, F*-Brown-Forsythe y F*.O'Brien Finalmente, se destacan las diferencias más relevantes entre los definidos tests.

Palabras clave: t de Student, F de Fisher-Snedecor, Prueba Paramétrica, Prueba No-Paramétrica, Robustez de un estadístico.

Abstract

The statistics t and F constitute a set of mathematical procedures that allow the determination of the level of statistical significance of one observed difference among the population means. These procedures depend basically on a series of basic assumptions such as: normality of the population distribution, homogeneity of population variances and independence of the observations. However, frequently we meet with a series of data that don't satisfy those assumptions, which indicates the existence of a need to identify which statistical tests are robust to the violation of the fundamental assumptions of the t-test. Consequently, this investigation fundamentally explores and describes the statistics t of Student and F of Fisher-Snedecor, examining their basic assumptions, the degree of robustness from the same ones to violations of normality and homogeneity of variances assumptions. At the same time, it describes some procedures designated with the name of non-parametric and robust regarding the violation of the assumptions, those procedures are considered as alternatives to comparison of means, as the t ' of Welch and t * of Yuen; F'-Folded, F*-Brown-Forsythe y F*.O'Brien. Finally, it highlights the most relevant differences among the tests defined.

Key Words: Student's t; Fisher-Snedecor's F, Parametric test, Non-Parametric test, Robustness of a statistic.

Recibido: 07-04-10 / Aprobado: 19-06-10

Introducción

Las estadísticas son técnicas matemáticas, de fácil manejo, utilizadas para analizar datos numéricos en cumplimiento de varios propósitos. Estas técnicas son objetivamente consideradas como elementos de apoyo que cuantifican y evalúan los resultados de una investigación, es decir, permiten investigar el comportamiento de diversos fenómenos de orden económico, social, político, educacional, biológicos u otros. Estos fenómenos de carácter estadístico deben reunir ciertos requisitos para poder ser observados; así, sus características deben ser cuantificables en forma tal que permitan determinar la intensidad con la cual se producen.

En muchas situaciones de toma de decisiones se requiere determinar si los parámetros de dos o más poblaciones son semejantes o diferentes. Así sucede, por ejemplo, cuando se contrastan hipótesis respecto a dos o más medias poblacionales independientes, donde surge la necesidad de decidir en relación a la selección de la prueba. En este caso particular, tales decisiones están encaminadas a probar la hipótesis respecto a la igualdad de dos o más medias poblacionales independientes. Estos procedimientos son aquéllos que forman parte de la inferencia estadística, la cual constituye una de las dos categorías principales del procedimiento estadístico.

Las pruebas de significancia estadística se dividen en paramétricas y no-paramétricas. Las paramétricas se refieren a los procedimientos en los cuales, en los cálculos, se recurre a alguna distribución de probabilidad. En ellas la prueba debe cumplir una serie de condiciones, tales como que las observaciones o mediciones sean de una población normalmente distribuida (normalidad); que las observaciones o mediciones sean independientes entre sí; que la selección de un caso no influya en la de otro (independencia); que las poblaciones tengan la misma varianza (homocedasticidad); y que las variables analizadas tengan el nivel de medición interval o proporcional.

Las pruebas no-paramétricas o de distribución libre son aquéllas que no especifican la distribución normal de la muestra; la diferencias entre estos procedimientos estadísticos radica en el nivel de medición, es decir las pruebas específicas a utilizar dependen de si los datos están en nivel de medición nominal, ordinal, interval o proporcional.

En este sentido, entre los métodos y procedimientos estadísticos más ampliamente conocidos y usados por los investigadores, para determinar el nivel de significancia estadística de una observada diferencia entre las medias, destacan los estadísticos t y F. Estos estadísticos identificados como pruebas paramétricas, constituyen un set de procedimientos matemáticos; y su uso en investigación depende básicamente de los supuestos básicos de los mismos, que en este caso son similares. Así, para producir conclusiones válidas a partir de un conjunto dado de datos es necesario satisfacer el conjunto de condiciones fundamentales de estos estadísticos. Sin embargo, frecuentemente nos encontramos con una serie de datos que no satisfacen esos supuestos, por lo que es necesario identificar y describir cuáles de estos tests son robustos a la violación de los supuestos establecidos.

Esta investigación esta dirigida, básicamente, a examinar y describir específicamente los procedimientos estadísticos paramétricos t de Student y F de Fisher-Snedecor, los cuales han sido diseñados para probar la igualdad de dos o más medias poblacionales independientes, respectivamente. Sin embargo, por cuanto el efecto de violación de los supuestos básicos de estos estadísticos se transforma en un punto fundamental en esta investigación, es necesario examinar también las alternativas de robustez de los tests para muestras independientes. Consecuentemente, esta investigación examina como tests alternativos al test t de Student, a los procedimientos estadísticos t'-Welch y t*- Yuen t; y como tests alternativos al test F de Fisher-Snedecor, a los procedimientos estadísticos F'-Folded, F*-Brown-Forsythe y F*.O'Brien. Estos estadísticos alternativos son denominados procedimientos no-paramétricos, robustos y poderosos.

Descripción del Estadístico t de Student

El estadístico "t", que para muestras normales provenientes de una población que tiene una variable contínua en un intervalo infinito, recibe el nombre t de Student y fue desarrollado por William Gosset (1876-1937) en el año 1908. Este procedimiento es considerado como una técnica paramétrica que requiere de un gran número de supuestos. La diferencia entre los diferentes estadísticos "t" es establecida por sus grados de libertad. Así, mientras más largos sean los grados de libertad menor es la diferencia entre la distribución t de Student y la distribución normal.

El estadístico t fue desarrollado, al igual que otros t-tests incondicionales, como una función de la magnitud de las diferencias entre las medias; también es una distribución de probabilidad que se define en función de los grados de libertad. Este tiene una distribución con $n_1 + n_2 - 2$ grados de libertad, así, para la hipótesis nula de no diferencia (Ho: $\mu_1 = \mu_2$) el t estadístico o t-test es establecido como el cociente de "la diferencia entre el estadístico observado y el hipotetizado parámetro" y "el error estándar de estimación de la diferencia". Sin embargo, bajo el supuesto de homogeneidad de las varianzas ($\sigma_1^2 = \sigma_2^2$), el error estándar de estimación de la diferencia esta basado en la varianza combinada estimada, S^2p, la cual es considerada como más precisa, ya que se espera que ésta contenga menor error de muestreo que cualquiera de las otras varianzas muestrales, S_1^2 ó S_2^2 (Glass & Hopkins, 1996). Así, la expresión general del estadístico t de Student viene dada por:

$$t = \left(\overline{X}_1 - \overline{X}_2\right) \Big/ \sqrt{S_p^2 (1/n_1 + 1/n_2)}$$

de donde:

$$S_p^2 = \left[(n_1 + 1)S_1^2 + (n_2 + 1)S_2^2\right] \Big/ n_1 + n_2 - 2$$

S^2_p es la varianza combinada estimada, la cual es expresada como un promedio ponderado de las varianzas muestrales, superior al promedio aritmético, que da igual ponderación a las varianzas muestrales. Estos promedios (ponderado y aritmético) coinciden cuando las medias son de igual tamaño.

El t-test es ampliamente usado para probar varias hipótesis, la más común es la hipótesis de la igualdad de dos medias poblacionales cuando muestras independientes son seleccionadas aleatoriamente de dos poblaciones que son normalmente distribuidas y tienen varianzas iguales. Este estadístico t es matemáticamente equivalente al F-test (ANOVA) cuando el número de grupos es igual a dos.

El uso de este procedimiento en investigación depende básicamente de los supuestos básicos. Glass y Hopkins (1996) han destacado que el estadístico t de Student está basado en las siguientes condiciones*:*

1) Normalidad: Este supuesto establece que la variable X_1 es distribuida normalmente con media μ_1 y varianza σ_1^2 y la variable X_2 también es distribuida normalmente con media

μ_2 y varianza σ_2^2. El estadístico t de Student es robusto con respecto a la violación del supuesto de normalidad, cuando las tamaños de muestras son grandes, lo cual indica que existe una insignificante diferencia entre las tasas de error tipo I nominal y observada o empírica. Favorablemente, muchas investigaciones (Glass, Peckham & Sanders, 1972), han revelado que la violación al supuesto de normalidad no ha tenido consecuencias prácticas en el uso del estadístico t.

2) Homogeneidad de varianzas: Esta asunción indica que las dos varianzas poblacionales son iguales ($\sigma_1^2 = \sigma_2^2$). La t de Student es robusta con respecto a la violación del supuesto de homogeneidad de varianzas cuando los tamaños de nuestras son iguales ($n_1 = n_2$). Así mismo, Glass, Peckham y Sanders (1972) han señalado que para prácticos propósitos no es necesario probar el supuesto de homogeneidad de varianzas cuando los tamaños de muestras son iguales.

3) Independencia de las observaciones: Esta condición simplemente establece que las observaciones dentro o entre los dos grupos no son dependientes, y se asume que una muestra de tamaño n_1 es obtenida aleatoriamente de la población 1 y una muestra independiente de tamaño n_2 es seleccionada aleatoriamente de la población 2. Esta asunción puede ser satisfecha a través del uso de una técnica de aleatorización. Sin embargo, Glass y Hopkins (1996) han señalado que en algunas situaciones es difícil evaluar si esta condición de independencia de las observaciones es reunida.

Robustez del estadístico t de Student

La robustez de un procedimiento estadístico puede ser entendida como la insensibilidad del mismo a desviaciones de sus asunciones o supuestos. En el caso del estadístico t, fundamentalmente es la insensibilidad respecto al no cumplimiento de los supuestos de normalidad y homogeneidad de varianzas.

Para evaluar la robustez del procedimiento estadístico t de Student, investigadores como Boneau, 1960 y Scheffé, 1959; han utilizado la Técnica de Simulación Monte Carlo, con la finalidad de comparar las tasas de error tipo I (nominal y observada) y el poder estadístico del test; es decir, la probabilidad de erróneamente concluir que la magnitud de las medias poblacionales es desigual y la probabilidad de correctamente concluir que la magnitud de las medias poblacionales es desigual, respectivamente.

Boneau (1960) estudiando el grado de robustez de la t de Student a violaciones de los supuestos de normalidad y homogeneidad de varianzas, encontró que en aquellos casos donde las varianzas poblacionales son desiguales, la relación entre la tasa observada de error tipo I (τ) y la tasa nominal de error tipo I (α) está influenciada por el tamaño de la muestra. Así, cuando las tamaños de muestras son iguales ($n_1 = n_2$) y suficientemente grandes, los valores de τ y α son aproximadamente iguales. Por otra parte, Scheffé (1959, p.339) también ha mostrado que el estadígrafo t de Student es aproximado a la distribución normal estandarizada, cuando las dos poblaciones son normales y con varianzas iguales.

En contraste, cuando los tamaños de muestras son desiguales ($n_1 \neq n_2$) y las varianzas desiguales ($\sigma^2_1 \neq \sigma^2_2$), se ha demostrado que la t de Student no siempre es asintóticamente aproximada a la normal. Consecuentemente, Boneau (1960), Scheffé (1959), y otros más, han demostrado que cuando los tamaños de muestras son desiguales ($n_1 \neq n_2$) y la muestra más grande es seleccionada de la población con la más grande varianza, llamada condición positiva, el estadígrafo t de Student es conservador, lo cual indica que τ es menor que α ($\tau < \alpha$). Este resultado por consiguiente: 1) incrementa el valor de la varianza ponderada, 2) incrementa el denominador de t, 3) disminuye el valor calculado de t y 4) disminuye el poder de la t de Student para detectar una diferencia entre las medias.

Inversamente, cuando la muestra más grande es obtenida de la población con la menor varianza, llamada condición negativa, el estadígrafo t de Student es liberal, lo cual indica que τ es mayor que α ($\tau > \alpha$). Stevens (1980) ofrece dos razones por las que un test con bajo poder para detectar una diferencia es un resultado no deseado: 1) con un alto poder, el investigador tiene más posibilidades de encontrar una verdadera diferencia si ésta existe, 2) el conocimiento de este poder incrementa la habilidad del investigador para interpretar correctamente los resultados no significantes. Por lo tanto, considerando estas razones, el investigador debe prestarle una mayor atención al poder del test que a la significancia estadística.

Estadísticos t´ de Welch y t* de Yuen como Alternativas al t de Student

En algunos casos es frecuente encontrar que la variable objeto de estudio carece de normalidad o no puede ser totalmente asumida por un tamaño muestral reducido; en estos casos se acostumbra recurrir a la transformación de la variable, con la finalidad de simetrizar su distribución, o utilizar procedimientos estadísticos como los t-test no-paramétricos, que solucionen tal situación

El estadístico t de Student es un procedimiento que está basado en asunciones que no siempre son comprobadas por los datos que se disponen, pero es uniformemente una de las pruebas más poderosas cuando se trata de muestras independientes seleccionadas en forma aleatoria de poblaciones distribuidas normalmente y con varianzas iguales. Sin embargo, bajo la violación de los supuestos de normalidad y homogeneidad de varianzas existe un set de tests denominados robustos que pueden ser considerados como alternativas en el procedimiento de comparación de medias en muestras independientes, bajo diferentes condiciones.

En el presente estudio se describen algunos de estos tests estadísticos, que son muy fáciles de implementar y además han sido utilizados en la práctica, tales como los estadísticos t´ de Welch y t* de Yuen. Estos t-tests han sido desarrollados con la finalidad de probar si la distribución de una variable X es igual en dos poblaciones independientes (Ho: $\mu_1 = \mu_2$), sin requerir el supuesto de la homogeneidad de varianzas poblacionales; o si ésta tiende a ser mayor o menor en alguno de los dos grupos. Sin embargo, cabe destacar que estas alternativas pueden tener un poder similar a ese de la t de Student, siempre y cuando prevalezcan las condiciones de normalidad y homogeneidad de varianzas.

Estadístico t’ de Welch

El test t’-Welch es una de las pruebas que tiene un poder estadístico y un control de las tasas de error tipo I, similares a la t de Student para muestras independientes, siempre y cuando las varianzas poblacionales sean iguales ($\sigma^2_1 = \sigma^2_2$); y un mejor control de las tasas de error tipo I cuando las varianzas poblacionales son desiguales ($\sigma^2_1 \neq \sigma^2_2$). El test de Welch es frecuentemente denotado como t’ y se concibe como una variable aleatoria distribuida según la t de Student, cuyos grados de libertad se obtienen mediante la expresión grados de libertad f. Este test fue descrito por Welch (1947) mediante la siguiente expresión:

$$t' = \left(\overline{X}_1 - \overline{X}_2\right) \Big/ \sqrt{S_1^2 / n_1 + S_2^2 / n_2}$$

de donde: X y S^2 son las medias y la varianzas muestrales.

La prima (`) indica que el estadígrafo t’-Welch no se distribuye estrictamente como la t de Student, sino como una aproximación de la misma con f grados de libertad expresados en función de S_1^2, S_2^2 y α. Estos grados de libertad se redondean al entero más próximo y son definidos como sigue:

$$f = \left[\left(s_1^2/n_1 + s_2^2/n_2\right)\right]^2 \Big/ \left\{\left[\left(s_1^2/n_1\right)^2/n_1 - 1\right] + \left[\left(s_2^2/n_2\right)^2/n_2 - 1\right]\right\}$$

Los grados de libertad así obtenidos oscilan entre un mínimo, que será el valor más pequeño de $n_1 - 1$ y $n_2 - 1$ y un máximo que es $n_1 + n_2 - 2$.

En la práctica, el estimador de f es obtenido por reemplazar parámetros por estadísticos, por ejemplo, S^2 reemplaza a σ^2. El test t'de Welch también está influenciado por el tamaño de las muestras, por tanto, cuando los tamaños de muestras son iguales ($n_1 = n_2$), los resultados de t de Student y t'-Welch son similares, sin embargo, cuando los tamaños de muestras son desiguales ($n_1 \neq n_2$), el t'-Welch se transforma en un test superior a la t de Student para muestras independientes. Scheffé (1970) concluyó que el t'-Welch es una solución práctica y satisfactoria al problema de Behrens-Fisher (un nombre dado al problema de cómo comparar dos medias independientes cuando no se puede asegurar la homogeneidad de varianzas poblacionales).

Estadístico t* de Yuen

El test t*-Yuen es una prueba de diferencia de dos medias propuesta y evaluada por Yuen (1974) con la finalidad de probar la hipótesis de igualdad de dos medias (Ho: $\mu_1 = \mu_2$), sin requerir el supuesto de la homogeneidad de varianzas poblacionales. Yuen ha sugerido remplazar el t'-Welch por un procedimiento basado sobre *medias reducidas (trimmed) y varianzas ajustadas (winsorized)*. La media reducida indica el valor medio de un parámetro para una muestra de datos ignorando los n valores extremos. Estos valores "n" determinan la cantidad de valores reducidos. La reducción de una muestra de datos consiste en remover los n miembros que tienen los n/2 valores más largos y los n/2 más pequeños de un parámetro dado. La media ajustada tiene un procedimiento similar al de la media reducida, pero en lugar de retirar los "n" valores extremos, éstos son reemplazados por dos de los restantes valores extremos.

El estadístico t* de Yuen es equivalente al test de Welch cuando no hay reducción de observaciones y es definido mediante la siguiente expresión:

$$t^* = \left(\bar{X}_{t1} - \bar{X}_{t2}\right) \Big/ \left[\left(S_{w1}^2/n_1 - 2k_1\right) + \left(S_{w2}^2/n_2 - 2k_2\right)\right]^{1/2}$$

de donde: el subíndice "t" se refiere a las medias reducidas y el subíndice "w" denota las varianzas ajustadas.

Este procedimiento es conocido como t*-Yuen trimmed test, y se distribuye aproximadamente como la t de Student, con ft grados de libertad correspondientes a la muestra reducida y definidos de la siguiente manera:

$$f_t = \left[\left(S_{w1}^2/h_1\right) + \left(S_{w2}^2/h_2\right)\right]^2 \Big/ \left[\left(S_{w1}^2/h_1\right)^2/h_1 - 1\right] + \left[\left(S_{w2}^2/h_2\right)^2/h_2 - 1\right]$$

de donde: $h_1 = n_1 - 2k_1$ y $h_2 = n_2 - 2k_2$

Medias reducidas (trimmed) y ajustadas (winsorized)

Las medias reducidas y ajustadas son conocidas como estimadores robustos de localización y escala. En los casos donde las varianzas son desiguales o cuando las observaciones presentan algunos valores extremos, estas medias representan los estimadores robustos de las medias poblacionales. Estas medidas son relativamente insensitivas a los valores extremos, por lo cual, son métodos utilizados para reducir los efectos de outliers o valores extremos en la muestra, que son los que afectan la homogeneidad de las varianzas.

La k-ésima media reducida es calculada después que las k observaciones más pequeñas y más grandes son retiradas de la muestra. Este procedimiento señala que las observaciones son reducidas en cada uno de los extremos. Para obtener esta media se aplica la siguiente fórmula:

$$\overline{X}_{tk} = (1/n - 2k)\sum X_i$$

De donde:

t: indica reducida

k: las observaciones

$X_{(i)}$: indica el i-ésimo orden estadístico, cuando las observaciones son arregladas en orden ascendente: $X_{(1)} \leq X_{(2)} \leq \ldots\ldots\ldots \leq X_{(n)}$

La k-ésima media ajustada es calculada después que las k observaciones más pequeñas son reemplazadas por las (k + 1)- ésima observaciones más pequeñas, y las k observaciones mas grandes son reemplazadas por las (n - k)- ésima observaciones más grandes, lo cual indica que las observaciones son reemplazadas y usadas en los cálculos. Para obtenerla se aplica la siguiente fórmula:

$$\overline{X}_{wk} = 1/n\left[(k+1)X_{k+1} + \sum X_i + (k+1)X_{n-k}\right]$$

De donde:

el subíndice "w" es ajustada

n: indica el número de observaciones

$X_{(i)}$: indica el i- ésimo orden estadístico cuando las observaciones son arregladas en orden ascendente: $X_{(1)} \leq X_{(2)} \leq \ldots\ldots\ldots \leq X_{(n)}$

Cuando la distribución es simétrica, estas medias *reducidas y ajustadas* son los estimadores insesgados de la media poblacional. Sin embargo, ellas no se distribuyen normalmente aunque los datos sean de poblaciones normales. En estos casos la distribución *t* reducida* o la t* ajustada pueden ser aproximadas a la distribución t de Student con n - 2k - 1 grados de libertad (Tukey & MnLaughlin 1963).

Varianzas ajustadas

La varianza ajustada representa un estimador robusto de la varianza poblacional. Esta varianza es define de la siguiente manera:

$$S_w^2 = \left[(k+1)(X_{k+1} - \overline{X}_w)^2 + (X_{k+2} - \overline{X}_w)^2 + \ldots + (X_{n-k-1} - \overline{X}_w)^2 + (k+1)(X_{n-k} - \overline{X}_w)^2)\right] \Big/ n-2k$$

Diferencias entre los t-tests analizados:

Las diferencias que se pueden destacar de los estadísticos descritos anteriormente son las siguientes:

1) En el caso de la t de Student de muestras independientes, la condición de homogeneidad de varianzas permite que las dos varianzas muestrales, S_1^2 y S_2^2, sean ponderadas al estimar el error estándar de las diferencia de medias. En contraste, la t'-Welch y la t*-Yuen usan una varianza estimada separada, debido a que la varianza ponderada no tiene sentido si el supuesto de homogeneidad de varianzas ha sido violado.

2) La t de Student tiene una distribución t con $n_1 + n_2 - 2$ grados de libertad, mientras que la t'-Welch tiene una distribución aproximada a la t de Student con f grados de libertad, definidos en función de las varianzas muestrales y la t*-Yuen tiene una distribución también aproximada a la t de Student con ft grados de libertad definidos en función de las varianzas ajustadas.

Distribución F de Fisher-Snedecor

Como en el caso de los t tests para dos muestras, la homogeneidad de las varianzas también es una asunción clave para conducir la validación de un test en análisis de varianza, para comparar tres o más medias. En estos casos, los tests de homogeneidad deben ser conducidos antes de llevar a cabo el análisis de varianza, con la finalidad de legitimar dicha asunción.

La distribución F de Snedecor es una distribución de probabilidad que se define en función de los grados de libertad y surge como el cociente de dos variables chi-cuadradas independientes divididas por sus grados de libertad.

$$F = X^2 v_1 / v_2 \Big/ X^2 v_2 / v_1$$

Así la razón que sigue una distribución F con v_1 grados de libertad del numerador y v_2 grados de libertad del denominador es relativamente robusta y muy sensible a la violación de normalidad, cuando es unimodal y los tamaños de las muestras son aproximadamente iguales. En este caso, el F test puede resultar con insignificantes p-values. La distribución F se utiliza principalmente en inferencia estadística.

Para inferencias respecto a tres o más medias poblacionales, mediante la técnica estadística análisis de varianza (ANOVA), el estadístico de prueba es la razón F del cuadrado medio de los tratamientos (CMt) y el cuadrado medio del error (CME):

$$F = CMT \Big/ CME$$

Estadísticos F'-Folded, F*-Brown-Forsythe y F*-O'Brien como Alternativas al F tests

Los tests de Folded, Brown-Forsythe, y O'Brien son considerados no-paramétricos y robustos. Los mismos han sido desarrollados con la finalidad de probar la hipótesis de igualdad de medias independientes, cuando los supuestos de normalidad y homogeneidad de las varianzas no son satisfechos, es decir, cuando la distribución de los datos dentro de cada uno de los grupos difiere significativamente de la distribución normal debido al error de muestreo. Además, cabe destacar que si las n observaciones dentro de cada una de las poblaciones consideradas fueron obtenidas aleatoriamente, ellas son independientes. Estos tests, considerados como alternativa al F-test tradicional han demostrado tener un buen control de la tasa de error tipo I para una amplia variedad de distribuciones poblacionales y un buen poder, especialmente para poblaciones leptocurticas.

El F'-*test de Folded* es una prueba de dos colas que no específica cuál de las varianzas es más grande. Este test es definido mediante la siguiente expresión:

$$Folded\,\mathrm{t} = \max(Var_1, Var_2) \big/ \min(Var_1, Var_2)$$

Donde:

$$Var_1 = S_1(n_1 - 1)$$
$$Var_2 = S_2(n_2 - 1)$$

Brown y Forsythe (1974) y O'Brien (1978) propusieron respectivos tests que son básicamente modificaciones del test propuesto por Levene (1960), el cual es considerado como el test estándar de homogeneidad de varianzas. Estos tests usan una variable transformada llamada de dispersión, BF_X y OB_X, respectivamente.

El F*-*test de Brown–Forsythe* es un procedimiento basado sobre un test estadístico, donde el numerador y denominador tienen el mismo valor esperado bajo la hipótesis nula. Ellos han sugerido usar la variable BF_X, la cual indica las desviaciones absolutas de la mediana. En este caso la mediana es utilizada en los cálculos de las desviaciones absolutas, como una alternativa de la media. Este test ha sido recomendado por varios investigadores (Olejnik & Algina, 1987) para distribuciones leptocurticas, el cual provee adecuados valores en las tasas de probabilidad del Error tipo I y excelente poder; además este test constituye un fuerte recurso si existen varios grupos o algunos grupos son muy grandes. Así, la variable BF_X es definida como:

$$BFX = \left| X_{ij} - m_i \right|$$

Donde: m_i es la mediana del i-th grupo.

El *test de Brown –Forsythe* es definido de la siguiente manera:

$$BF\_t = \left(BF\_X_1 - BF\_X_2\right) \Big/ \sqrt{\left[(S\_BFX_1 + S\_BFX_2)\right] / \left[(n_1 + n_2 - 2)(1/n_1 + 1/n_2)\right]}$$

El F*-*test de O'Brien* es un procedimiento basado sobre una variable transformada diferente, la cual es computada para cada observación. Este test es recomendado para distribuciones asimétricas y platicurticas (O'Brien, 1978, 1981; Olejnik & Algina, 1987). El F*-*test de O'Brien* es definido de la siguiente forma:

$$OB\_t = \left(OB\_X_1 - OB\_X_2\right) \Big/ \sqrt{\left[(S\_OBX_1 + S\_OBX_2)\right] / \left[(n_1 + n_2 - 2)(1/n_1 + 1/n_2)\right]}$$

Este test sugiere usar la variable OB_X, la cual es definida de la siguiente manera:

$$OB\_X = \left[(W + n_i - 2) n_i (X_{ij} - X_i)^2 - W(n_i - 1) S_i^2\right] \big/ (n_i - 1)(n_i - 2)$$

Donde:

i: en este estudio toma los valores 1 y 2 (número de grupos independientes)

S_i^{2}: representa las varianzas muestrales

W = .05, este valor es sugerido por O'Brien, el cual empareja la supuesta kurtosis de la distribución.

BF_t y OB_t son distribuidos aproximadamente como una F-test (ANOVA), con G – 1, y g_2 grados de libertad definidos así:

$$g_2 = \left[\sum (1 - n_i / N) S_i^2\right]^2 \Big/ \sum \left\{\left[\left(1 - n_i / N\right) S_i^2\right]^2 / n_i - 1\right\}$$

Donde:

g: número de grupos independientes

$N = \sum n_i$

S_i^2 = representa las varianzas muestrales

El F-test conocido con el nombre de Fisher-Snedecord, es uniformemente más poderoso e invariable a las transformaciones lineales cuando muestras aleatorias independientes son seleccionadas de poblaciones distribuidas normalmente y con varianzas iguales. Bajo esas mismas condiciones, los poderes de los tests de Brown-Forsythe, Folded y O'Brien son similares a la F-Snedecord. Sin embargo, cuando las varianzas poblacionales son desiguales, el poder del test de F-Snedecord no es relevante debido a la falta de control de la tasa de error tipo I, por lo tanto, los test de Brown-Forsythe, Folded y O'Brien son estadísticamente más poderosos, debido a una función de relación entre medias y varianzas poblacionales. Sin embargo, cuando estos tests rechazan la asunción de la homogeneidad de varianzas al nivel de significación del 1 %, Milliken y Johnson (1984) recomiendan usar el Welch's ANOVA en lugar del usual ANOVA, en comparación de Medias.

Referencias Bibliográficas:

Boneau, C. A. (1960). The effects of violations of assumptions underlying the t-test. *Psychological Bulletin*, *57*, 49-64.

Brown, M. & Forsythe, A. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*, 69, 364-367.

Glass, G. V. & Hopkins, K. D. (1996). *Modern elementary statistics*. Octavia Edition. Englewood Cliffs, NJ: Prentice Hall.

Glass, G. V., Peckham, P. D. & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237-88.

Levene, H. (1960). Robust tests for equality of variances, in contributions to probability and statistics. Ed. I. Olkin, Palo Alto, C.A.: Stanford University Press.

Milliken, G.A. & Johnson, D.E. (1984). Analysis of messy data, volume I: *Designed experiments*. Belmont, CA: Lifetime Learning Publications.

O'Brien, R.G. (1978). Robust techniques for testing heterogeneity of variances effects in factorial designs. *Psychometrik*a, 43, 327-344.

O'Brien, R.G. (1981). A simple test for variances effects in experimental designs. *Psychological Bulletin*, 89, 570-574.

Olejnik, S.F. & Algina, J. (1987). Type I error rates and power estimates of selected parametric and non-parametric tests of scale. *Journal of Educational Statistics*, 12, 45-61.

Scheffe, H. (1959). *The analysis of variance.* New York: Wiley.

Scheffe, H. (1970). Practical solutions to the Behrens-Fisher problem. *Journal of the American Statistical Association, 65*, 1051-1058.

Stevens, J. P. (1980). Power of the multivariate analysis of variance tests. *Psychological Bulletin, 88*, 728-737.

Tukey, J. W. & McLaughlin, D. H. (1963). Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming /Winsorization I. *Sankhya, Ser. A., 25*, 3, 331-352.

Welch, B. L. (1947). The generalization of Students problem when several different population variances are involved. *Biometrika*, *34*, 23-35.

Yuen, K. K. (1974). The two-sample Trimmed t for unequal population variances. *Biometrika, 61*, 1, 165-170.