

COMPARACIÓN DE ALGUNAS PRUEBAS ESTADÍSTICAS ASINTÓTICAS DE NO-INFERIORIDAD PARA DOS PROPORCIONES INDEPENDIENTES

COMPARISON OF SOME NON-INFERIORITY ASYMPTOTIC STATISTICAL TESTS FOR TWO INDEPENDENT PROPORTIONS

Félix Almendra-Arao¹, David Sotres-Ramos^{2*}

¹UPIITA del Instituto Politécnico Nacional. Avenida Instituto Politécnico Nacional 2580. 07340 México, D. F. (falmendra@ipn.mx). ²Colegio de Postgraduados, Km. 36.5. Carretera México- Texcoco. 56230. Montecillo, México. (sotres.davida@kendle.com)

RESUMEN

En este trabajo se compararon las pruebas asintóticas de no-inferioridad de Blackwelder, Farrington-Manning, Böhning-Viwatwongkasen, Hauck-Anderson; la prueba de razón de verosimilitudes y dos variantes de estas pruebas con base en sus niveles de significancia reales y en sus potencias. La prueba de Farrington-Manning resultó con la mejor aproximación del nivel de significancia real al nivel nominal para tamaños de muestra $30 \leq n \leq 100$ y para los tres límites de no-inferioridad más frecuentemente utilizados en la práctica. Además, la potencia de la prueba de Farrington-Manning fue muy similar a las potencias de aquellas pruebas con buena aproximación del nivel de significancia real al nominal.

Palabras clave: No-inferioridad, proporción, prueba asintótica.

INTRODUCCIÓN

Las pruebas estadísticas asintóticas de no-inferioridad se utilizan muy frecuentemente en ensayos clínicos. Estas pruebas sirven para demostrar que una terapia nueva (con menores efectos secundarios o menor costo) no es sustancialmente inferior en eficacia a la estándar (Chen *et al.*, 2000).

El objetivo de este trabajo fue comparar las pruebas asintóticas para no-inferioridad de Blackwelder, Farrington-Manning, Böhning-Viwatwongkasen, Hauck-Anderson, la prueba de razón de verosimilitudes y dos variantes de estas pruebas con base en sus niveles de significancia reales y en sus potencias.

Se han reportado algunas comparaciones de pruebas para no-inferioridad, pero están basadas en simulaciones o en aproximaciones gruesas de los niveles de significancia exactos y de las potencias (Tu, 1997; Chen *et al.*, 2000; Li y Chuang-Stein, 2006). En la literatura consultada no se encontró análisis comparativo de pruebas de no-inferioridad donde sean calculados los niveles de significancia y de las potencias.

*Autor responsable ♦ Author for correspondence.

Recibido: Mayo, 2008. Aprobado: Enero, 2009.

Publicado como ARTÍCULO en *Agrociencia* 43: 163-172, 2009.

ABSTRACT

In this study, the asymptotic tests of non inferiority of Blackwelder, Farrington-Manning, Böhning-Viwatwongkasen and Hauck-Anderson, as well as the likelihood ratio test and two variants of these tests were compared based on their real significance levels and their powers. The Farrington-Manning test turned out to be the one with the best approximation to the real significance level to the nominal level for sample sizes $30 \leq n \leq 100$ and for the three most frequently used non-inferiority margins. In addition, the power of the Farrington-Manning test was very similar to the power of those tests with good approximation to the real significance level of the nominal level.

Key words: Non-inferiority, proportion, asymptotic test.

INTRODUCTION

Non-inferiority asymptotic statistical tests are frequently used in clinical trials. These tests are used to demonstrate that a new therapy (with less secondary effects or less cost) is not substantially inferior in efficacy to the standard one (Chen *et al.*, 2000).

The objective of this paper was to compare the Blackwelder, Farrington-Manning, Böhning-Viwatwongkasen and Hauck-Anderson tests, the likelihood ratio test and two variants of these non-inferiority asymptotic tests based on their real significance levels and powers.

Some comparisons for non-inferiority tests have been reported, but they are based on simulations or gross approximations to the exact significance levels and powers (Tu, 1997; Chen *et al.*, 2000; Li and Chuang-Stein, 2006). In the consulted literature no comparative studies were found between non-inferiority tests based on the calculation of the significance levels and the powers of the tests. This is probably due to the long computing time required to calculate those values. In the present study the Barnard convexity condition and the property of symmetry on the same tail (Röhmel, 2005) were utilized to calculate the

Esto probablemente se debe al muy extenso tiempo de cómputo requerido para calcular dichos valores. En el presente trabajo se usaron las condiciones de convexidad de Barnard y de simetría en la misma cola (Röhmel, 2005), para calcular los niveles de significancia y las potencias de las pruebas y compararlas para los tamaños de muestra $30 \leq n \leq 100$, así como para los tres límites de no-inferioridad más frecuentemente utilizados.

MATERIALES Y MÉTODOS

Pruebas estadísticas consideradas

En este trabajo se utilizó el modelo estándar (Bernoulli) para comparar dos tratamientos con base en una variable dicotómica. Este modelo supone que las observaciones correspondientes al primer tratamiento (el estándar) provienen de una muestra aleatoria $\{X_{11}, X_{12}, \dots, X_{1n_1}\}$ de una distribución Bernoulli con probabilidad de éxito p_1 y que las observaciones del segundo tratamiento (el nuevo o experimental) provienen de otra muestra aleatoria $\{X_{21}, X_{22}, \dots, X_{2n_2}\}$ de una distribución Bernoulli con probabilidad de éxito p_2 ; se supone además que estas dos muestras son independientes. La hipótesis de interés (hipótesis de no-inferioridad) que se desea probar es la alternativa (H_a) en el siguiente juego de hipótesis:

$$[H_0 : d \geq d_0] \text{ vs. } [H_a : d < d_0] \quad (1)$$

donde, $d = p_1 - p_2$ y d_0 es el límite de no-inferioridad, que es una constante positiva y conocida. En el contexto de ensayos clínicos los valores usuales para d_0 son 0.10, 0.15 y 0.20.

Seis de las estadísticas de prueba consideradas son del tipo:

$$T(X_1, X_2) = \frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}} \quad (2)$$

donde, $X_1 = \sum X_{1j}$, $X_2 = \sum X_{2j}$, y $\hat{p}_i = \frac{X_i}{n_i}$ es el estimador de máxima verosimilitud de p_i para $i = 1, 2$ y $\hat{\sigma}$ es un estimador consistente de la desviación estándar de $\hat{d} = \hat{p}_1 - \hat{p}_2$; la séptima estadística es aquella para la prueba de razón de verosimilitudes:

$$\lambda(X_1, X_2) = \frac{\sup_{\Theta_0} L(d | (X_1, X_2))}{\sup_{\Theta} L(d | (X_1, X_2))} \quad (3)$$

La diferencia entre las seis estadísticas del tipo (2) radica en la estimación elegida para la desviación estándar de $\hat{d}$. Se consideran seis estimadores:

significance levels and the power of these tests and compare them for sample sizes of $30 \leq n \leq 100$, as well as for the three most frequently utilized non-inferiority margins.

MATERIALS AND METHODS

Statistical tests considered

In this paper the standard model (Bernoulli) was used to compare two treatments based on a dichotomous variable. This model assumes that the observations that correspond to the first treatment (the standard) come from a random sample $\{X_{11}, X_{12}, \dots, X_{1n_1}\}$ following the Bernoulli distribution with success probability p_1 , and that the observations from the second treatment (the new or experimental one) come from another random sample $\{X_{21}, X_{22}, \dots, X_{2n_2}\}$ following the Bernoulli distribution with a success probability p_2 ; in addition, it is assumed that these two samples are independent. The hypothesis of interest (hypothesis of non-inferiority), which we wish to prove, is the alternative one (H_a) in the following set of hypotheses:

$$[H_0 : d \geq d_0] \text{ vs. } [H_a : d < d_0] \quad (1)$$

where $d = p_1 - p_2$ and d_0 is the non-inferiority margin, which is a positive and known constant. In the context of clinical trials, the values most frequently used for d_0 are 0.10, 0.15 and 0.20.

Six of the test statistics considered are of the type:

$$T(X_1, X_2) = \frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}} \quad (2)$$

where $X_1 = \sum X_{1j}$, $X_2 = \sum X_{2j}$, and $\hat{p}_i = \frac{X_i}{n_i}$ is the maximum likelihood estimator of p_i for $i = 1, 2$ y $\hat{\sigma}$ is a consistent estimator of the standard deviation of $\hat{d} = \hat{p}_1 - \hat{p}_2$; the seventh test statistic is the one corresponding to the likelihood ratio test:

$$\lambda(X_1, X_2) = \frac{\sup_{\Theta_0} L(d | (X_1, X_2))}{\sup_{\Theta} L(d | (X_1, X_2))} \quad (3)$$

The difference between the six test statistics of type (2) is in the estimators chosen for the standard deviation of $\hat{d}$. Six estimators have been considered:

$$\hat{\sigma}_1 = \left(\frac{\hat{p}_1 \left(1 - \hat{p}_1\right)}{n_1} + \frac{\hat{p}_2 \left(1 - \hat{p}_2\right)}{n_2} \right)^{1/2}$$

$$\hat{\sigma}_1 = \left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \right)^{1/2}$$

$$\hat{\sigma}_2 = \left(\frac{\check{p}_1(1-\check{p}_1)}{n_1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2} \right)^{1/2}$$

$$\hat{\sigma}_3 = \left(\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2} \right)^{1/2}$$

$$\hat{\sigma}_4 = \left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2-1} \right)^{1/2}$$

$$\hat{\sigma}_5 = \left(\frac{\check{p}_1(1-\check{p}_1)}{n_1-1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2-1} \right)^{1/2}$$

$$\hat{\sigma}_6 = \left(\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1-1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2-1} \right)^{1/2}$$

donde, $\hat{p}_i = \frac{X_i}{n_i}$ es el estimador de máxima verosimilitud de p_i , $\check{p}_i$ es el estimador de máxima verosimilitud restringida bajo la hipótesis nula de p_i (Miettinen y Nurminen, 1985; Farrington y Manning, 1990); y $\tilde{p}_i = \frac{X_i + 1}{n_i + 2}$ (Böhning y Viwatwongkasen, 2005).

Cuando $\lim_{\min(n_1, n_2) \rightarrow \infty} \frac{n_1}{n_1 + n_2} = q \in (0, 1)$ la estadística T en (2) tiene distribución asintótica normal estándar para cualquier estimador consistente $\hat{\sigma}$ de la desviación estándar de $\hat{d}$; para la estadística λ , la distribución asintótica de $-2 \ln \lambda$ es $\frac{1}{2} + \frac{1}{2} F_{\chi^2_1}$ donde $F_{\chi^2_1}$ denota la función de distribución acumulada de una variable aleatoria ji-cuadrada con un grado de libertad.

Las pruebas asintóticas para las estadísticas del tipo (2) para un nivel de significancia nominal α tienen región de rechazo de la forma $R_T(\alpha) = \{ (x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\} : T(x_1, x_2) < -z_\alpha \}$ donde z_α es el percentil superior α de la distribución normal estándar, es decir, $\Phi(z_\alpha) = 1 - \alpha$, donde Φ es la función de distribución acumulada de una variable aleatoria normal estándar.

La región de rechazo para la prueba asintótica correspondiente a la estadística (3) es:

$$R_T(\alpha) = \{ (x_1, x_2) : -2 \ln \lambda(x_1, x_2) > \chi^2_{1-2\alpha}(1) \}$$

$$\hat{\sigma}_2 = \left(\frac{\check{p}_1(1-\check{p}_1)}{n_1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2} \right)^{1/2}$$

$$\hat{\sigma}_3 = \left(\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2} \right)^{1/2}$$

$$\hat{\sigma}_4 = \left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2-1} \right)^{1/2}$$

$$\hat{\sigma}_5 = \left(\frac{\check{p}_1(1-\check{p}_1)}{n_1-1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2-1} \right)^{1/2}$$

$$\hat{\sigma}_6 = \left(\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1-1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2-1} \right)^{1/2}$$

where $\hat{p}_i = \frac{X_i}{n_i}$ is the maximum likelihood estimator of p_i , $\check{p}_i$ is the maximum likelihood estimator of p_i restricted under the null hypothesis (Miettinen and Nurminen, 1985; Farrington and Manning, 1990); and finally $\tilde{p}_i = \frac{X_i + 1}{n_i + 2}$ (Böhning y Viwatwongkasen, 2005).

When $\lim_{\min(n_1, n_2) \rightarrow \infty} \frac{n_1}{n_1 + n_2} = q \in (0, 1)$, the T statistic in (2) has an asymptotic normal standard distribution for any consistent estimator $\hat{\sigma}$ of the standard deviation of $\hat{d}$, for the statistic λ , the asymptotic distribution of $-2 \ln \lambda$ is $\frac{1}{2} + \frac{1}{2} F_{\chi^2_1}$ where $F_{\chi^2_1}$ denotes the cumulative distribution function of a chi-square random variable with one degree of freedom.

The asymptotic tests corresponding to the type (2) test statistics, with nominal level equal to α , have a rejection region of the form $R_T(\alpha) = \{ (x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\} : T(x_1, x_2) < -z_\alpha \}$, where z_α is the upper α -percentile of the standard normal distribution, that is, $\Phi(z_\alpha) = 1 - \alpha$, where Φ is the cumulative distribution function of the standard normal random variable.

The rejection region for the asymptotic test corresponding to the test statistic (3) is:

$$R_T(\alpha) = \{ (x_1, x_2) : -2 \ln \lambda(x_1, x_2) > \chi^2_{1-2\alpha}(1) \}$$

where $\chi^2_{1-2\alpha}(1)$ is the $1 - 2\alpha$ percentile of the chi-square distribution with one degree of freedom, that is, $P(\chi^2_1 > \chi^2_{1-2\alpha}(1)) = 1 - 2\alpha$.

donde $\chi^2_{1-2\alpha}(1)$ es el percentil superior $1 - 2\alpha$ de la distribución ji-cuadrada con un grado de libertad, i.e. $P(\chi^2_1 > \chi^2_{1-2\alpha}(1)) = 1 - 2\alpha$.

Las correcciones por continuidad analizadas son:

$$C_0 = 0, \quad C_1 = \frac{1}{4 \min(n_1, n_2)}, \quad C_2 = 2C_1$$

$$C_3 = \frac{1}{2n_1} + \frac{1}{2n_2}, \quad C_4 = \frac{3}{2 \min(n_1, n_2)}, \quad C_5 = \frac{2}{\min(n_1, n_2)},$$

donde C_0 , C_2 y C_3 son consideradas por Hauck y Anderson (1986) para el caso de las estadísticas T_1 y T_4 .

El análisis se realizó para diseños balanceados, es decir, para $n_1 = n_2 = n$. En esta situación es claro que $C_i < C_{i+1}$ para $i = 0, 1, 2, 3, 4$.

Así, las estadísticas de prueba consideradas son:

$$T_{iC_j}(X_1, X_2) = \frac{\hat{p}_1 - \hat{p}_2 - d_0 + C_j}{\hat{\sigma}_i} \quad (4)$$

para $i = 0, 1, 2, 3, 4, 5, 6$ y $j = 0, 1, 2, 3, 4, 5$ y

$$T_{7C_j}(X_1, X_2) = \lambda(X_1, X_2) + C_j = \frac{\sup_{\Theta} L(d | (X_1, X_2))}{\sup_{\Theta} L(d | (X_1, X_2))} + C_j \quad (5)$$

para $j = 0, 1, 2, 3, 4, 5$.

Las pruebas T_{iC_j} fueron propuestas en los siguientes artículos: T_{1C_0} en Blackwelder (1982), T_{2C_0} en Farrington y Manning (1990), T_{3C_0} en Böhning y Viwatwongkasen (2005), T_{4C_0} en Hauck y Anderson (1986). T_{5C_0} se obtiene de T_{2C_0} al reemplazar n_1 por $n_1 - 1$ y n_2 por $n_2 - 1$ en el denominador del estimador de la desviación estándar, mientras que T_{6C_0} se obtiene de T_{3C_0} al reemplazar n_1 por $n_1 - 1$ y n_2 por $n_2 - 1$ en el denominador del estimador de la desviación estándar. Finalmente, T_{7C_0} es la conocida estadística de razón de verosimilitudes (Casella y Berger, 2002).

El nivel de significancia nominal usado en todo este trabajo fue $p = 0.05$. Las pruebas estadísticas serán simbolizadas de la misma forma que sus correspondientes estadísticas de prueba.

Estrategia para el cálculo del nivel de significancia

De acuerdo con el modelo Bernoulli usado en este trabajo, el espacio muestral es $\chi = \{(x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\}\}$, el espacio paramétrico es $\Theta = \{(p_1, p_2) \in [0, 1]^2\}$, y dado que X_i tiene distribución binomial con parámetros (n_i, p_i) para $i = 1, 2$, la función de verosimilitud conjunta es:

$$L(p_1, p_2; x_1, x_2) = \binom{n_1}{x_1} p_1^{x_1} (1 - p_1)^{n_1 - x_1} \binom{n_2}{x_2} p_2^{x_2} (1 - p_2)^{n_2 - x_2},$$

The corrections for continuity analyzed are:

$$C_0 = 0, \quad C_1 = \frac{1}{4 \min(n_1, n_2)}, \quad C_2 = 2C_1$$

$$C_3 = \frac{1}{2n_1} + \frac{1}{2n_2}, \quad C_4 = \frac{3}{2 \min(n_1, n_2)}, \quad C_5 = \frac{2}{\min(n_1, n_2)},$$

where C_0 , C_2 and C_3 are considered by Hauck and Anderson (1986) for the case of statistics T_1 and T_4 .

The analysis was carried out for balanced designs, that is, for $n_1 = n_2 = n$. In this situation it is clear that $C_i < C_{i+1}$ for $i = 0, 1, 2, 3, 4$.

Thus, the test statistics considered are:

$$T_{iC_j}(X_1, X_2) = \frac{\hat{p}_1 - \hat{p}_2 - d_0 + C_j}{\hat{\sigma}_i} \quad (4)$$

for $i = 0, 1, 2, 3, 4, 5, 6$ and $j = 0, 1, 2, 3, 4, 5$ and

$$T_{7C_j}(X_1, X_2) = \lambda(X_1, X_2) + C_j = \frac{\sup_{\Theta} L(d | (X_1, X_2))}{\sup_{\Theta} L(d | (X_1, X_2))} + C_j \quad (5)$$

for $j = 0, 1, 2, 3, 4, 5$.

The test statistics T_{iC_j} were proposed in the following papers: T_{1C_0} in Blackwelder (1982), T_{2C_0} in Farrington and Manning (1990), T_{3C_0} in Böhning and Viwatwongkasen (2005), T_{4C_0} in Hauck and Anderson (1986). T_{5C_0} is obtained from T_{2C_0} by replacing n_1 by $n_1 - 1$ and n_2 by $n_2 - 1$ in the denominator of the standard deviation estimator, while T_{6C_0} is obtained from T_{3C_0} by replacing n_1 by $n_1 - 1$ and n_2 by $n_2 - 1$ in the denominator of the standard deviation estimator. Finally, T_{7C_0} is the well-known likelihood ratio test statistic (Casella and Berger, 2002).

The nominal significance level used for the whole study was $p = 0.05$. The statistical tests will be symbolized in the same way than their corresponding test statistics.

Strategy for calculation of the significance level

According to the Bernoulli model used in this work, the sample space is $\chi = \{(x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\}\}$, the parameter space is $\Theta = \{(p_1, p_2) \in [0, 1]^2\}$, and since X_i has a binomial distribution with (n_i, p_i) parameters for $i = 1, 2$, the joint likelihood function is:

$$L(p_1, p_2; x_1, x_2) = \binom{n_1}{x_1} p_1^{x_1} (1 - p_1)^{n_1 - x_1} \binom{n_2}{x_2} p_2^{x_2} (1 - p_2)^{n_2 - x_2},$$

and the power function is $\beta_T(p_1, p_2) = \sum_{(x_1, x_2) \in R_T(\alpha)} L(p_1, p_2; x_1, x_2);$

y la función de potencia es $\beta_T(p_1, p_2) = \sum_{(x_1, x_2) \in R_T(\alpha)} L(p_1, p_2; x_1, x_2)$ además, el espacio nulo es $\Theta_0 = \{(p_1, p_2) \in \Theta : p_1 - p_2 \geq d_0\}$ y el nivel de significancia está dado por $\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2)$.

Chan (1998) calculó el nivel de significancia para la prueba de Farrington-Manning (T_{2C_0}) tomando el supremo no en todo el espacio nulo (Θ_0) , sino calculando el máximo únicamente en $\Theta_0^* = \{(p_1, p_2) \in \Theta : p_1 - p_2 = d_0\}$, el cual es sólo una parte de la frontera del espacio nulo. Computacionalmente esto representa una inmensa ventaja, pues el tiempo de cómputo se reduce aproximadamente a 0.22% del original. Sin embargo, el autor mencionado no justificó formalmente la validez de este argumento. Röhmel (2005) presentó una prueba formal que justifica el procedimiento usado por Chan (1998). En este trabajo se siguió la misma estrategia de Chan (1998). Por tanto, en lo que resta de esta sección, se verifica la validez de la llamada condición de convexidad de Barnard y de la condición de simetría en la misma cola (ver definiciones abajo) para todas las pruebas asintóticas consideradas.

Definición. Una prueba estadística para el problema en (1) con región de rechazo R_T cumple la condición de convexidad de Barnard (C) si satisface las dos propiedades siguientes:

- a) $(x, y) \in R_T \Rightarrow (x-1, y) \in R_T \quad \forall \quad 1 \leq x \leq n_1, \quad 0 \leq y \leq n_2$
- b) $(x, y) \in R_T \Rightarrow (x, y+1) \in R_T \quad \forall \quad 0 \leq x \leq n_1, \quad 0 \leq y \leq n_2 - 1$

Röhmel y Mansmann (1999) demostraron el resultado ya mencionado para pruebas exactas. A continuación se demuestra el resultado correspondiente para pruebas asintóticas.

Proposición 1. Sea T una estadística con distribución asintótica A_T para el problema de prueba de hipótesis:

$$[H_0 : p_2 \leq g(p_1)] \text{ vs. } [H_a : p_2 > g(p_1)]$$

con región crítica dada por $R_T^A(\alpha) = \{(x_1, x_2) : T(x_1, x_2) \leq c_\alpha\}$ y sean $\Xi_0 = \{(p_1, p_2) \in [0, 1]^2 : p_2 \leq g(p_1)\}$, $\Xi_0^* = \{(p_1, p_2) \in [0, 1]^2 : p_2 = g(p_1)\}$, $a \leq p_1 \leq 1$, $0 \leq a$ y $g(a) = 0$ cuando $a > 0$ con g creciente y no interseca $p_1 = p_2$ si la región crítica es no vacía y satisface la condición de convexidad de Barnard, entonces el supremo $\sup_{(p_1, p_2) \in \Xi_0} \beta_T(p_1, p_2)$ es un máximo y se alcanza en un punto frontera sobre la curva $p_2 = g(p_1)$.

Demostración. Puesto que por hipótesis $R_T^A(\alpha) \neq \emptyset$, entonces existe $(x_0, y_0) \in R_T^A(\alpha)$ tal que $\max_{(x_1, x_2) \in R_T^A(\alpha)} T(x_1, x_2) = T(x_0, y_0) \leq c_\alpha$, es decir:

$$R_T(\alpha) = \{(x_1, x_2) \in \chi : T(x_1, x_2) \leq T(x_0, y_0)\}$$

de donde, por el teorema demostrado para pruebas exactas por Röhmel y Mansmann (1999) se sigue el resultado.

also, the null space is $\Theta_0 = \{(p_1, p_2) \in \Theta : p_1 - p_2 \geq d_0\}$ and the significance level is given by $\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2)$.

Chan (1998) calculated the significance level of the Farrington-Manning test (T_{2C_0}) , taking the supreme not from the whole null space (Θ_0) but rather calculating the maximum solely in $\Theta_0^* = \{(p_1, p_2) \in \Theta : p_1 - p_2 = d_0\}$, which is only part of the null space borderline. Computationally, this represents an enormous advantage, because the computing time is reduced to approximately 0.22% of the original time. However, the author mentioned did not formally justify the validity of this argument. Röhmel (2005) presented a formal proof that justifies the procedure used by Chan (1998). In this article we have used the same strategy used by Chan (1998). Therefore, in the remainder of this section, the validity of the so-called Barnard convexity condition and the condition of symmetry on the same tail (see definitions below) are verified for all the asymptotic tests considered here.

Definition. A statistical test, for the problem in (1), with rejection region R_T fulfills the Barnard convexity condition (C) if satisfies both of the following properties:

- a) $(x, y) \in R_T \Rightarrow (x-1, y) \in R_T \quad \forall \quad 1 \leq x \leq n_1, \quad 0 \leq y \leq n_2$
- b) $(x, y) \in R_T \Rightarrow (x, y+1) \in R_T \quad \forall \quad 0 \leq x \leq n_1, \quad 0 \leq y \leq n_2 - 1$

Röhmel and Mansmann (1999) demonstrated the result already mentioned for exact tests. Next, the result for the asymptotic tests will be proved.

Proposition 1. Let T be a test statistic with an asymptotic distribution A_T for the hypothesis testing problem:

$$[H_0 : p_2 \leq g(p_1)] \text{ vs. } [H_a : p_2 > g(p_1)]$$

with a critical region given by $R_T^A(\alpha) = \{(x_1, x_2) : T(x_1, x_2) \leq c_\alpha\}$ and suppose $\Xi_0 = \{(p_1, p_2) \in [0, 1]^2 : p_2 \leq g(p_1)\}$, $\Xi_0^* = \{(p_1, p_2) \in [0, 1]^2 : p_2 = g(p_1)\}$, $a \leq p_1 \leq 1$, $0 \leq a$ and $g(a) = 0$ when $a > 0$ with g an increasing function which do not intersect $p_1 = p_2$, if the critical region is not empty and satisfies the Barnard convexity condition, then the supreme $\sup_{(p_1, p_2) \in \Xi_0} \beta_T(p_1, p_2)$ is a maximum and is reached in a border point on the curve $p_2 = g(p_1)$.

Demonstration. By hypothesis $R_T^A(\alpha) \neq \emptyset$, thus there exists $(x_0, y_0) \in R_T^A(\alpha)$ such that $\max_{(x_1, x_2) \in R_T^A(\alpha)} T(x_1, x_2) = T(x_0, y_0) \leq c_\alpha$, that is:

$$R_T(\alpha) = \{(x_1, x_2) \in \chi : T(x_1, x_2) \leq T(x_0, y_0)\}$$

Using this fact and, by the theorem proved for exact tests by Röhmel and Mansmann (1999) the result follows.

In this work, we have used the function $g(p_1) = p_1 - d_0$.

Note that, by definition, if $R_T \neq \emptyset$ and it fulfills condition (C), then necessarily $(0, n_2) \in R_T$; besides, $\hat{\sigma}_1(0, n_2) = \hat{\sigma}_4(0, n_2) = 0$

En este trabajo se usa la función $g(p_1) = p_1 - d_0$.

Nótese que, por definición, si $R_T \neq \emptyset$ y se cumple la condición (C), entonces necesariamente $(0, n_2) \in R_T$; además, $\hat{\sigma}_1(0, n_2) = \hat{\sigma}_4(0, n_2) = 0$ y por tanto $T_1(0, n_2)$ y $T_4(0, n_2)$ no están definidas, entonces $(0, n_2) \notin R_{T_1} \cup R_{T_4}$ y en consecuencia R_{T_1} y R_{T_4} no cumplen (C). Una forma de solventar esta dificultad es redefiniendo la estimación de la desviación estándar en los puntos donde se anula.

Observación 1. Sea $\phi(n, x) = \frac{1}{n} \frac{x}{n} \left(1 - \frac{x}{n}\right)$, entonces $\phi(n, x) = \phi(n, n - x)$

Como $\hat{\sigma}_1(x, y) = 0$ para $(x, y) \in \{0, n_1\} \times \{0, n_2\}$ se puede redefinir de la siguiente forma:

$$\hat{\sigma}_1(x, y) = \begin{cases} \sqrt{\phi(n_1, .01) + \phi(n_2, .01)} & \text{si } x = y = 0 \\ \sqrt{\phi(n_1, .01) + \phi(n_2, n_2 - .01)} & \text{si } x = 0, y = n_2 \\ \sqrt{\phi(n_1, n_1 - .01) + \phi(n_2, .01)} & \text{si } x = n_1, y = 0 \\ \sqrt{\phi(n_1, n_1 - .01) + \phi(n_2, n_2 - .01)} & \text{si } x = n_1, y = n_2 \end{cases}$$

de manera similar se redefine $\hat{\sigma}_4(x, y)$.

La demostración del siguiente resultado se establece en Almendra (2007).

Proposición 2. Sea T una estadística con distribución asintótica A_T . Sean las regiones críticas para las pruebas exacta y asintótica R_T y R_T^A . Si R_T cumple la condición de convexidad de Barnard, entonces R_T^A también la cumple.

Como se mencionó, Röhmle (2005) demostró que la versión exacta de la región crítica para T_{2C_0} cumple la condición de convexidad de Barnard. Entonces, por la proposición anterior, la versión asintótica también cumple la condición de convexidad de Barnard.

Definición. Si $n_1 = n_2 = n$, una región de rechazo R cumple la condición de simetría en la misma cola si $(x, y) \in R \Rightarrow (n - y, n - x) \in R$.

Proposición 3. Si $n_1 = n_2 = n$ entonces $R_{T_{iC_j}}$ cumple la condición de simetría en la misma cola para $i = 1, 3, 4, 6$ y $j = 0, 1, 2, 3, 4, 5$.

Demostración. Almendra (2007).

Dado el problema de prueba de hipótesis (1) si se considera una estadística de prueba $T(X_1, X_2)$ con región de rechazo

$R_T = \{(x_1, x_2) \in \chi : T(x_1, x_2) < c\}$, se define el espacio muestral lícito como $\chi' = \{(x_1, x_2) \in \chi : \hat{d} < d_0\}$; la región crítica restringida al espacio muestral lícito es $R'_T = \{(x_1, x_2) \in \chi' : T(x_1, x_2) < c\}$.

La importancia del concepto anterior es reiterada por Martín y Herranz (2002, 2004a, 2004b) quienes enfatizan la necesidad de tal concepto, ya que de no restringirse el espacio muestral al espacio muestral lícito, podrían obtenerse inferencias inválidas en algunos casos. La razón de esta restricción es que de otra forma la prueba podría no tener significancia: por ejemplo, si $\hat{p}_1 - \hat{p}_2 \geq d_0$, entonces es inferencialmente ilógico concluir que $p_1 - p_2 < d_0$.

and therefore $T_1(0, n_2)$ and $T_4(0, n_2)$ are not defined, then $(0, n_2) \notin R_{T_1} \cup R_{T_4}$ and as consequence R_{T_1} and R_{T_4} do not fulfill (C). A way of solving this difficulty is by redefining the estimation of the standard deviation in the points in which is equal to zero.

Observation 1. Let $\phi(n, x) = \frac{1}{n} \frac{x}{n} \left(1 - \frac{x}{n}\right)$, then $\phi(n, x) = \phi(n, n - x)$

Since $\hat{\sigma}_1(x, y) = 0$ for $(x, y) \in \{0, n_1\} \times \{0, n_2\}$ it can be redefined in the following way:

$$\hat{\sigma}_1(x, y) = \begin{cases} \sqrt{\phi(n_1, .01) + \phi(n_2, .01)} & \text{if } x = y = 0 \\ \sqrt{\phi(n_1, .01) + \phi(n_2, n_2 - .01)} & \text{if } x = 0, y = n_2 \\ \sqrt{\phi(n_1, n_1 - .01) + \phi(n_2, .01)} & \text{if } x = n_1, y = 0 \\ \sqrt{\phi(n_1, n_1 - .01) + \phi(n_2, n_2 - .01)} & \text{if } x = n_1, y = n_2 \end{cases}$$

in a similar way $\hat{\sigma}_4(x, y)$ is redefined.

The proof of the following result is established in Almendra (2007).

Proposition 2. Let T be a test statistic with asymptotic distribution A_T . Suppose that the exact and asymptotic critical regions corresponding to this test statistic are R_T and R_T^A . Then, if R_T fulfills the Barnard convexity condition, R_T^A also fulfills it.

As mentioned, Röhmle (2005) demonstrated that the exact version of the critical region for T_{2C_0} fulfills the Barnard convexity condition. Then, using the previous proposition, the asymptotic version also fulfills the Barnard convexity condition.

Definition. If $n_1 = n_2 = n$, a rejection region R fulfills the condition of symmetry on the same tail if $(x, y) \in R \Rightarrow (n - y, n - x) \in R$.

Proposition 3. If $n_1 = n_2 = n$, then $R_{T_{iC_j}}$ fulfills the condition of symmetry on the same tail for $i = 1, 3, 4, 6$ and $j = 0, 1, 2, 3, 4, 5$.

Demonstration. See Almendra (2007).

Given the hypothesis testing problem in (1), let $T(X_1, X_2)$ be a test statistic with a rejection region

$R_T = \{(x_1, x_2) \in \chi : T(x_1, x_2) < c\}$, then the licit sample space is defined as $\chi' = \{(x_1, x_2) \in \chi : \hat{d} < d_0\}$; the critical region restricted to the licit sample space is $R'_T = \{(x_1, x_2) \in \chi' : T(x_1, x_2) < c\}$.

The importance of the previous concept is reiterated by Martín and Herranz (2002, 2004a, 2004b), who emphasize the need of such a concept since if the sample space is not restricted in this manner, to the licit sample space, invalid inferences could be obtained in some cases. The reason for this restriction is that otherwise the test could possibly not reach significance, for example, if $\hat{p}_1 - \hat{p}_2 \geq d_0$ then it is inferentially illogical to conclude that $p_1 - p_2 < d_0$.

Proposition 4. For type (2) test statistics, the sample space is equal to the licit sample space, that is, if T is a test statistic

of the form $T(x_1, x_2) = \frac{\hat{d} - d_0}{\hat{\sigma}}$ for a test with a rejection region $R_T = \{(x_1, x_2) \in \chi : T(x_1, x_2) < -z_\alpha\}$, then $R_T = R'_T$.

Proposición 4. Para estadísticas tipo (2), el espacio muestral coincide con el espacio muestral lícito, es decir, si T es una estadística de la forma $T(x_1, x_2) = \frac{\hat{d} - d_0}{\hat{\sigma}}$ para una prueba con región de rechazo $R_T = \{(x_1, x_2) \in \mathcal{X} : T(x_1, x_2) < -z_\alpha\}$, entonces $R_T = R'_T$.

Para una demostración de la proposición anterior, ver Almendra (2007).

No obstante, para la estadística de razón de verosimilitudes los espacios muestral y muestral lícito no coinciden en general. Por esta razón el cálculo de los niveles de significancia y las potencias para T_7 se realizó con base en el espacio muestral lícito.

Proposición 5. Sean $n_1 = n_2 = n$ y $R(\alpha)$ una región crítica para el problema de prueba de hipótesis $[H_0 : d \geq d_0]$ vs. $[H_a : d < d_0]$, si $R(\alpha)$ cumple la condición de convexidad de Barnard y la condición de simetría en la misma cola, entonces el nivel de significancia está dado por:

$$\alpha^* = \max_{\substack{p_2 = p_1 - d_0 \\ p_1 \in \left[d_0, \frac{1+d_0}{2}\right]}} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} \quad (6)$$

$$(1-p_2)^{n_2-x_2} I_{[(x_1, x_2) \in R(\alpha)]}$$

Demostración. Almendra (2007).

Observación 2. Todas las pruebas asintóticas T_{iC_j} , $i = 1, 2, \dots, 7$; $j = 0, 1, \dots, 5$, con estadísticas de prueba definidas en (4) y (5) satisfacen la condición de convexidad de Barnard y la condición de simetría en la misma cola.

La condición de convexidad de Barnard para T_{2C_0} se probó analíticamente en la proposición 2. La condición de simetría en la misma cola para T_{iC_j} para $i = 1, 3, 4, 6$; $j = 0, 1, 2, 3, 4, 5$; se probó analíticamente en la proposición 3. Para las demás pruebas, la verificación se efectuó numéricamente mediante programas de cómputo elaborados en S-PLUS®.

Con base en la proposición 5 y la observación 2, se usó la fórmula en (6) para calcular una aproximación del nivel de significancia exacto de todas las pruebas consideradas. La aproximación del nivel de significancia exacta α^* , en la fórmula (6), se hizo reemplazando el intervalo continuo $[d_0, (1+d_0)/2]$ por el conjunto discreto de puntos: $\{d_0 + (.001)i : i = 0, 1, 2, \dots, 500(1-d_0)\}$, y al valor obtenido de esta aproximación de α^* usando tal conjunto discreto se le ha llamado nivel de significancia real de la prueba, el cual será denotado por α_R .

Es importante subrayar que la fórmula en (6), para el cálculo del nivel de significancia (α^*) de cualquiera de las pruebas consideradas aquí, es una fórmula exacta. Así que la única aproximación que se hace para calcular el nivel de significancia real (α_R) es al reemplazar al intervalo continuo $[d_0, (1+d_0)/2]$ por el conjunto de puntos $\{d_0 + (.001)i : i = 0, 1, 2, \dots, 500(1-d_0)\}$. El grado de error en que incurre esta aproximación se estimó de la siguiente manera: para todas las pruebas consideradas y para cada una de las doce

For a demonstration of the latter proposition, see Almendra (2007).

Nevertheless, for the likelihood ratio test statistic the sample space and licit sample space do not generally coincide. This is why the calculation of the significance levels and the powers for T_7 were performed using the licit sample space.

Proposition 5. Let $n_1 = n_2 = n$ and $R(\alpha)$ a critical region for the hypothesis testing problem $[H_0 : d \geq d_0]$ vs. $[H_a : d < d_0]$, if $R(\alpha)$ fulfills the Barnard convexity condition and the condition of symmetry on the same tail, then the significance level of $R(\alpha)$ is given by:

$$\alpha^* = \max_{\substack{p_2 = p_1 - d_0 \\ p_1 \in \left[d_0, \frac{1+d_0}{2}\right]}} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} \quad (6)$$

$$(1-p_2)^{n_2-x_2} I_{[(x_1, x_2) \in R(\alpha)]}$$

Demonstration. See Almendra (2007).

Observation 2. All the asymptotic tests T_{iC_j} , $i = 1, 2, \dots, 7$; $j = 0, 1, \dots, 5$, with test statistics defined in (4) and (5), satisfy the Barnard convexity condition and the condition of symmetry on the same tail.

The Barnard convexity condition for T_{2C_0} was analytically proved in proposition 2. The condition of symmetry on the same tail for T_{iC_j} for $i = 1, 3, 4, 6$; $j = 0, 1, 2, 3, 4, 5$; was analytically proved in proposition 3. For the other tests, the verification was performed numerically through computer programs elaborated in S-PLUS®.

Based on proposition 5 and observation 2, the formula in (6) was used to calculate an approximation to the exact significance level for all tests considered. The approximation of the exact significance level (α^*), in formula (6), was performed by replacing the continuous interval $[d_0, (1+d_0)/2]$ by the discrete set of points $\{d_0 + (.001)i : i = 0, 1, 2, \dots, 500(1-d_0)\}$, and the value obtained from this approximation to α^* , using such discrete set of points, has been called real significance level of the test, which will be denoted by α_R .

It is important to underline that the formula in (6), for the calculation of the significance level (α^*) of any of the tests considered here, is an exact formula. So the only approximation that is made in order to calculate the real significance level (α_R) is by replacing the continuous interval $[d_0, (1+d_0)/2]$ by the set of points $\{d_0 + (.001)i : i = 0, 1, 2, \dots, 500(1-d_0)\}$. The degree of error in which this approximation incurs was estimated as follows: for all the tests considered and for each one of the twelve combinations of parameters ($n_1 = n_2 = 30, 50$ and 100), ($\alpha = 0.05$ and 0.10), ($d_0 = 0.10$ and 0.15), the error incurred for the estimation of α^* by α_R was analyzed, and in all cases it was verified that in the values adjacent to the maximum estimated (α_R), the variation in the power function was less than or equal to 0.00001, that is, our estimation of the error $|\alpha^* - \alpha_R|$ is that it is less than or equal to 0.00001. For

combinaciones de parámetros ($n_1 = n_2 = 30, 50$ y 100), ($\alpha=0.05$ y 0.10), ($d_0 = 0.10$ y 0.15) se analizó el error incurrido al estimar α^* por α_R y en todos los casos se verificó que en los valores adyacentes al máximo estimado (α_R) la variación en la función de potencia fue menor o igual a 0.00001 , es decir, nuestra estimación del error $|\alpha^* - \alpha_R|$ es que es menor o igual a 0.00001 . Por ejemplo, para $n_1 = n_2 = 50$, $\alpha = 0.05$ y $d_0 = 0.10$ los valores adyacentes al máximo estimado (α_R) de la función de potencia son:

$$\begin{aligned}\beta_T(0.549, 0.449) &= 0.0441077 \\ \beta_T(0.550, 0.450) &= 0.0441080 = \alpha_R \text{ (máximo aproximado)} \\ \beta_T(0.551, 0.451) &= 0.0441077,\end{aligned}$$

en este caso $\beta_T(0.550, .0450) - \beta_T(0.549, .449) = 0.0000003$, el cual es menor a 0.00001 . Con base en este análisis es razonable concluir que el error incurrido al estimar α^* por α_R es aceptable desde el punto de vista práctico, ya que el nivel de significancia real (α_R) difiere del nivel de sinificancia exacto (α^*) cuando mucho en la quinta cifra decimal.

RESULTADOS Y DISCUSIÓN

Comparación de las pruebas con base en sus niveles de significancia reales

Para comparar las 42 pruebas estadísticas consideradas en este trabajo (T_{iC_j} con $1 \leq i \leq 7$ y $0 \leq j \leq 5$), parece razonable considerar que el nivel de significancia real (α_R) de una prueba tiene un buen comportamiento cuando $\alpha_R \in [0.04, 0.05]$, es decir cuando α_R es menor o igual al nivel nominal ($\alpha=0.05$), y cuando α_R es una buena aproximación al nivel nominal (α): $|R - \alpha| \leq 0.01$. Con este propósito, para cada prueba T_{iC_j} se calculó el porcentaje de α_R s que pertenece al intervalo $[0.04, 0.05]$ con base en los 71 tamaños de muestra n en el rango $30 \leq n \leq 100$, y para cada uno de los valores de $d_0 = 0.10, 0.15$ y 0.20 . En el Cuadro 1 se reportan estos resultados, pero únicamente para las pruebas que obtuvieron porcentajes mayores o iguales a 90%. T_{2C_2} y T_{5C_2} resultaron las pruebas con mejor comportamiento del nivel de significancia real (α_R), para los tres valores considerados: $d_0 = 0.10, 0.15$ y 0.20 .

Comparación de las pruebas con base en sus potencias reales

Para cada valor de d_0 ($0.10, 0.15, 0.20$) se compararon las potencias de las pruebas T_{2C_2} y T_{5C_2} para aquellos tamaños de muestra n , $30 \leq n \leq 100$, donde la máxima diferencia entre los niveles de significancia reales resultó menor o igual que 0.0001 y donde al menos una de las potencias a comparar fue mayor o igual que 0.7 . La comparación de las

example, for $n_1 = n_2 = 50$, $\alpha = 0.05$ and $d_0 = 0.10$, the values adjacent to the maximum (α_R) estimated for the power function, are:

$$\begin{aligned}\beta_T(0.549, 0.449) &= 0.0441077 \\ \beta_T(0.550, 0.450) &= 0.0441080 = \alpha_R \text{ (aproximate maximum)} \\ \beta_T(0.551, 0.451) &= 0.0441077,\end{aligned}$$

in this case $\beta_T(0.550, .0450) - \beta_T(0.549, .449) = 0.0000003$, which is less than 0.00001 . Based on this analysis it is reasonable to conclude that the error incurred, when estimating α^* by α_R , is acceptable from a practical point of view, since the real significance level of (α_R) differs from the exact significance level (α^*) at most in the fifth decimal figure.

RESULTS AND DISCUSSION

Comparison of the tests based on their real levels of significance

In order to compare the 42 statistical tests considered in this study (T_{iC_j} with $1 \leq i \leq 7$ and $0 \leq j \leq 5$), it seems reasonable to consider that the real significance level (α_R) of a test has a good performance when $\alpha_R \in [0.04, 0.05]$, that is, when α_R is less or equal to the nominal level ($\alpha=0.05$) and when α_R is a good approximation to the nominal level (α): $|R - \alpha| \leq 0.01$. With this purpose, for each test T_{iC_j} the percentage of α_R 's that belong to the interval $[0.04, 0.05]$ was calculated, based on the 71 sample sizes n in the $30 \leq n \leq 100$ range, and for each one of the values for $d_0=0.10, 0.15$ and 0.20 . In Table 1 these results are reported, but only for those tests where percentages were greater than or equal to 90%. Clearly, T_{2C_2} and T_{5C_2} were the tests with the best performance in terms of the real significance level (α_R), for the three values considered: $d_0 = 0.10, 0.15$ and 0.20 .

Cuadro 1. Porcentaje de niveles de significancia reales que pertenecen al intervalo $[.04, .05]$ con base en los 71 tamaños de muestra n en el rango $30 \leq n \leq 100$.

Table 1. Percentage of real significance levels that belong to the interval $[.04, .05]$ based on the 71 sample sizes n in the $30 \leq n \leq 100$ range.

d_0	Prueba	Porcentaje de niveles de significancia reales que pertenecen al intervalo $[.04, .05]$
0.1	T2C2	98.59
	T5C2	100.00
0.15	T2C2	91.55
	T5C2	90.14
0.2	T2C2	98.59
	T5C2	95.77

potencias se realizó en los puntos (p_1, p_2) con $p_1 = 0(.5)1$ y $p_2 \in [\max(p_1 - d_0, 0), 1]$ iniciando en $\max(p_1 - d_0, 0)$ y tomando incrementos de 0.05, es decir, con $p_2 = \max(p_1 - d_0, 0) + 0.05i$ para aquellos i tales que $p_2 \in [\max(p_1 - d_0, 0), 1]$.

En resumen las potencias de T_{2C_2} y T_{5C_2} se compararon para un total de (40, 26, y 28) tamaños de muestra en los casos ($d_0 = 0.10, 0.15$, y 0.20) respectivamente. Para todos estos tamaños de muestra en donde se realizó la comparación, las potencias de las pruebas T_{2C_2} y T_{5C_2} resultaron iguales en el 100% de los casos. Estos resultados dan como ganadoras a las pruebas T_{2C_2} y T_{5C_2} con ventaja de T_{2C_2} para $d_0 = 0.15$ y 0.20 . Para el caso $d_0 = 0.10$, T_{5C_2} supera a T_{2C_2} , aunque la diferencia es sólo 1.41%. Para una recomendación práctica se preferirá a la prueba T_{2C_2} en todos los casos ($d_0 = 0.10, 0.15$, y 0.20).

CONCLUSIONES

Para los tamaños de muestra $30 \leq n \leq 100$, considerando un nivel de significancia nominal $\alpha = 0.05$ y para los tres límites de no-inferioridad analizados ($d_0 = 0.10, 0.15$, y 0.20), se recomienda usar T_{2C_2} ; es decir, la prueba de Farrington-Manning con el factor de corrección C_2 .

AGRADECIMIENTOS

El primer autor desea agradecer a la UPIITA del Instituto Politécnico Nacional su apoyo al otorgarle licencia para la realización del presente trabajo.

LITERATURA CITADA

- Almendra A., F. 2007. Comparación de algunas pruebas estadísticas asintóticas de no-inferioridad para contrastar dos proporciones independientes. Tesis Doctoral, Especialidad en Estadística, ISEI, Colegio de Postgraduados, México. 87 p.
- Blackwelder, W. 1982. "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials* 3:345-353.
- Böhning, D., and C. Wiwatwongkasen. 2005. Revisiting proportion estimators. *Statistical Methods in Medical Res.* 14: 1-23.
- Casella, G., and L. Berger. 2002. *Statistical Inference*. Second Edition. Duxbury, Thompson Learning. USA. 660 p.
- Chan, I. S. F. 1998. Exact tests of equivalence and efficacy with a non zero lower bound for comparative studies. *Statistics in Medicine* 17: 1403-1413.
- Chen, J., Y. Tsong, and S. Kang. 2000. Tests for equivalence or noninferiority between two proportions. *Drug Information J.* 34: 569-578.
- Farrington, C., and G. Manning. 1990. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* 9: 1447-1454.
- Hauck, W., and S. Anderson. 1986. A comparison of large-sample confidence interval methods for the difference of two binomial probabilities. *The Am. Statistician* 40: 318-322.

Comparison of the tests based on their real powers

For each value of d_0 (0.10, 0.15, 0.20), the powers of the tests T_{2C_2} and T_{5C_2} were compared for the sample sizes n , $30 \leq n \leq 100$, where the maximum difference between the real significance levels was less than or equal to 0.0001 and where at least one of the powers to be compared was greater or equal to 0.7. The comparison of the powers was carried out on points (p_1, p_2) with $p_1 = 0(.5)1$ and $p_2 \in [\max(p_1 - d_0, 0), 1]$ starting at $\max(p_1 - d_0, 0)$ and with increments of 0.05; that is, with $p_2 = \max(p_1 - d_0, 0) + 0.05i$ for those i so that $p_2 \in [\max(p_1 - d_0, 0), 1]$.

Summarizing, the powers of T_{2C_2} and T_{5C_2} were compared for a total of (40, 26, and 28) sample sizes corresponding to the values ($d_0 = 0.10, 0.15$, and 0.20), respectively. For all these sample sizes the comparison of these tests was carried out, and the powers of the tests T_{2C_2} and T_{5C_2} were equal in 100% of the cases. These results give as winners tests T_{2C_2} and T_{5C_2} with an advantage by T_{2C_2} for $d_0 = 0.15$ and 0.20 . For the case $d_0 = 0.10$, T_{5C_2} outperforms T_{2C_2} , although the difference is only 1.41%. For a practical recommendation, the T_{2C_2} test would be preferred in all cases ($d_0 = 0.10, 0.15$, and 0.20).

CONCLUSIONS

For the sample sizes $30 \leq n \leq 100$, considering the nominal significance level $\alpha = 0.05$ and for the three non-inferiority margins analyzed ($d_0 = 0.10, 0.15$, and 0.20), it is recommended to use T_{2C_2} ; that is, the Farrington-Manning test with a C_2 correction factor.

—End of the English version—



- Li, Z., and C. Chuang-Stein. 2006. A note on comparing two binomial proportions in confirmatory noninferiority trials. *Drug Information J.* 40: 203-208.
- Martin A., A., and I. Herranz T. 2002. Equivalence testing for binomial random variables: which test to use?. *The Am. Statistician* 56(3): 253-254, Letter to the editor.
- Martin A., A., and I. Herranz T. 2004a. Asymptotical test on the equivalence, substantial difference and non-inferiority problems with two proportions. *Biometrical J.* 46: 305-319.
- Martin A., A., and I. Herranz T. 2004b. Exact unconditional non-classics tests on the difference of two proportions. *Computational Statistics & Data Analysis* 45: 373-388.
- Miettinen, O., and M. Nurminen. 1985. Comparative analysis of two rates. *Statistics in Medicine* 4: 213-226.
- Röhmel, J. 2005. Problems with existing procedures to calculate exact unconditional p-values for noninferiority/superiority and confidence intervals for two binomials and how to resolve them. *Biometrical J.* 47: 37-47.

Röhmel, J., and U. Mansmann. 1999. Unconditional nonasymptotic one sided tests for independent binomial proportions when the interest lies in showing noninferiority and or superiority. *Biometrical J.* 2: 149-170.

Tu, D. 1997. A comparative study of some statistical procedures in establishing therapeutic equivalence of nonsystemic drugs with binary endpoints. *Drug Information J.* 31: 1291-1300.