



# **COLEGIO DE POSTGRADUADOS**

**INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS**

**CAMPUS MONTECILLO**

**SOCIOECONOMÍA, ESTADÍSTICA E  
INFORMÁTICA**

**ESTADÍSTICA**

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

**FÉLIX ALMENDRA ARAO**

**T E S I S**

**PRESENTADA COMO REQUISITO PARCIAL  
PARA OBTENER EL GRADO DE:**

**DOCTOR EN CIENCIAS**

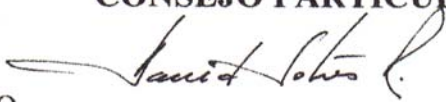

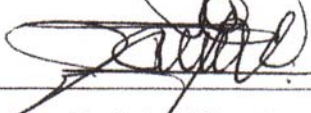
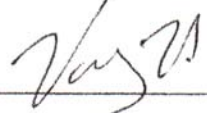
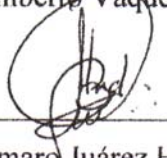
**MONTECILLO, TEXCOCO, EDO. DE MÉXICO**

**2007**

La presente tesis titulada: **Comparación de algunas pruebas estadísticas asintóticas de no-inferioridad para contrastar dos proporciones independientes** realizada por el alumno **Félix Almendra Arao** bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

**DOCTOR EN CIENCIAS**  
**SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA**  
**ESTADÍSTICA**

**CONSEJO PARTICULAR**

-   
CONSEJERO \_\_\_\_\_  
Dr. David Sotres Ramos
-   
ASESOR \_\_\_\_\_  
Dr. Ignacio Méndez Ramírez
-   
ASESOR \_\_\_\_\_  
Dr. José A. Villaseñor Alva
-   
ASESOR \_\_\_\_\_  
Dr. Humberto Vaquera Huerta
-   
ASESOR \_\_\_\_\_  
Dr. Bulmaro Juárez Hernández

## Dedicatoria

*A toda mi familia, amigos, compañeros y maestros, quienes de alguna forma contribuyeron a la culminación de esta meta.*

**especialmente**

**A mi esposa.**

*Por su cariño, comprensión y apoyo incondicional.*

**A mis hijos.**

*Mis pequeños maestros, por su impulso entusiasta.*

**A mis padres.**

*Por su motivación incansable y su gran entusiasmo ante la vida.*

**A mis hermanos.**

*Por alentarme siempre a seguir adelante.*

## Agradecimientos

*Al Instituto Politécnico Nacional por su generoso apoyo al concederme licencia para la realización de estudios doctorales.*

*Al Colegio de Postgraduados, especialmente a mis profesores, por darme la posibilidad de ampliar mi formación académica.*

*Al Dr. David Sotres Ramos por su ayuda, valiosa cooperación, apoyo, paciencia y acertada dirección así como por el enorme tiempo y dedicación empleados.*

*A los miembros de mi Consejo Particular, doctores Ignacio Méndez Ramírez, José A. Villaseñor Alva, Humberto Vaquera Huerta y Bulmaro Juárez Hernández así como a los doctores Félix V. González Cossío y Sergio Pérez Elizalde, sinodales; por sus sugerencias e invaluable contribución para la realización del presente trabajo.*

*A todo el personal administrativo, de quienes siempre obtuve amabilidad y apoyo.*

*A todos mis compañeros y amigos del Colegio de Postgraduados por su apoyo y amistad, les deseo un merecido éxito profesional.*

Comparación de algunas pruebas estadísticas  
asintóticas de no-inferioridad para contrastar dos  
proporciones independientes

**Félix Almendra Arao**

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Objetivos . . . . .	3
1.3. Revisión de la literatura . . . . .	3
1.4. Organización de la tesis . . . . .	6
<b>2. Pruebas estadísticas</b>	<b>9</b>
2.1. Aspectos preliminares . . . . .	9
2.2. El modelo teórico . . . . .	13
2.3. Propiedades asintóticas de las estadísticas de prueba . . . . .	21
2.4. Pruebas estadísticas y regiones de rechazo . . . . .	25
2.5. Correcciones por continuidad . . . . .	28
<b>3. Niveles de significancia</b>	<b>33</b>
3.1. La función de potencia . . . . .	34
3.2. Simplificación del cálculo de los niveles de significancia reales . . . . .	36
3.3. Procedimiento utilizado para calcular los niveles de significancia reales	51
3.4. Propiedades asintóticas de los niveles de significancia . . . . .	52
3.5. Clasificación de las pruebas con base en su nivel de significancia . . . . .	54
<b>4. Comparación de las pruebas</b>	<b>56</b>
4.1. Pruebas tipo $\mathcal{A}$ . . . . .	57
4.1.1. Recomendaciones. . . . .	60
4.2. Pruebas tipo $\mathcal{B}$ . . . . .	61
4.2.1. Recomendaciones. . . . .	63
<b>5. Conclusiones y posibles líneas de investigación futuras</b>	<b>65</b>
5.1. Conclusiones . . . . .	65
5.2. Posibles líneas de investigación futuras . . . . .	67

A. Validación de los cálculos	68
B. Resultados básicos y simbología	72
C. Programas de cómputo	75
C.1. Programa para verificar si una región crítica para la prueba $T_7$ cumple o no la condición (C) . . . . .	75
C.2. Programa para determinar los niveles de significancia reales para la prueba $T_7$ . . . . .	78
Bibliografía	83

## RESUMEN

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

Para pruebas exactas de no-inferioridad, Röhmel y Mansmann [25] probaron que si la región de rechazo cumple la condición de convexidad de Barnard, entonces el nivel de significancia en vez de calcularse como el supremo en todo el espacio nulo puede calcularse como el máximo en una parte de la frontera del espacio nulo. Esto tiene particular importancia debido al extenso tiempo de cómputo requerido para calcular niveles de significancia para pruebas de no-inferioridad, ver por ejemplo Röhmel [26]. En este trabajo se generaliza el teorema demostrado por Röhmel y Mansmann [25] en dos direcciones, en primer lugar se extiende el resultado para pruebas estadísticas en general (incluyendo pruebas exactas y asintóticas), en segundo lugar se relaja la condición de convexidad de Barnard a una condición menos restrictiva. El resultado incluye hipótesis de no-inferioridad para parámetros como la diferencia, la razón y la razón de momios. Este resultado permite calcular los niveles de significancia para pruebas como la de Blackwelder y la de Hauck-Anderson obteniendo el máximo en una parte de la frontera con una reducción sustancial del tiempo de cómputo.



# ABSTRACT

In this work are compared the asymptotic tests for non-inferiority of Blackwelder, Farrington-Manning, Böhning-Viwatwongkasen, Hauck-Anderson, generalized likelihood ratio test and two variants of these tests, comparison was made based in their real levels of significance and in their power. The test of Farrington-Manning has best approximation of the real significance level to the nominal one for sample size  $30 \leq n \leq 100$  and for the three non-inferiority limits more often used in clinical trials. Power of the Farrington-Manning test is very similar to power of tests with good approximation of the real level of significance to nominal.

For exact tests of non-inferiority, Röhmel and Mansmann [25] proved that if the rejection region fulfills the Barnard convexity condition, then the level of significance can be computed as the maximum in a part of the boundary of the null space instead of the supremum in the whole null space. This is particularly important due to the great amount of time required to compute levels of significance in non-inferiority tests, e.g. see Röhmel [26]. In this work, the theorem demonstrated by Röhmel and Mansmann [25] is generalized in two directions, firstly the result for general statistical tests is extended (including exact and asymptotic tests), secondly the Barnard convexity condition is relaxed to a less restrictive condition. The result includes hypotheses of non-inferiority for parameters such as difference, ratio, and odds ratio. This result allows the computing of levels of significance for tests such as the Blackwelder and the Hauck-Anderson, obtaining the maximum in one part of the boundary with a substantial reduction in computing time.

# Capítulo 1

## Introducción

### 1.1. Introducción

Las pruebas estadísticas asintóticas de no-inferioridad se utilizan frecuentemente en investigaciones clínicas con el objetivo de evaluar nuevas terapias, son especialmente usadas en el desarrollo de medicamentos. Estas pruebas sirven para demostrar que una terapia nueva (con menores efectos secundarios, menor costo o mayor facilidad de aplicación que una terapia considerada estándar) no es sustancialmente inferior en eficacia a la terapia estándar [9]. Aunque la aplicación de estas pruebas es usual en el ámbito de ensayos clínicos su aplicación en otras áreas donde se requiera contrastar dos tratamientos es también natural.

Existen diversas pruebas estadísticas para no-inferioridad que usan como medida de disimilaridad ya sea a la diferencia de proporciones, a la razón de proporciones o

bien a la razón de momios. En ninguno de estos casos hay estudios comparativos de estas pruebas donde se determinen los niveles de significancia reales de las pruebas y sus potencias, esto probablemente se debe al muy extenso tiempo de cómputo requerido para calcular explícitamente los niveles de significancia.

Debido a lo anterior es de gran utilidad un teorema demostrado por Röhmel y Mansmann [25] el cual permite reducir sustancialmente el tiempo de cómputo para calcular los niveles de significancia reales en pruebas exactas de no-inferioridad, cuando la región crítica es de una forma especial.

No obstante, hay pruebas de no-inferioridad muy utilizadas que poseen regiones críticas de un tipo más general que las referidas en dicho teorema, para las cuáles no existen resultados teóricos que permitan reducir el cálculo de sus niveles de significancia reales, haciendo complicada una posible comparación de las pruebas.

Además tampoco hay resultados teóricos que permitan la reducción del tiempo de cómputo de los niveles reales de significancia para pruebas asintóticas, incluso cuando la correspondiente región crítica es de la forma especificada en el teorema de Röhmel y Mansmann [25].

Por lo anterior y con el objeto de hacer factible la comparación de pruebas de no-inferioridad tanto exactas como asintóticas cuyas regiones críticas tengan una forma más general que la especificada en el teorema de Röhmel y Mansmann [25], se plantean en el presente trabajo los objetivos descritos en la siguiente sección.

## 1.2. Objetivos

Los objetivos que se persiguen en esta investigación son los siguientes:

1. Comparar siete pruebas asintóticas para no-inferioridad: las pruebas de Blackwelder, Farrington-Manning, Böhning-Viwatwongkasen, Hauck-Anderson, la prueba de razón de verosimilitudes generalizada y dos variantes de estas pruebas, realizar la comparación con base en sus niveles de significancia reales y sus potencias, considerando para cada prueba seis correcciones por continuidad; para los tres límites de no-inferioridad más frecuentemente utilizados en el contexto de ensayos clínicos. En tales pruebas se considera como medida de disimilaridad a la diferencia de proporciones.
2. Generalizar el teorema de Röhmel y Mansmann [25] a regiones críticas más generales que las consideradas en tal teorema.
3. Generalizar el teorema de Röhmel y Mansmann [25] a pruebas de no-inferioridad tanto exactas como asintóticas.

## 1.3. Revisión de la literatura

La literatura reporta algunas comparaciones de pruebas para no inferioridad que son basadas en simulaciones o en aproximaciones gruesas de los niveles de significancia reales y de las potencias exactas ([9], [14] y [28]). Sin embargo, en la literatura consultada no se encontró algún análisis comparativo de pruebas de no-inferioridad donde

se efectúe el cálculo explícito de los niveles de significancia reales y de las potencias de las pruebas comparadas. Esto probablemente se debe al muy extenso tiempo de cómputo requerido para calcular explícitamente dichos valores.

En 1998 Chan en [6] calculó el nivel de significancia para pruebas exactas de no inferioridad tomando el supremo no en todo el espacio nulo sino únicamente en una parte de su frontera. Chan no justificó formalmente el motivo de esta restricción, quedando un vacío en su procedimiento. El hueco en las demostraciones de Chan fué identificado en 1999 por Röhmel y Mansmann [24] quienes en una carta al editor de la revista *Statistics in Medicine* exhiben razones por las que permanece abierto el problema de si buscar el máximo solamente en la frontera es suficiente para determinar el supremo en todo el espacio nulo. Desde entonces varios intentos fueron realizados para resolver tal situación. En 1999 Röhmel y Mansmann [25] también notaron que es suficiente buscar solamente en la frontera del espacio nulo si la región crítica cumple la condición de convexidad de Barnard. En su respuesta Chan [7] (1999) ofreció un argumento heurístico del porqué en su caso la condición de convexidad de Barnard podría ser válida. Un argumento heurístico similar fué dado en 2003 por Mehrotra et al. [19], Martin y Herranz en 2004 [18] quienes presentaron un bosquejo de una demostración pero no una demostración formal. No es sino hasta 2005 en que queda zanjada tal problemática en [26] donde el autor presenta una prueba formal que justifica el procedimiento utilizado por Chan en 1998.

Para pruebas exactas de no-inferioridad, Röhmel y Mansmann [25] probaron que si

la región de rechazo cumple la condición de convexidad de Barnard, entonces el nivel de significancia en vez de calcularse como el supremo en todo el espacio nulo puede calcularse como el máximo en una parte de la frontera del espacio nulo. Posteriormente, Röhmel [26] demostró que la prueba exacta de Farrington-Manning cumple la condición de convexidad de Barnard, el teorema probado por Röhmel y Mansmann [25] permite reducir el tiempo de cómputo para calcular niveles de significancia reales a menos del 1%. Debido al enorme requerimiento en tiempo de cómputo necesario para el cálculo de niveles de significancia reales, antes del teorema de Röhmel y Mansmann [25] no era práctico el cálculo de los niveles de significancia reales para pruebas de no-inferioridad.

En esta tesis se generaliza el teorema demostrado por Röhmel y Mansmann [25] en dos aspectos, uno de ellos es que se extiende el resultado para pruebas estadísticas en general (incluyendo pruebas exactas y asintóticas), el otro aspecto es la generalización a una condición menos restrictiva que la condición de Barnard, por lo tanto esta condición es más fácil de cumplirse que la de Barnard. Este nuevo resultado permite entre otras cosas calcular los niveles de significancia reales para pruebas como la de Blackwelder y la de Hauck-Anderson para diseños desbalanceados (diferentes tamaños de muestra), obteniendo el máximo en una parte de la frontera y por tanto en un tiempo de cómputo razonable, en este caso el tiempo de cómputo se reduce aproximadamente al 1% del tiempo original.

## 1.4. Organización de la tesis

En el presente trabajo se utilizaron las condiciones de convexidad de Barnard y de simetría en la misma cola, con lo cual fue posible calcular explícitamente los niveles de significancia reales así como las potencias de las pruebas y compararlas para los tamaños de muestra  $30 \leq n \leq 100$ , para los tres límites de no-inferioridad más frecuentemente utilizados en el contexto de ensayos clínicos (.1, .15 y .2).

Para cubrir los objetivos ya mencionados, el trabajo se organizó de la siguiente forma:

En el presente capítulo se proporciona una introducción al tema de investigación, se presentan los objetivos del trabajo así como una revisión de la literatura; además se presenta la estructura del trabajo.

En el capítulo dos se presentan conceptos básicos acerca de pruebas de no inferioridad y el modelo teórico utilizado, se presentan las estadísticas de prueba que se analizarán, a saber, la estadística de Blackwelder que utiliza la desviación estándar no ponderada, la cual tiene grandes ventajas debido a su simplicidad y al hecho de que los estimadores de proporciones son los estimadores máximo verosímiles; la estadística de prueba de Hauck-Anderson, difiere de la anterior en que se disminuye en uno el tamaño de cada muestra en la estimación de la desviación estándar. La estadística de Farrington-Manning, que antes estudiaron Mietening y Nurminen, estima la desviación estándar tomando como estimadores de las proporciones a los estimadores máximo verosímiles restringidos a la hipótesis nula. La estadística de

Böhning-Viwatwongkasen en la cual la desviación estándar se estima considerando como estimadores de las proporciones una generalización de los estimadores de máxima verosimilitud. Las combinaciones de la estadística de Hauck-Anderson con la de Farrington-Manning y la de Böhning-Viwatwongkasen parecen razonables, se proponen en este trabajo y se comparan con las otras. Se considera también la estadística de razón de verosimilitudes generalizada. Además, se obtienen algunas propiedades de las pruebas estadísticas consideradas.

En el capítulo tres se calculan los niveles de significancia reales de las siete pruebas evaluadas para  $\alpha = 0.05$  y  $30 \leq n \leq 100$ , se presenta una manera de simplificar su cálculo, el uso de algunas condiciones que cumplen las regiones de rechazo de las estadísticas en cuestión, se estudia el comportamiento asintótico de los niveles de significancia reales y se realiza una comparación de los niveles de significancia reales.

El capítulo cuatro está dedicado a la comparación de las potencias de las pruebas considerando los tres límites de no-inferioridad usuales en ensayos clínicos, se hacen recomendaciones acerca de las pruebas que presentan mejor desempeño considerando sus niveles de significancia reales y potencias.

En el capítulo cinco se dan las conclusiones del trabajo.

En la parte final del trabajo se incluyó una sección con cuatro apéndices: en el primero de ellos (apéndice A) se validan los cálculos de este trabajo, comparándolos con los realizados en diversos artículos sobre el tema, en el apéndice B se da la terminología básica así como la notación utilizada en todo el trabajo y se presentan



algunos resultados conocidos, los cuales se utilizan para la obtención de resultados de interés en este trabajo, en el apéndice C se dan algunos de los programas de cómputo escritos para algunos de los cálculos realizados, tales programas se hicieron en el lenguaje S-PLUS.

# Capítulo 2

## Pruebas estadísticas

### 2.1. Aspectos preliminares

En esta sección se presentan algunos conceptos esenciales para el desarrollo de este trabajo.

Sea  $X_1, \dots, X_n$  una muestra aleatoria de  $f_\theta(x)$  donde  $\theta \in \Theta \subseteq \mathbb{R}$ , se dice que una estadística  $T_n = T_n(X_1, \dots, X_n)$  es un estimador puntual de  $\theta$  si  $T_n(\mathbb{R}^n) \subseteq \Theta$ . En general  $E(T_n) = \theta + b_n(\theta)$ ,  $b_n(\theta)$  es llamado sesgo,  $T_n$  es insesgado si  $b_n(\theta) = 0$ ,  $T_n$  es asintóticamente insesgado si  $b_n(\theta) \xrightarrow{n \rightarrow \infty} 0$ .

Sea  $\{X_n\}$  una sucesión de variables aleatorias y sea  $X$  una variable aleatoria definidas en un espacio de probabilidad. Sea  $F_{X_n}$  la función de distribución de  $X_n$  y sea  $F_X$  la función de distribución de  $X$ , se dice que  $\{X_n\}$  converge en distribución (o en ley) a  $X$  si  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$  para cada  $x$  punto de continuidad de  $F_X$ .

Se escribe  $X_n \xrightarrow{d} X$  o  $X_n \xrightarrow{L} X$ . Se dice que  $\{X_n\}$  converge en probabilidad a  $X$  si

$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$  para cada  $\varepsilon > 0$ . Se escribe  $X_n \xrightarrow{p} X$ .

Una sucesión de estimadores  $T_n$  de una función paramétrica  $g(\theta)$  es consistente si  $T_n \xrightarrow{P} g(\theta)$ .

La función de verosimilitud de  $n$  variables aleatorias  $X_1, \dots, X_n$  es la densidad conjunta de las  $n$  variables aleatorias  $f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$  considerada como función de  $\theta$  y se denota por  $L(\theta; x_1, \dots, x_n)$ .

El estimador de máxima verosimilitud  $\hat{\theta}$  de  $\theta$  es tal que  $L(\hat{\theta}; \mathbf{x}) = \max_{\theta \in \Theta} L(\theta; \mathbf{x})$ , donde  $\mathbf{x} = (x_1, \dots, x_n)$ .

Dada una muestra aleatoria  $X_1, \dots, X_n$  de la densidad  $f(x; \theta)$ , una estadística  $S(\mathbf{X})$ , donde  $\mathbf{X} = (X_1, \dots, X_n)$ , es suficiente si la distribución condicional de  $T$  dado  $S$  no depende de  $\theta$  para toda estadística  $T(\mathbf{X})$ .

Bajo ciertas condiciones de regularidad que si  $T$  es un estimador de  $g(\theta)$ , entonces  $Var_{\theta}(T) \geq \frac{[g'(\theta)]^2}{nI(\theta)}$ , donde  $I(\theta) = E \left[ \frac{\partial \ln L(\theta; \mathbf{X})}{\partial \theta} \right]^2 \cdot \frac{[g'(\theta)]^2}{nI(\theta)}$  es llamada la cota inferior de Crámer-Rao (CICR).

Una sucesión  $\{T_n\}$  de estimadores de  $g(\theta)$  se dice mejor asintóticamente normal (MAN) si es consistente, asintóticamente normal y la varianza de su distribución límite es mínima en la clase de todas las sucesiones de estimadores asintóticamente normales.

Los estimadores de máxima verosimilitud tienen las propiedades siguientes:

1) No necesariamente son insesgados. Si no son insesgados, son asintóticamente

insesgados.

2) Si existe un estimador cuya varianza alcanza la cota inferior de Crámer-Rao, es el estimador de máxima verosimilitud. [27]

3) Bajo condiciones de regularidad, son consistentes

4) Toda sucesión de estimadores de máxima verosimilitud es MAN.

5) Son funciones de las estadísticas suficientes.

6) Son invariantes bajo transformaciones 1-1 del parámetro.

Si el espacio de parámetros  $\Theta$  se particiona en dos subconjuntos  $\Theta_0$  y  $\Theta_1$  tal que  $\Theta = \Theta_0 \cup \Theta_1$ , un procedimiento estadístico de prueba o simplemente prueba es una función sobreyectiva  $T : \chi \rightarrow \{\Theta_0, \Theta_1\}$ , donde  $\chi$  es el espacio muestral. En otras palabras, un procedimiento estadístico de prueba es una regla para seleccionar una y sólo una de las alternativas sobre la base de los datos observados. Frecuentemente tales alternativas se representan como  $H_0 : \theta \in \Theta_0$  (hipótesis nula) y  $H_a : \theta \in \Theta_1$  (hipótesis alternativa). Esencialmente un procedimiento de prueba puede verse como una partición del espacio muestral, donde  $T^{-1}(\Theta_0)$ , denotado por  $\chi_0$  representa la región de aceptación de la hipótesis nula, mientras que  $\chi - T^{-1}(\Theta_0) = T^{-1}(\Theta_1)$ , denotado por  $\chi_1$  representa la región de rechazo o región crítica de la hipótesis nula. Usualmente esta partición del espacio muestral es inducida por una estadística, llamada estadística de prueba.

Se especifica como hipótesis nula aquélla hipótesis que implica el valor existente del parámetro, o la que se supone más estable, siendo necesaria una elevada evidencia

para rechazarla.

Un problema de prueba de hipótesis implica la posibilidad de equivocarse o de acertar en la elección al no saber con certeza cuál es la verdadera, generando con ello dos posibles decisiones erróneas. Hay dos posibilidades de equivocarse. Una sucede si la hipótesis nula se rechaza y es cierta, llamado error tipo I; la otra posibilidad es cuando la hipótesis alternativa se rechaza y es cierta, en este caso es llamado error tipo II. Existen otras dos posibilidades en las cuales la decisión es correcta: si la hipótesis nula no se rechaza y es cierta, la otra es cuando la hipótesis alternativa no se rechaza y es cierta.

Puesto que el experimentador desconoce qué hipótesis es la correcta, no sabe en cuál de las cuatro situaciones descritas se encuentra. Ante la posibilidad de cometer un error, se controla la ocurrencia de errores en términos probabilísticos.

Sea  $T$  una prueba de la hipótesis nula  $H_0$ , la función de potencia de la prueba  $T$ , denotada por  $\beta_T(\theta)$  se define como  $\beta_T(\theta) = P_\theta(\text{rechazar } H_0)$ .

Sea  $T$  una prueba de la hipótesis  $H_0 : \theta \in \Theta_0 \subseteq \Theta$  el nivel de significancia o tamaño de la prueba  $T$  de  $H_0$  se define como  $\sup_{\theta \in \Theta_0} \beta_T(\theta)$ .

Una prueba  $T$  de tamaño  $\alpha$  de  $H_0 : \theta \in \Theta_0$  contra la hipótesis alternativa  $H_1 : \theta \in \Theta_1$  es una prueba insesgada si  $\alpha \leq \beta_T(\theta) \forall \theta \in \Theta_1$ .

Una sucesión de pruebas estadísticas  $\{T_n\}$  para el problema de prueba de hipótesis  $H_0 : \theta \in \Theta_0$  vs  $H_a : \theta \in \Theta_1$  se dice consistente contra la alternativa si  $\beta_{T_n}(\theta) \xrightarrow[n \rightarrow \infty]{} 1 \forall \theta \in \Theta_1$ .

## 2.2. El modelo teórico

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

$$[H_0 : d \geq d_0] \text{ vs } [H_a : d < d_0] \quad (2.1)$$

donde  $d = p_1 - p_2$  y  $d_0$  es el límite de no-inferioridad que es una constante positiva.

En el contexto de ensayos clínicos los valores usuales para  $d_0$  son 0.10, 0.15 y 0.20.

Sean  $X_1 = \sum_{j=1}^{n_1} X_{1j}$ ,  $X_2 = \sum_{j=1}^{n_2} X_{2j}$ , de acuerdo al modelo arriba mencionado. El espacio muestral con base en las estadísticas  $X_1$  y  $X_2$  se define como  $\chi = (x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\}$ .

Se define el **espacio muestral lícito** como  $\chi' = \{(x_1, x_2) \in \chi : \hat{p}_1 - \hat{p}_2 < d_0\}$ , donde  $\hat{p}_j = X_j/n_j$  es el estimador de máxima verosimilitud de  $p_j$  para  $j = 1, 2$ . La importancia del concepto anterior es reiterada por Martin y Herranz en [16], [17] y [18] quienes enfatizan la necesidad de tal concepto ya que de no restringirse el espacio

muestral de esta manera, 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  $\widehat{p}_1 - \widehat{p}_2 \geq d_0$ , entonces es inferencialmente ilógico concluir que  $p_1 - p_2 < d_0$ , en la sección 2.4 se incorpora un resultado relacionado con este concepto y se ilustra mediante un ejemplo la diferencia que puede haber entre las correspondientes regiones de rechazo.

El espacio de parámetros es  $\Theta = \{(p_1, p_2) : p_1, p_2 \in [0, 1]\}$  y en virtud de que la variable aleatoria  $X_i$  tiene distribución binomial con parámetros  $(n_i, p_i)$  para  $i = 1, 2$ , la función de verosimilitud conjunta es

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

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

En este trabajo (excepto en la sección 3.2), se supone que  $T$  es una estadística de prueba para el problema de prueba de hipótesis (2.1), con región crítica de la forma  $R_T = \{(x_1, x_2) \in \chi : T(x_1, x_2) \leq t_0\}$ , donde  $t_0$  es una constante; nótese que con ello quedan incluidas tanto pruebas exactas como asintóticas. Cuando se desee enfatizar el nivel nominal de la prueba, la región crítica se denotará por  $R_T(\alpha)$  o simplemente por  $R(\alpha)$  cuando se sobreentienda la estadística de prueba  $T$ .

En lo sucesivo la función masa de probabilidad de una variable aleatoria binomial

$X$  con parámetros  $n$  y  $p$  será denotada por  $f(x; n, p)$ , es decir

$$f(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, x \in \{0, 1, \dots, n\}; p \in [0, 1]$$

Seis de las estadísticas de prueba consideradas son del tipo

$$T_i(X_1, X_2) = \frac{\hat{p}_1 - \hat{p}_2 - d_0}{\hat{\sigma}_i} \quad (2.2)$$

para  $i = 1, 2, 3, 4, 5, 6$ , donde  $\hat{p}_j = X_j/n_j$  es el estimador de máxima verosimilitud de  $p_j$  para  $j = 1, 2$  y  $\hat{\sigma}_i$  es un estimador 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 generalizada, es decir,

$$T_7(X_1, X_2) = \Lambda(X_1, X_2) = \frac{\sup_{\theta \in \Theta_0} L(p_1, p_2; X_1, X_2)}{\sup_{\theta \in \Theta} L(p_1, p_2; X_1, X_2)} \quad (2.3)$$

La diferencia entre las seis estadísticas del tipo (2.2) radica en la estimación que se elige para la desviación estándar de  $\hat{d}$ . Se consideran los siguientes seis estimadores.

$$\begin{aligned} \hat{\sigma}_1 &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} & \hat{\sigma}_4 &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2-1}} \\ \hat{\sigma}_2 &= \sqrt{\frac{\check{p}_1(1-\check{p}_1)}{n_1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2}} & \hat{\sigma}_5 &= \sqrt{\frac{\check{p}_1(1-\check{p}_1)}{n_1-1} + \frac{\check{p}_2(1-\check{p}_2)}{n_2-1}} \\ \hat{\sigma}_3 &= \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2}} & \hat{\sigma}_6 &= \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1-1} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2-1}} \end{aligned}$$

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 de  $p_i$  restringido bajo la hipótesis nula, además,  $\tilde{p}_i = \frac{X_i+1}{n_i+2}$  para  $i = 1, 2$ .



La estadística  $T_1$  se construye usando la estimación clásica para la desviación estándar, nótese que  $\hat{\sigma}_1 = 0$  para  $\hat{p}_i = 0, 1$ , en estos casos  $T_1$  no está definida, en breve se da una redefinición de  $\hat{\sigma}_1$  para estos valores con el fin de evitar dicha dificultad.

**Observación 1** Sea  $\varphi(n, x) = \frac{\frac{x}{n}(1-\frac{x}{n})}{n} = \frac{x(n-x)}{n^3}$ , entonces

1.  $\varphi(n, n-x) = \varphi(n, x)$
2.  $\hat{\sigma}_1(x, y) = \sqrt{\varphi(n_1, x) + \varphi(n_2, y)}$
3.  $\varphi$  es creciente en  $n \Leftrightarrow x > \frac{2}{3}n$
4.  $\varphi$  es decreciente en  $n \Leftrightarrow x < \frac{2}{3}n$
5.  $\varphi$  es creciente en  $x \Leftrightarrow x < \frac{n}{2}$
6.  $\varphi$  es decreciente en  $x \Leftrightarrow x > \frac{n}{2}$

Como  $\hat{\sigma}_1(x, y) = 0$  para  $(x, y) \in \{0, n_1\} \times \{0, n_2\}$  resulta que  $T_1(x, y)$  no está definida para  $(x, y) \in \{0, n_1\} \times \{0, n_2\}$ , por ello en este trabajo se propone redefinir a  $\hat{\sigma}_1(x, y)$  en esos puntos del espacio muestral, de la siguiente forma

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

de la observación 1(1) se obtiene de inmediato que todos los valores redefinidos son iguales, es decir,  $\hat{\sigma}_1(0, 0) = \hat{\sigma}_1(0, n_2) = \hat{\sigma}_1(n_1, 0) = \hat{\sigma}_1(n_1, n_2)$ .

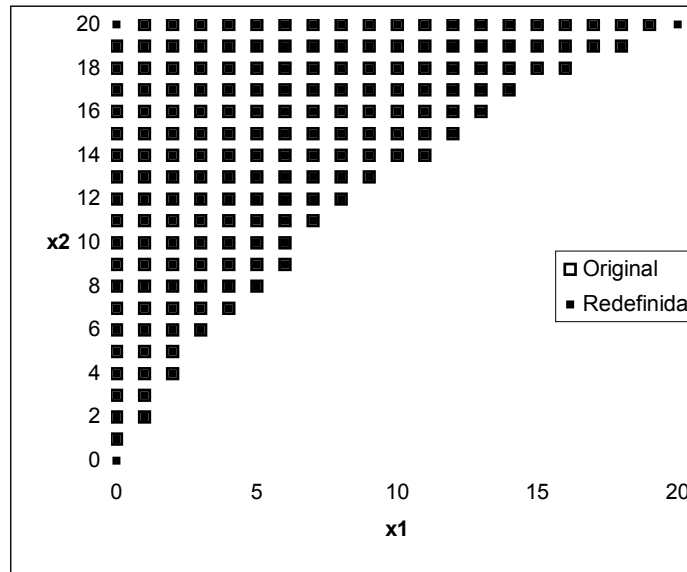


Figura 2.1: Regiones de rechazo original y redefinida para la estadística  $T_1$  para los tamaños de muestra  $n_1 = n_2 = 20$  tomando el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no inferioridad  $d_0 = 0.1$ .

La anterior redefinición de  $\hat{\sigma}_1$  es esencial como se verá en la sección 3.2 debido a que con ella se logra que la región crítica sea de un tipo especial que permite el cálculo del nivel de significancia de una forma mucho más simple.

Nótese que al igual que  $\hat{\sigma}_1, \hat{\sigma}_4 = 0$  para  $\hat{p}_i = 0, 1$ , en estos casos  $T_4$  no está definida, en este trabajo se propone redefinir a  $\hat{\sigma}_4$  de la misma forma que se hizo con  $\hat{\sigma}_1$ .

El método para estimar a  $\hat{\sigma}$  mediante  $\hat{\sigma}_2$  fue propuesto por Miettinen y Nurminen en [20] y se basa en la idea de estimar a  $p_1$  y  $p_2$  tomando los estimadores de máxima verosimilitud de  $p_1$  y  $p_2$  bajo la hipótesis nula, denotados respectivamente por  $\hat{p}_1^{\vee}$  y  $\hat{p}_2^{\vee}$ .

Munk et al. [21] demostraron el siguiente resultado

**Lema 2** *Sea  $g$  una función continua y creciente, y sea el problema de prueba de*

hipótesis

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

con  $g : [0, 1] \rightarrow [0, 1]$  función continua y no decreciente con  $g \neq 1$ .

a) Entonces el estimador de máxima verosimilitud restringido a la hipótesis nula ( $\hat{\theta}^{\vee}$ ) existe y está dado como  $\hat{\theta}^{\vee} = \hat{\theta}$  (el estimador de máxima verosimilitud), si  $\hat{\theta} \in \Theta_0$  y  $\hat{\theta}^{\vee} = \arg \left\{ \max_{p_2=g(p_1)} L(\theta) \right\}$  si  $\hat{\theta} \notin \Theta_0$ .

b) Sea  $h \in C^2[0, 1]$ , entonces el estimador de máxima verosimilitud restringido a la hipótesis nula ( $\hat{\theta}^{\vee}$ ) es único si y sólo si las siguientes condiciones i) y ii) se cumplen o si se cumple iii) sobre el conjunto  $\Theta_0^* = \{(p_1, p_2) \in \Theta : p_2 = g(p_1)\}$

$$i) -(h')^2 + hh'' \leq 0$$

$$ii) -(h')^2 - h'' + hh'' \leq 0$$

iii)  $h$  es convexa.

Usando este resultado se calculan los estimadores de máxima verosimilitud restringidos en la curva  $p_2 = p_1 - d_0$  (es decir, cuando  $g(p_1) = p_1 - d_0$ ) la cual es una parte de la frontera del espacio nulo.

De acuerdo con [20] estos estimadores están dados de la siguiente forma:

Sean

$$r = -\frac{x_1 + x_2 + n_1(1 + 2d_0) + n_2(1 + d_0)}{n_1 + n_2}$$

$$s = \frac{x_2 + x_1(1 + 2d_0) + d_0(n_2 + n_1(1 + d_0))}{n_1 + n_2}$$

$$t = \frac{-x_1 d_0(1 + d_0)}{n_1 + n_2}$$

$$\check{p}_1 = 2 \frac{\sqrt{r^2 - 3s}}{3} \cos \left( \frac{1}{3} \arccos \left( -\frac{\frac{2r^3}{27} - rs/3 + t}{2 \left( \frac{\sqrt{r^2 - 3s}}{3} \right)^3} \right) + \frac{4}{3}\pi \right) - \frac{r}{3}$$

$$\check{p}_2 = \check{p}_1 - d_0$$

Agresti y Caffo [1] entre otros han considerado el problema de estimar el parámetro binomial  $p$  destacando que  $\frac{X}{n}$  podría no ser una buena elección como un estimador de  $p$  cuando el tamaño de muestra es pequeño. En efecto, el estimador de máxima verosimilitud  $\hat{p} = \frac{X}{n}$  tiene varios inconvenientes aún cuando el tamaño de muestra no sea pequeño. Por ejemplo, la varianza de  $\hat{p}$  cuando se estima por  $\hat{p}(1-\hat{p})/n$  es 0 para  $x = 0$  y  $x = n$ , este problema es más severo cuando se combinan estimadores de proporciones diferentes como es el caso del estimador  $Var(\hat{d})$ , como se vió antes (algunos otros inconvenientes del uso de estimadores de máxima verosimilitud para estimar proporciones cuando se desea estimar la desviación estándar de  $\hat{d}$  se establecerán en la sección 3.2). A la luz de esa dificultad, Böhning y Viwatwongkasen en [4] sugieren usar una clase de estimadores de forma paramétrica, a saber,  $\tilde{p}_b = \frac{X+b}{n+2b}$ ,  $b \geq 0$ . Varias elecciones para  $b$  como  $1/6$ ,  $1/2$ ,  $1$  o  $2$  han sido sugeridas, estas sugerencias frecuentemente tienen bases intuitivas y una de ellas tiene una justificación bayesiana. Aquí se usará el valor  $b = 1$  pues es adecuado para las pruebas donde se ocupará, recuérdese además que el estimador así obtenido corresponde al estimador de Bayes cuando se toma como distribución a priori a la distribución  $Beta(1, 1)$ .

Si se usa  $\tilde{p}_{i,b} = \frac{x_i+b}{n_i+2b}$  para  $i = 1, 2$  se tiene como estimador de la desviación estándar  $\hat{\sigma}_3 = \sqrt{\frac{\tilde{p}_{1,b}(1-\tilde{p}_{1,b})}{n_1} + \frac{\tilde{p}_{2,b}(1-\tilde{p}_{2,b})}{n_2}}$ , nótese que  $\hat{\sigma}_3 \neq 0$  para todo  $b > 0$ , desde esta perspectiva, esta es una propiedad muy conveniente ya que evita la indeterminación de la correspondiente estadística  $T$ . Además la estadística  $T_1$  es el caso particular de ésta cuando  $b = 0$ . Böhning y Viwatwongkasen solamente sugieren el uso del estimador  $\hat{\sigma}_3$  pero no realizan análisis alguno en relación al comportamiento de la correspondiente estadística de prueba. En el presente trabajo se analiza la prueba asintótica construída usando esta estadística de prueba.

**Observación 3** *Note que*

1.  $\tilde{p}_{i,b} \in (0, 1)$  para toda  $b > 0$
2.  $\lim_{b \rightarrow 0} \tilde{p}_{i,b} = \hat{p}_i$
3.  $\lim_{b \rightarrow 0} E(\tilde{p}_{i,b}) = \lim_{b \rightarrow 0} \frac{n_i p_i + b}{n_i + 2b} = p_i = E(\hat{p}_i)$
4.  $\lim_{n_i \rightarrow \infty} E(\tilde{p}_{i,b}) = \lim_{n_i \rightarrow \infty} \frac{n_i p_i + b}{n_i + 2b} = p_i = E(\hat{p}_i)$
5.  $Var(\tilde{p}_{i,b}) = \frac{n_i p_i (1 - p_i)}{(n_i + 2b)^2} \leq \frac{p_i (1 - p_i)}{n_i} = var(\hat{p}_i)$
6.  $\lim_{b \rightarrow 0} Var(\tilde{p}_{i,b}) = var(\hat{p}_i)$
7.  $\lim_{n_i \rightarrow \infty} Var(\tilde{p}_{i,b}) = 0 = \lim_{n_i \rightarrow \infty} Var(\hat{p}_i)$

Para los objetivos del presente trabajo se considera el valor  $b = 1$ , en todos los cálculos realizados usando a este estimador. Una vez hecha esta aclaración en lugar de  $\tilde{p}_{i,b}$  se escribirá simplemente  $\tilde{p}_i$  para denotar a tal estimador.

El estimador de  $\sigma$ ,  $\hat{\sigma}_4$  fue propuesto por Hauck y Anderson en [12].

$\hat{\sigma}_5$  es una combinación de  $\hat{\sigma}_2$  y  $\hat{\sigma}_4$  mientras que  $\hat{\sigma}_6$  es una combinación de  $\hat{\sigma}_3$  y  $\hat{\sigma}_4$ .

## 2.3. Propiedades asintóticas de las estadísticas de prueba

En esta sección se utilizan varios resultados asintóticos (teoremas 32 al 37) cuyos enunciados se presentan en el apéndice B.

Sea  $N = n_1 + n_2$ , en relación a la hipótesis  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$  de la proposición 37, un caso muy simple en el que se cumple es, por ejemplo cuando  $n_2 = kn_1$  con  $k \in \mathbb{N}$ , entonces  $\frac{n_1}{N} = \frac{n_1}{n_1 + kn_1} = \frac{1}{1+k}$  por lo tanto  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \frac{1}{1+k} \in (0, 1)$ .

En la proposición siguiente así como en su prueba se usará la notación  $\hat{\sigma}_N = \sqrt{\frac{\hat{\sigma}_{n_1}^2}{n_1} + \frac{\hat{\tau}_{n_2}^2}{n_2}}$ .

**Proposición 4** Sean  $V_1, \dots, V_{n_1}$  y  $W_1, \dots, W_{n_2}$  independientes con distribuciones  $F_1$  y  $F_2$ , respectivamente; medias respectivas  $\mu_1$  y  $\mu_2$  y varianzas finitas respectivas  $\sigma^2 > 0$  y  $\tau^2 > 0$ . Sean  $N = n_1 + n_2$ ,  $\sigma_N = \sqrt{\frac{\sigma^2}{n_1} + \frac{\tau^2}{n_2}}$ . Supóngase que  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$  sean  $\{\hat{\sigma}_{n_1}\}$  una sucesión consistente de estimadores de  $\sigma$  y  $\{\hat{\tau}_{n_2}\}$  una sucesión consistente de estimadores de  $\tau$  y  $\hat{\sigma}_N = \sqrt{\frac{\hat{\sigma}_{n_1}^2}{n_1} + \frac{\hat{\tau}_{n_2}^2}{n_2}}$ . Entonces

$$\frac{\bar{V} - \bar{W} - (\mu_1 - \mu_2)}{\hat{\sigma}_N} \xrightarrow{L} N(0, 1)$$

**Demostración.** Por hipótesis  $(\hat{\sigma}_{n_1}, \hat{\tau}_{n_2}) \xrightarrow{P} (\sigma, \tau)$ , también por hipótesis se tiene

que  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$ , entonces por el teorema 32

$$\frac{N\hat{\sigma}_N^2}{\frac{\hat{\sigma}_{n_1}^2}{\zeta} + \frac{\hat{\tau}_{n_2}^2}{1-\zeta}} = \frac{N \left( \frac{\hat{\sigma}_{n_1}^2}{n_1} + \frac{\hat{\tau}_{n_2}^2}{n_2} \right)}{\frac{\hat{\sigma}_{n_1}^2}{\zeta} + \frac{\hat{\tau}_{n_2}^2}{1-\zeta}} \xrightarrow{P} \frac{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}}{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}} = 1$$

además

$$\frac{N\sigma_N^2}{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}} = \frac{N \left( \frac{\sigma^2}{n_1} + \frac{\tau^2}{n_2} \right)}{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}} \xrightarrow{P} \frac{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}}{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}} = 1$$

así de las dos expresiones anteriores y el teorema 32 se tiene

$$\frac{\frac{N\hat{\sigma}_N^2}{\frac{\hat{\sigma}_{n_1}^2}{\zeta} + \frac{\hat{\tau}_{n_2}^2}{1-\zeta}}}{\frac{N\sigma_N^2}{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}}} = \frac{\hat{\sigma}_N^2 \left( \frac{\hat{\sigma}_{n_1}^2}{\zeta} + \frac{\hat{\tau}_{n_2}^2}{1-\zeta} \right)}{\hat{\sigma}_N^2 \left( \frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta} \right)} \xrightarrow{P} 1$$

además

$$\frac{\frac{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}}{\frac{\hat{\sigma}_{n_1}^2}{\zeta} + \frac{\hat{\tau}_{n_2}^2}{1-\zeta}}}{\frac{\sigma^2}{\zeta} + \frac{\tau^2}{1-\zeta}} \xrightarrow{P} 1$$

aplicando el teorema 32 a las dos últimas expresiones se obtiene

$$\frac{\sigma_N^2}{\hat{\sigma}_N^2} \xrightarrow{P} 1$$

y por el teorema 33  $\frac{\sigma_N}{\hat{\sigma}_N} \xrightarrow{P} 1$ , finalmente de la proposición 37 y el teorema 34 se

obtiene

$$\frac{\bar{V} - \bar{W} - (\mu_1 - \mu_2)}{\hat{\sigma}_N} = \frac{\bar{V} - \bar{W} - (\mu_1 - \mu_2)}{\sigma_N} \frac{\sigma_N}{\hat{\sigma}_N} \xrightarrow{L} N(0, 1)$$

■

**Proposición 5** Sean  $V_1, \dots, V_{n_1} \sim \text{Ber}(p_1)$ ;  $W_1, \dots, W_{n_2} \sim \text{Ber}(p_2)$ , independientes

$X_1 = \sum_{i=1}^{n_1} V_i$   $X_2 = \sum_{i=1}^{n_2} W_i$  supóngase que  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$ . Entonces

$$\frac{\frac{X_1}{n_1} - \frac{X_2}{n_2} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \xrightarrow{L} N(0, 1)$$

es decir,  $\frac{\hat{d}-d}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \xrightarrow{L} N(0, 1)$

**Demostración.**

$$X_1 \sim \text{Bin}(n_1, p_1), \quad X_2 \sim \text{Bin}(n_2, p_2)$$

$E[\bar{V}] = E\left[\frac{X}{n_1}\right] = p_1$ ;  $E[\bar{W}] = E\left[\frac{Y}{n_2}\right] = p_2$  y como  $\text{Var}[V_i] = p_1(1-p_1)$ ,  $\text{Var}[W_i] = p_2(1-p_2)$  entonces de la proposición 37 se tiene

$$\frac{\frac{X_1}{n_1} - \frac{X_2}{n_2} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \xrightarrow{L} N(0, 1)$$

■

**Proposición 6** Sean  $V_1, \dots, V_{n_1} \sim \text{Ber}(p_1)$ ;  $W_1, \dots, W_{n_2} \sim \text{Ber}(p_2)$ , independientes,

$X_1 = \sum_{i=1}^{n_1} V_i$  y  $X_2 = \sum_{i=1}^{n_2} W_i$  y supóngase que  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$ . Entonces

1.  $\{\hat{\sigma}_{n_1} = \hat{p}_1(1 - \hat{p}_1)\}$  y  $\{\tilde{\sigma}_{n_1} = \tilde{p}_1(1 - \tilde{p}_1)\}$  son sucesiones consistentes de estimadores de  $\sigma = p_1(1 - p_1)$ ; donde  $\hat{p}_1 = \frac{X_1}{n_1}$  es el estimador de máxima verosimilitud de  $p_1$  y  $\tilde{p}_1 = \frac{X_1+b}{n_1+2b}$  con  $b$  un número real positivo.
2.  $\{\hat{\sigma}_{n_2} = \hat{p}_2(1 - \hat{p}_2)\}$  y  $\{\tilde{\sigma}_{n_2} = \tilde{p}_2(1 - \tilde{p}_2)\}$  son sucesiones consistentes de estimadores de  $\tau = p_2(1 - p_2)$ ; donde  $\hat{p}_2 = \frac{X_2}{n_2}$  es el estimador de máxima verosimilitud de  $p_2$  y  $\tilde{p}_2 = \frac{X_2+b}{n_2+2b}$  con  $b$  un número real positivo.

**Demostración.**

1. Que  $\{\hat{\sigma}_{n_1}\}$  es una sucesión consistente de estimadores de  $\tau$  se sigue de los teoremas 33 y 35 y de que  $\hat{p}_1$  son estimadores consistentes de  $p_1$ .



Para probar que  $\{\tilde{\sigma}_{n_1}\}$  es una sucesión consistente de estimadores de  $\sigma$ , por los teoremas 33 y 35 basta verificar que  $E[\tilde{p}_1] \rightarrow p_1$  y  $Var[\tilde{p}_1] \rightarrow 0$ . En efecto,

$$E[\tilde{p}_1] = E\left[\frac{X+b}{n_1+2b}\right] = \frac{n_1 p_1 + b}{n_1 + 2b} \rightarrow p_1 \text{ y } Var[\tilde{p}_1] = Var\left[\frac{X+b}{n_1+2b}\right] = \frac{n_1 p_1 (1-p_1)}{(n_1+2b)^2} \rightarrow 0.$$

2. Análoga al inciso anterior.

■

**Proposición 7** Sean  $V_1, \dots, V_{n_1} \sim Ber(p_1)$ ;  $W_1, \dots, W_{n_2} \sim Ber(p_2)$ , independientes,

$X_1 = \sum_{i=1}^{n_1} V_i$  y  $X_2 = \sum_{i=1}^{n_2} W_i$  y supongamos que  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$ . Entonces

$$\frac{\frac{X_1}{n_1} - \frac{X_2}{n_2} - (p_1 - p_2)}{\hat{\sigma}_i} \xrightarrow{L} N(0, 1)$$

para  $i = 1, 3, 4, 6$ .

**Demostración.** Inmediata de las tres proposiciones anteriores. ■

**Observación 8** Si  $f(x; n, p)$  es la función masa de probabilidad de una variable aleatoria  $X$  con distribución  $B(n, p)$ , entonces

1.  $f(n-x; n, 1-p) = f(x; n, p)$
2.  $\frac{\partial f(x; n, p)}{\partial p} = n[f(x-1; n-1, p) - f(x; n-1, p)]$
3.  $\frac{\partial f(x; n, p)}{\partial p} = f(x, n, p) \frac{n(\hat{p}-p)}{p(1-p)}$

**Demostración.** Inmediata. ■

Puesto que para que se cumpla la convergencia asintótica, esencialmente se ha impuesto en términos generales la condición  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{n_1 + n_2} = \zeta \in (0, 1)$ , de aquí en adelante siempre que se trate con resultados asintóticos en relación a las estadísticas mencionadas, se supondrá que se cumple tal condición, nótese también que tal condición es equivalente a  $\lim_{\min(n_1, n_2) \rightarrow \infty} \frac{n_1}{n_2} = k \in (0, 1)$  con  $k = \frac{\zeta}{1-\zeta}$ .

De los resultados previos se tiene que las estadísticas de interés, es decir  $T_i$  para  $i = 1, \dots, 6$  convergen en distribución a la distribución normal estándar.

Se verá ahora el comportamiento asintótico de la estadística de razón de verosimilitudes generalizada.

Munk et al. [21] demostraron que si  $h : [0, 1] \rightarrow [0, 1]$  es creciente y  $h \in C^{(1)}[0, 1]$ , entonces bajo las hipótesis del Lema 2 se tiene para  $p_2 = g(p_1)$  y para cualquier solución  $\check{\theta}$ ,  $-2 \ln \Lambda \xrightarrow{D} U \sim \frac{1}{2} + \frac{1}{2} F_{\chi_1^2}$  cuando  $\min(n_1, n_2) \rightarrow \infty$  y  $\frac{n_1}{n_2} \rightarrow k > 0$  donde  $F_{\chi_1^2}$  denota la función de distribución acumulada de una variable  $\chi_1^2$ .

## 2.4. Pruebas estadísticas y regiones de rechazo

Las pruebas asintóticas para las estadísticas del tipo (2.2) para un nivel de significancia nominal  $\alpha$  son:

Prueba  $T_i$  : rechazar  $[H_0 : d \geq d_0]$  en favor de  $[H_a : d < d_0]$  (ver 2.1) si y sólo si  $T_i(x_1, x_2) < -z_\alpha$  y en consecuencia tienen región de rechazo de la forma  $R_{T_i}(\alpha) = \{(x_1, x_2) \in \chi : T_i(x_1, x_2) < -z_\alpha\}$  donde  $z_\alpha$  es el percentil superior  $\alpha$  de la distribución normal estándar, es decir,  $P(Z > z_\alpha) = \alpha$ ; para  $i = 1, 2, 3, 4, 5, 6$  donde  $Z$  es una

variable aleatoria con distribución normal estándar.

Nótese que estas pruebas son consistentes para el caso  $n = n_1 = n_2$  ya que si se denota a  $\hat{d}$  mediante  $\hat{d}_n$ , a  $\hat{\sigma}$  mediante  $\hat{\sigma}_n$  y a cualquiera de estas seis pruebas por  $W_n$ , entonces  $W_n = \frac{\hat{d}_n - d_0}{\hat{\sigma}_n} = \frac{\hat{d}_n - d}{\hat{\sigma}_n} + \frac{d - d_0}{\hat{\sigma}_n} \xrightarrow{n \rightarrow \infty} -\infty$  ya que  $\frac{\hat{d}_n - d}{\hat{\sigma}_n}$  tiene distribución normal estándar; por lo tanto  $P(W_n < -z_\alpha) \xrightarrow{n \rightarrow \infty} 1$ .

La prueba asintótica para la estadística del tipo (2.3) es:

Prueba  $T_7$  : rechazar  $[H_0 : d \geq d_0]$  en favor de  $[H_a : d < d_0]$  (ver 2.1) si y sólo si  $-2 \ln \Lambda(x_1, x_2) > \chi_{1-2\alpha}^2(1)$  y en consecuencia tiene región de rechazo de la forma  $R_{T_7}(\alpha) = \{(x_1, x_2) \in \chi : -2 \ln \Lambda(x_1, x_2) > \chi_{1-2\alpha}^2(1)\}$  donde  $\chi_{1-2\alpha}^2(1)$  es el percentil superior  $1 - 2\alpha$  de la distribución ji-cuadrada con un grado de libertad, es decir,  $P(\chi_1^2 > \chi_{1-2\alpha}^2(1)) = 1 - 2\alpha$ .

A manera de ilustración, en las figuras 2.2 y 2.3 se presenta la forma de las regiones de rechazo para las pruebas  $T_1, T_2, T_3$  y  $T_7$  para el tamaño de muestra 30 y límite de no inferioridad  $d_0 = 0.1$  considerando el nivel de significancia nominal  $\alpha = 0.05$ .

**Proposición 9** *Para estadísticas tipo (2.2), el espacio muestral coincide con el espacio muestral lícito, es decir, si  $T$  es una estadística del tipo (2.2) y  $t_0 < 0$  si  $R_T = \{(x_1, x_2) \in \chi : T(x_1, x_2) \leq t_0\}$  y  $R'_T = \{(x_1, x_2) \in \chi' : T(x_1, x_2) \leq t_0\}$  son la región de rechazo y la región de rechazo restringida al espacio muestral lícito, respectivamente, entonces  $R_T = R'_T$ .*

**Demostración.** Por definición  $R'_T \subseteq R_T$ . Supóngase que  $R_T - R'_T \neq \emptyset$  entonces existe  $(x, y) \in R_T - R'_T$  puesto que  $(x, y) \in R_T$  entonces  $\frac{\hat{d} - d_0}{\hat{\sigma}} < t_0$ , además, como

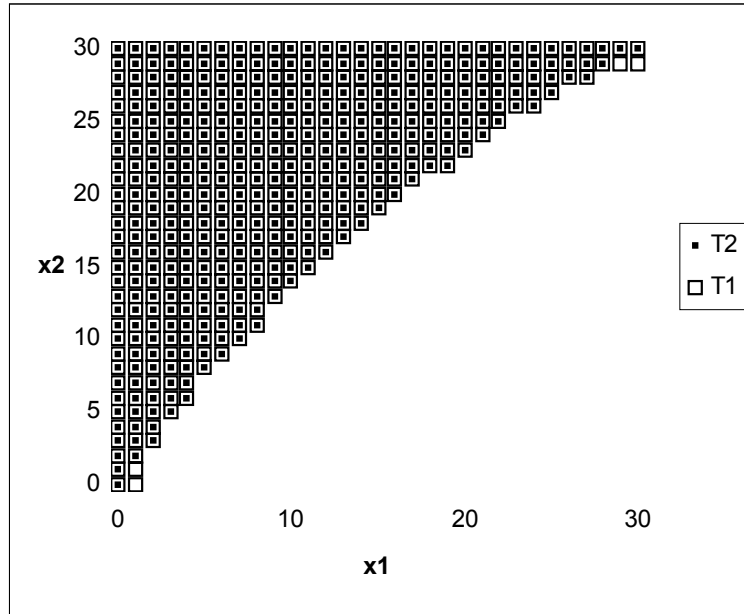


Figura 2.2: Regiones de rechazo para las pruebas  $T_1$  y  $T_2$  para el tamaño de muestra 30, límite de no inferioridad  $d_0 = 0.1$  y nivel de significancia nominal  $\alpha = 0.05$ .

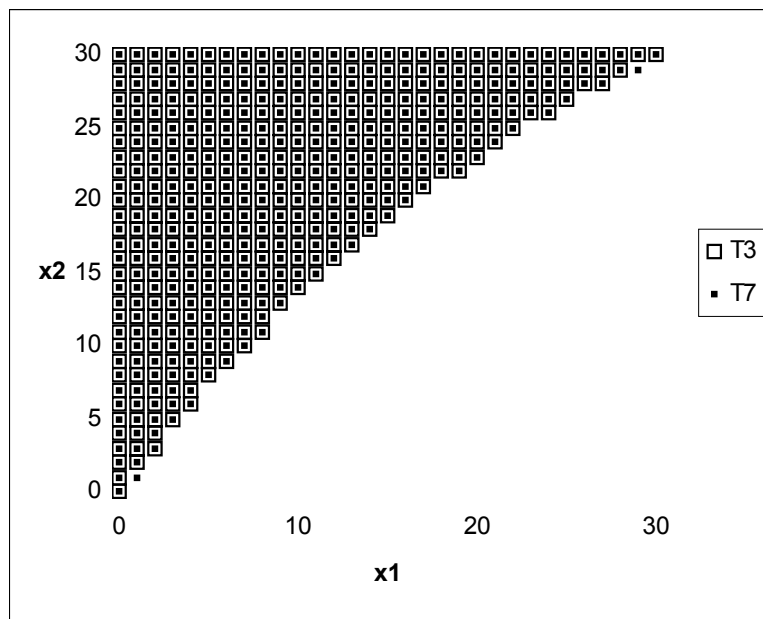


Figura 2.3: Regiones de rechazo para las pruebas  $T_3$  y  $T_7$  para el tamaño de muestra 30, límite de no inferioridad  $d_0 = 0.1$  y nivel de significancia nominal  $\alpha = 0.05$ .

$(x, y) \notin R'_T$  entonces  $\hat{d} \geq d_0$  por lo tanto  $\frac{\hat{d}-d_0}{\hat{\sigma}} < 0$  y  $\frac{\hat{d}-d_0}{\hat{\sigma}} \geq 0$ , esto es una contradicción, por lo cual debe ser cierto que  $R_T - R'_T = \emptyset$ , así se ha probado la igualdad deseada. ■

No obstante, para la estadística de razón de verosimilitudes generalizada los espacios muestral y muestral lícito no coinciden en general, como se ilustra en la figura 2.4

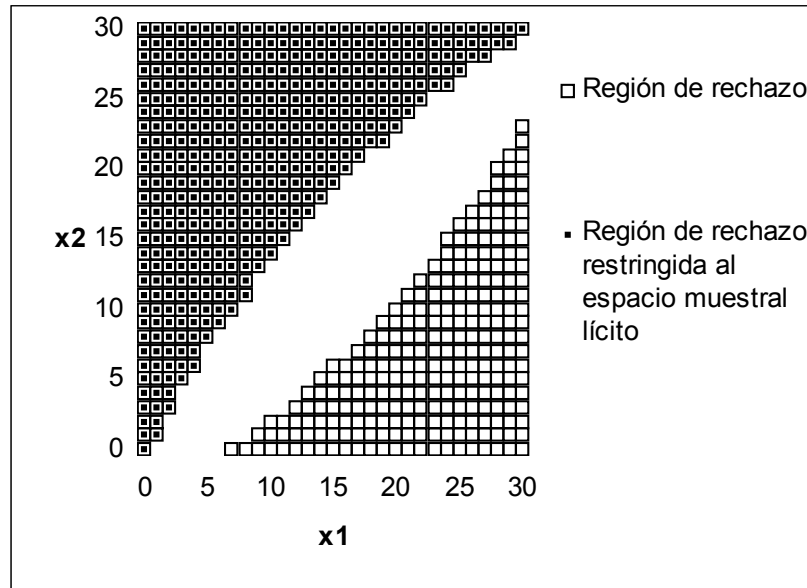


Figura 2.4: Región de rechazo y región de rechazo restringida al espacio muestral lícito para  $T_7$ ,  $n = 30$ ,  $\alpha=0.05$  y  $d_0 = 0.1$ .

## 2.5. Correcciones por continuidad

Las pruebas estadísticas que se van a comparar en este trabajo son asintóticas, sin embargo, la ejecución de tales pruebas es incierta para muestras moderadas (tamaños

de muestra  $\leq 100$ ). Para ilustrarlo, en las figuras 2.5 a 2.7 se presentan los niveles de significancia reales (en el capítulo siguiente se dan los detalles acerca de la forma en que se realizó el cálculo de los niveles de significancia reales) para las siete estadísticas para los límites de no-inferioridad  $d_0 = 0.1, 0.15$  y  $0.2$  para el nivel de significancia nominal  $\alpha = 0.05$  considerando diseños balanceados, es decir, con  $n_1 = n_2 = n$ .

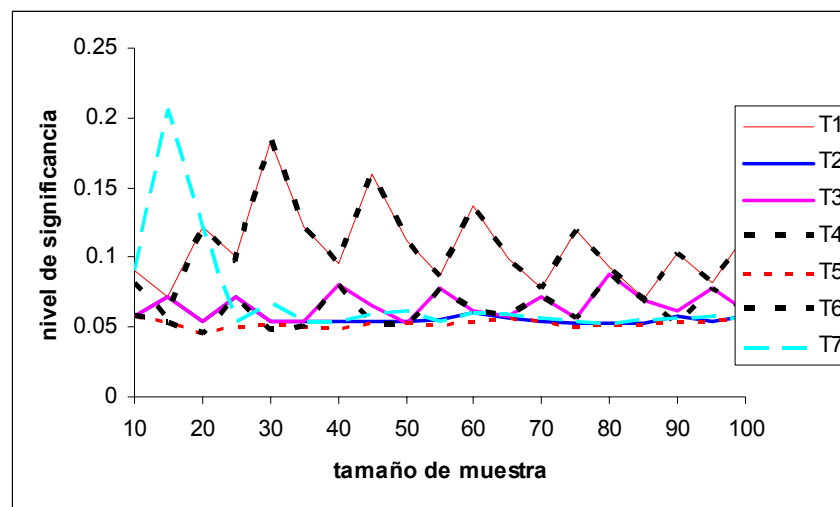


Figura 2.5: Niveles de significancia para las estadísticas  $T_1$  a  $T_7$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad de  $d_0 = 0.1$  para los tamaños de muestra  $n = 10(5)100$ .

En las figuras 2.5 a 2.7 se aprecia que todas las pruebas estadísticas son liberales, es decir, siempre exceden el nivel nominal, tres de ellas ( $T_2$ ,  $T_5$  y  $T_7$ ) son las más próximas al nivel nominal, esto es, son las de mejor comportamiento en este sentido. El uso de una corrección por continuidad puede ayudar a solucionar esta irregularidad pues en muchos casos, procedimientos aproximados que usan corrección por continuidad se transforman en *conservadores*, es decir, mantienen el nivel de significancia real por

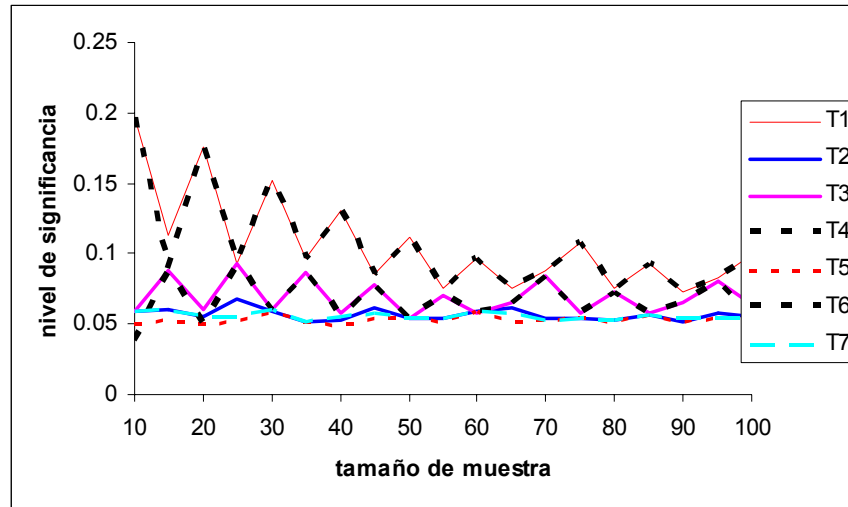


Figura 2.6: Niveles de significancia para las estadísticas  $T_1$  a  $T_7$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad de  $d_0 = 0.15$  para los tamaños de muestra  $n = 10(5)100$ .

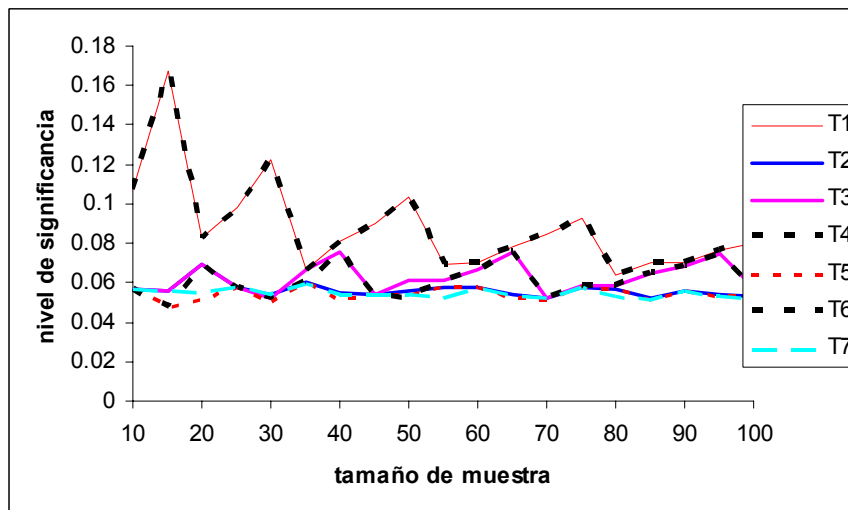


Figura 2.7: Niveles de significancia para las estadísticas  $T_1$  a  $T_7$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad de  $d_0 = 0.2$  para los tamaños de muestra  $n = 10(5)100$ .

debajo del nivel nominal  $\alpha$ . Debido a lo anterior, surge la necesidad de adaptar las estadísticas mediante la introducción de correcciones por continuidad.

Con base en lo anterior, se consideró pertinente también incluir en la comparación de las pruebas consideradas a las mismas pruebas pero modificadas por corrección por continuidad, específicamente, se consideraron las siguientes estadísticas de prueba

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

para  $i = 1, 2, 3, 4, 5, 6$  y  $j = 0, 1, 2, 3, 4, 5$ , donde las correcciones por continuidad son

$$\begin{aligned} C_0 &= 0 & C_1 &= \frac{1}{4 \min(n_1, n_2)} & C_2 &= 2C_1 \\ C_3 &= \frac{1}{2n_1} + \frac{1}{2n_2} & C_4 &= 6C_1 & C_5 &= 8C_1 \end{aligned}$$

para el caso de diseño balanceados, es decir, para  $n_1 = n_2 = n$ , se tiene

$$\begin{aligned} C_0 &= 0 & C_1 &= \frac{1}{4n} & C_2 &= 2C_1 \\ C_3 &= \frac{1}{n} = 4C_1 & C_4 &= 6C_1 & C_5 &= 8C_1 \end{aligned}$$

de donde  $C_i < C_{i+1}$  para  $i = 0, 1, 2, 3, 4$ .

Hauck y Anderson [12] comentan varias correcciones por continuidad, los autores en ese artículo recomiendan  $C_2$  y  $C_3$  las cuales han probado ser adecuadas en la construcción de intervalos de confianza. Además se consideraron también

$$T_{7C_j}(X_1, X_2) = \Lambda(X_1, X_2) + C_j = \frac{\sup L(d; (X_1, X_2))}{\frac{\Theta_0}{\Theta}} + C_j$$

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



El considerar estos seis factores de corrección por continuidad llevó a ampliar la comparación de pruebas a un número total de 42, siendo estas  $T_{iCj}$  para  $i = 1, 2, 3, 4, 5, 6, 7$  y  $j = 0, 1, 2, 3, 4, 5$ .

## Capítulo 3

### Niveles de significancia

En este capítulo se realiza el cálculo de los niveles de significancia de las pruebas consideradas en esta tesis. Se presentan varios resultados los cuales permitirán calcular más eficientemente, desde el punto de vista computacional, los niveles de significancia reales de las pruebas analizadas, esto es de suma importancia en virtud de que el tiempo de cómputo necesario es considerable.

Por ejemplo, supóngase que se desea calcular el nivel de significancia para la prueba  $T_{1C}$  (con o sin corrección por continuidad) en un experimento balanceado (independientemente del tamaño de muestra) si no se conocieran los resultados que se presentarán en este capítulo se tendría que calcular la función potencia en todo el espacio nulo, es decir en  $p_1 - p_2 \geq d_0$ , tomando  $d_0 = .2$  si considerásemos incrementos de .001 debería evaluarse la función de potencia en un total de 321, 201 puntos  $(p_1, p_2)$ . Tomando en consideración los resultados presentados en [25] estos puntos se reducirían

a 801, esto corresponde a solamente el .25 % de los puntos originales  $(p_1, p_2)$ ; mientras que mediante los resultados de este capítulo en este mismo caso se debe evaluar la función de potencia en tan sólo 401 puntos  $(p_1, p_2)$ , lo cual representa .12 % de la cantidad original de puntos, es decir, se obtiene un ahorro de 99.8 % de trabajo de cómputo (puntos donde ha de evaluarse la función potencia).

### 3.1. La función de potencia

Recuérdese que la función de potencia de la prueba está dada por

$$\beta_T(p_1, p_2) = \sum_{(x_1, x_2) \in R_T} L(p_1, p_2; x_1, x_2) \quad (3.1)$$

donde

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

A continuación se presenta un resultado que facilita la comparación de las funciones de potencia de algunas de las estadísticas.

**Proposición 10** *Sea  $A$  una estadística y sean  $T(x, y) = \frac{A(x, y)}{s(x, y)}$ ,  $T'(x, y) = \frac{A(x, y)}{s'(x, y)}$ ,*

*para un problema de prueba de hipótesis con regiones de rechazo respectivas*

$$R_T = \{(x, y) \in \chi : T(x, y) < -c_\alpha\}$$

$$R'_T = \{(x, y) \in \chi : T'(x, y) < -c_\alpha\}$$

*con  $c_\alpha > 0$  fijo, si  $0 < s(x, y) < s'(x, y)$ , entonces  $R'_T \subseteq R_T$ .*

**Demostración.** Sea  $(x, y) \in R'_T(\alpha)$  entonces  $T'(x, y) < -c_\alpha$  por lo tanto  $A(x, y) < 0$ ; por otra parte, como  $s(x, y) < s'(x, y)$  entonces se tiene que  $\frac{1}{s'(x, y)} < \frac{1}{s(x, y)}$  y en consecuencia  $T(x, y) = \frac{A(x, y)}{s(x, y)} < \frac{A(x, y)}{s'(x, y)} = T'(x, y) < -c_\alpha$  de donde  $(x, y) \in R_T$ . ■

**Proposición 11** Para  $p_1, p_2 \in [0, 1]^2$ ;  $\alpha \in (0, 1)$  se tiene que  $\beta_{T_{i+3, C}}(p_1, p_2) \leq \beta_{T_i C}(p_1, p_2)$

**Demostración.** Se probará el resultado para  $i = 1$ .

Si  $p_1, p_2 \in [0, 1]^2$  entonces  $\hat{\sigma}_1 \leq \hat{\sigma}_4$  así por la proposición anterior  $R_{T_4}(\alpha) \subseteq R_{T_1}(\alpha)$  y de la ecuación (3.1) se sigue el resultado.

Los otros casos se prueban de forma similar. ■

**Proposición 12** Si  $\hat{p}_1, \hat{p}_2 \in (0, 1/2)$  o  $\hat{p}_1, \hat{p}_2 \in (1/2, 1)$  entonces  $\beta_{T_{3C}}(p_1, p_2) \leq \beta_{T_1 C}(p_1, p_2)$  para  $p_1, p_2 \in [0, 1]^2$ ;  $\alpha \in (0, 1)$ .

**Demostración.** Primero nótese que  $\hat{p}_i < \frac{1}{2} \Leftrightarrow \tilde{p}_{i,b} < \frac{1}{2}$ . En efecto,

$$\frac{x_i}{n_i} < \frac{1}{2} \Leftrightarrow 2x_i < n_i \Leftrightarrow 2x_i + 2b < n_i + 2b \Leftrightarrow \frac{x_i + b}{n_i + 2b} < \frac{1}{2}.$$

$$\text{a) } \hat{p}_i = \frac{x_i}{n_i} < \frac{1}{2} \Leftrightarrow 2x_i < n_i \Leftrightarrow n_i x_i + 2b x_i < n_i x_i + n_i b \Leftrightarrow \frac{x_i}{n_i} < \frac{x_i + b}{n_i + 2b} \Leftrightarrow \hat{p}_i < \tilde{p}_{i,b}$$

y como la función  $x(1-x)$  es creciente en  $(0, 1/2)$  entonces se tiene que  $\hat{p}_i(1-\hat{p}_i) < \tilde{p}_{i,b}(1-\tilde{p}_{i,b})$  por lo tanto  $\hat{\sigma}_1 \leq \hat{\sigma}_3$

b)  $\hat{p}_i = \frac{x_i}{n_i} > \frac{1}{2} \Leftrightarrow \hat{p}_i > \tilde{p}_{i,b}$  y como la función  $x(1-x)$  es decreciente en  $(1/2, 1)$  entonces se tiene que  $\hat{p}_i(1-\hat{p}_i) < \tilde{p}_{i,b}(1-\tilde{p}_{i,b})$  por lo tanto  $\hat{\sigma}_1 \leq \hat{\sigma}_3$ .

Se ha visto que en cualquier caso se tiene  $\hat{\sigma}_1 \leq \hat{\sigma}_3$ , de donde por la proposición 10,  $R_{T_3} \subseteq R_{T_1}$  en consecuencia se sigue el resultado. ■

## 3.2. Simplificación del cálculo de los niveles de significancia reales

Chan [6] calculó el nivel de significancia para la prueba que aquí se ha llamado  $T_{2C0}$  tomando el supremo no en todo el espacio nulo ( $\Theta_0$ ) sino calculando el máximo únicamente en  $\Theta_0^* = \{(p_1, p_2) : p_1 - p_2 = d_0\}$  el cual es solamente una parte de su frontera, computacionalmente esto representa una inmensa ventaja; para dar una idea de la diferencia en tiempo de cómputo que representa el sólo calcular el nivel de significancia sobre esta parte de la frontera del espacio nulo, basta mencionar que el programa que calcula el nivel de significancia con base en todo el espacio nulo utiliza aproximadamente 450 veces el tiempo de cómputo que se utiliza para calcular el mismo nivel de significancia pero sólo sobre la parte referida de la frontera del espacio nulo.

En la presente sección y a menos que se especifique otra cosa,  $T$  es una estadística de prueba para el problema de contrastar las hipótesis

$$[H_0 : p_2 \leq g(p_1)] \quad vs \quad [H_a : p_2 > g(p_1)] \quad (3.2)$$

con  $g : [0, 1] \rightarrow [0, 1]$  función continua y no decreciente con  $g \neq 1$ . Nótese que esto incluye funciones como  $g(p_1) = p_1 - \theta_0$ ,  $g(p_1) = p_1/\theta_0$ ,  $g(p_1) = \frac{p_1}{\theta_0(1-p_1)+p_1}$ , estas funciones corresponden a las medidas de disimilaridad diferencia de proporciones, razón de proporciones y razón de momios, respectivamente.

**Definición 13** *Sea  $T$  una prueba estadística para el problema de prueba de hipótesis*

en (3.2), con región crítica  $R_T$ . Se dice que  $T$  (o que  $R_T$ ) satisface la **condición de convexidad de Barnard (C)** si se cumplen las dos condiciones siguientes:

$$a) (x, y) \in R \Rightarrow (x - 1, y) \in R \text{ para toda } 1 \leq x \leq n_1, 0 \leq y \leq n_2 \text{ y}$$

$$b) (x, y) \in R \Rightarrow (x, y + 1) \in R \text{ para toda } 0 \leq x \leq n_1, 0 \leq y \leq n_2 - 1$$

Röhmel y Mansmann en [25] demostraron el siguiente resultado para pruebas exactas.

**Teorema 14** Sea  $T$  una estadística de prueba 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 de rechazo dada por  $R = \{(x, y) : T(x, y) \leq T(x_0, y_0)\}$  y sean

$$\Psi_0 = \{\theta = (p_1, p_2) : p_2 \leq g(p_1)\}$$

$a \leq p_1 \leq 1$ ,  $0 \leq a$  y  $g(a) = 0$  si  $a > 0$  con  $g$  creciente y no intersecta  $p_2 = p_1$ . Si la región  $R$  satisface la condición de convexidad de Barnard, entonces el supremo

$$\sup_{(p_1, p_2) \in \Psi_0} \beta_T(p_1, p_2)$$

es un máximo y se alcanza en un punto frontera sobre la curva  $p_2 = g(p_1)$ .

Es decir, si  $\Psi_0^* = \{(p_1, p_2) : p_2 = g(p_1)\}$ , entonces

$$\text{nivel de significancia} = \sup_{\theta \in \Psi_0} \beta_T(p_1, p_2) = \max_{\theta \in \Psi_0^*} \beta_T(p_1, p_2)$$

En virtud de que la demostración del teorema anterior dada por Röhmel y Mansmann adolece de algunas formalidades y tiene pequeños errores, sería conveniente

presentar en este trabajo una demostración detallada del resultado, no obstante se prefiere demostrar dos resultados que generalizan el teorema anterior, uno de ellos lo generaliza a pruebas tanto asintóticas como exactas y el otro lo generaliza a una clase de regiones de rechazo más amplia que la clase de las regiones críticas que cumplen la condición (C).

La condición (C) es bastante restrictiva por lo que hay una gran cantidad de ejemplos de regiones críticas que no la cumplen.

En la figura 2.1 se puede apreciar que para la estadística  $T_1$ , la región de rechazo original no cumple la condición (C) pero que la región de rechazo redefinida sí la cumple. Este no es siempre el caso como se ve en el siguiente

**Ejemplo 15** *En la figura 3.1 puede verse que  $(3, 3) \in R_{T_1} \cap R_{T_4}$  pero  $(2, 3) \notin R_{T_1} \cup R_{T_4}$  por lo tanto ni  $R_{T_1}$  ni  $R_{T_4}$  cumplen la propiedad (a) de la definición de condición de convexidad de Barnard, es decir, ninguna de esas pruebas (equivalentemente regiones de rechazo) cumple la condición (C). Puede verse también que  $R_{T_1}$  y  $R_{T_4}$  tampoco cumplen la propiedad (b) de la definición de condición (C).*

Con el ejemplo anterior queda ilustrado que debe tenerse cuidado con la redefinición que se hizo de las estadísticas  $T_1$  y  $T_4$  ya que genera ciertas dificultades para tamaños de muestra pequeños pues incluso para diseños balanceados a veces no se cumple la condición (C). En realidad esto no representa una dificultad para el caso de pruebas asintóticas ya que como se mencionó, este problema se observa para tamaños de muestra pequeños.

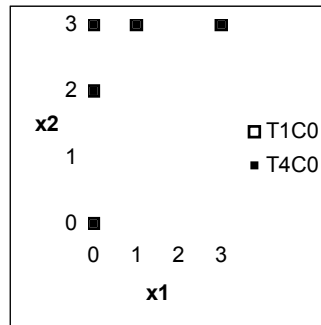


Figura 3.1: Regiones de rechazo correspondientes a las pruebas  $T_1$  y  $T_4$  para  $n_1 = n_2 = 3$ ,  $\alpha = 0.05$  y  $d_0 = 0.1$

**Ejemplo 16** *En relación a la figura 3.2, nótese que  $(2, 0) \in R_{T_1} \cap R_{T_4}$  pero  $(2, 1) \notin R_{T_1} \cup R_{T_4}$  por tal motivo ni  $R_{T_1}$  ni  $R_{T_4}$  cumplen la propiedad (b) de la definición de condición de convexidad de Barnard, es decir, ninguna de esas pruebas (equivalentemente regiones de rechazo) cumple la condición (C).*

Como es natural esperar, se pueden construir innumerables ejemplo de este tipo.

Cuando se desea calcular los niveles de significancia reales para pruebas con regiones críticas como las del ejemplo anterior se presenta un problema ya que al no cumplir la condición (C) no se tiene un resultado que permita simplificar su cálculo, esto es, para calcular el nivel de significancia debería calcularse el máximo en todo el espacio nulo con el consecuente elevado consumo de tiempo de cómputo.

No obstante, en este trabajo se generaliza el multicitado teorema de Röhmel y Mansmann a una clase de regiones críticas en la cual quedan incluidas regiones como las del ejemplo anterior.

Se usará la convención usual  $\binom{n}{x} = 0$  si  $x < 0$  ó si  $x > n$ .



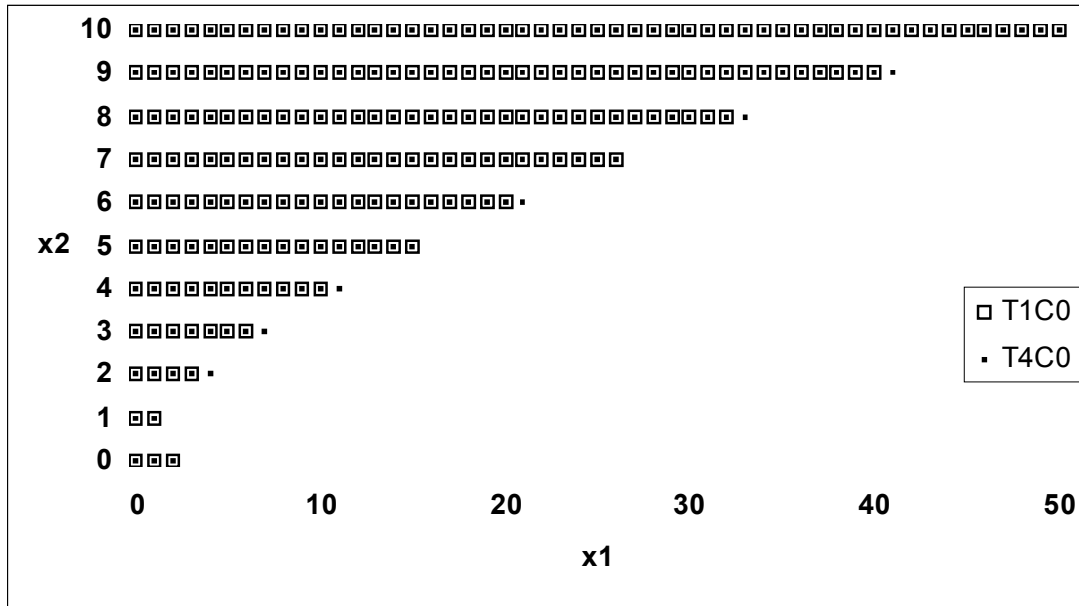


Figura 3.2: Regiones de rechazo para  $n_1 = 50$ ,  $n_2 = 10$ ,  $\alpha = 0.05$  y  $d_0 = 0.1$  para las pruebas  $T_1$  y  $T_4$ .

La evaluación de  $0^0$  representa un problema debido a que diferentes enfoques llegan a diferentes resultados. Básicamente hay dos acercamientos, el del análisis matemático, en el cual se considera  $0^0$  como indefinido y el de matemáticas discretas, naturalmente aquí se adopta la convención usual de este último enfoque según el cual se define  $0^0 = 1$ .

De aquí en adelante, se usará la notación descrita a continuación.

Para  $x \in \{0, \dots, n_1\}$  sea  $R_x = \{y : (x, y) \in R_T\}$ , si  $R_x \neq \emptyset$ , sea  $a(x) = \min R_x$ ; similarmente, para  $y \in \{0, \dots, n_2\}$  sea  $S_y = \{x : (x, y) \in R_T\}$ , si  $S_y \neq \emptyset$ , sea  $b(y) = \max S_y$ .

**Definición 17** Las siguientes condiciones (SC1) y (SC2) son llamadas condiciones de semiconvexidad:

(**SC1**) (SC1a)  $(x, y) \in R \Rightarrow (x-1, y) \in R$  para toda  $1 \leq x \leq n_1, 0 \leq y \leq n_2$ .

(SC1b)  $S_y \neq \emptyset \Rightarrow S_{y'} \neq \emptyset$  para toda  $y \leq y' \leq n_2$ .

(**SC2**) (SC2a)  $(x, y) \in R \Rightarrow (x, y+1) \in R$  para toda  $0 \leq x \leq n_1, 0 \leq y \leq n_2 - 1$ .

(SC2b)  $R_x \neq \emptyset \Rightarrow R_{x'} \neq \emptyset$  para toda  $0 \leq x' \leq x$ .

El siguiente resultado caracteriza la condición de convexidad de Barnard en términos de las condiciones de semiconvexidad.

**Proposición 18**  $R_T$  cumple la condición (C) si y sólo si  $R_T$  cumple las condiciones (SC1) y (SC2).

**Demostración.** ( $\Rightarrow$ ) Supóngase que  $S_y \neq \emptyset$  y sea  $y'$  con  $y \leq y' \leq n_2$ , como por hipótesis se cumple la condición (C), se tiene que  $S_y \subseteq S_{y'}$  por lo tanto  $S_{y'} \neq \emptyset$  y en consecuencia se sigue la propiedad (SC1b).

Similarmente, supóngase que  $R_x \neq \emptyset$  y sea  $x'$  con  $0 \leq x' \leq x$ , también por la condición (C), se tiene que  $R_x \subseteq R_{x'}$  así  $R_{x'} \neq \emptyset$  de donde se sigue la propiedad (SC2b).

( $\Leftarrow$ ) Inmediata. ■

**Proposición 19** Si  $R_T \neq \emptyset$  y cumple alguna de las dos condiciones de semiconvexidad (SC1) o (SC2), entonces  $\text{grad}(\beta_T(p_1, p_2)) \neq \mathbf{0} \forall (p_1, p_2) \in \Theta_0^o$ .

**Demostración.** Supóngase que se cumple la condición de semiconvexidad (SC1), entonces por la propiedad (SC1b), existe  $n_2^* = \min \{y : S_y \neq \emptyset\}$ , por lo tanto por la propiedad (SC1a)

$$\beta_T(p_1, p_2) = \sum_{(x,y) \in R_T} L(p_1, p_2; x, y) = \sum_{y=n_2^*}^{n_2} \sum_{x=0}^{b(y)} f(x; n_1, p_1) f(y; n_2, p_2)$$

y por la observación 8 (2)

$$\frac{\partial \beta_T(p_1, p_2)}{\partial p_1} = \sum_{y=n_2^*}^{n_2} \sum_{x=0}^{b(y)} f(y; n_2, p_2) n_1 [f(x-1; n_1-1, p_1) - f(x; n_1-1, p_1)]$$

así

$$\frac{\partial \beta_T(p_1, p_2)}{\partial p_1} = -n_1 \sum_{y=n_2^*}^{n_2} f(y; n_2, p_2) f(b(y); n_1-1, p_1) < 0$$

análogamente, supóngase que se cumple la condición de semiconvexidad (SC2), entonces por la propiedad (SC2b) existe  $n_1^* = \max \{x : A_x \neq \emptyset\}$ , por lo tanto por la propiedad (SC2a)

$$\beta_T(p_1, p_2) = \sum_{(x,y) \in R_T} L(p_1, p_2; x, y) = \sum_{x=0}^{n_1^*} \sum_{y=a(x)}^{n_2} f(x; n_1, p_1) f(y; n_2, p_2)$$

y por la observación 8 (2)

$$\frac{\partial \beta_T(p_1, p_2)}{\partial p_2} = \sum_{x=0}^{n_1^*} \sum_{y=a(x)}^{n_2} f(x; n_1, p_1) n_2 [f(y-1; n_2-1, p_2) - f(y; n_2-1, p_2)]$$

así

$$\frac{\partial \beta_T(p_1, p_2)}{\partial p_2} = n_2 \sum_{x=0}^{n_1^*} f(x; n_1, p_1) f(a(x)-1; n_2-1, p_2) > 0$$

por lo tanto se tiene que  $grad(\beta_T(p_1, p_2)) \neq \mathbf{0}$ . ■

En lo que sigue se usará la notación descrita a continuación

$$\mathcal{F}_1 = \{(p_1, 0) : a \leq p_1 \leq 1, 0 \leq a \leq 1\}$$

$$\mathcal{F}_2 = \{(1, p_2) : 0 \leq p_2 \leq b, 0 \leq b \leq 1\}$$

$$\mathcal{F}_3 = \{(0, p_2) : 0 \leq p_2 \leq c, 0 \leq c \leq 1\}$$

$$\mathcal{F}_4 = \{(p_1, 1) : d \leq p_1 \leq 1, 0 \leq d \leq 1\}$$

$$\Theta_0^* = \{(p_1, p_2) \in \Theta : p_2 = g(p_1)\}$$

en consecuencia la frontera de  $\Theta_0 = \{(p_1, p_2) \in \Theta : p_2 \leq g(p_1)\}$  se puede representar mediante

$$\partial\Theta_0 = \Theta_0^* \cup \left( \bigcup_{i=1}^4 \mathcal{F}_i \right)$$

Los tres resultados siguientes generalizan el teorema de Röhmel y Mansmann a pruebas tanto exactas como asintóticas, así como a regiones críticas más generales.

**Teorema 20** Sean  $\Psi_0$  cualquier subconjunto cerrado de  $\Theta$ . Si  $R_T \neq \emptyset$  y satisface alguna de las condiciones de semiconvexidad (SC1) o (SC2), entonces el supremo

$$\sup_{(p_1, p_2) \in \Psi_0} \beta_T(p_1, p_2)$$

es un máximo y se alcanza en un punto frontera, es decir,

$$\text{nivel de significancia} = \sup_{\theta \in \Psi_0} \beta_T(p_1, p_2) = \max_{\theta \in \partial\Psi_0} \beta_T(p_1, p_2)$$

**Demostración.** Como  $\beta_T(p_1, p_2)$  es una función continua en el conjunto compacto  $\Psi_0$ , por el criterio de la primera derivada para funciones de varias variables y por la proposición anterior, ningún punto en el interior de  $\Psi_0$  ( $\Psi_0^\circ$ ) es un máximo, por lo tanto el máximo debe alcanzarse en la frontera de  $\Psi_0$  ( $\partial\Psi_0$ ). ■

El siguiente resultado generaliza el teorema de Röhmel y Mansmann a pruebas tanto exactas como asintóticas.

**Teorema 21** *Si  $R_T \neq \emptyset$ , entonces se cumplen las afirmaciones siguientes*

1. *Si  $R_T$  satisface la condición (SC1), entonces el supremo  $\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2)$  es un máximo y se alcanza en un punto frontera sobre  $\Theta_1^* = \partial\Theta_0 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_4 = \Theta_0^* \cup \mathcal{F}_3$ , es decir,  $\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2) = \sup_{(p_1, p_2) \in \Theta_1^*} \beta_T(p_1, p_2)$ .*
2. *Si  $R_T$  satisface la condición (SC2), entonces el supremo  $\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2)$  es un máximo y se alcanza en un punto frontera sobre  $\Theta_2^* = \partial\Theta_0 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 = \Theta_0^* \cup \mathcal{F}_4$ , es decir,  $\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2) = \sup_{(p_1, p_2) \in \Theta_2^*} \beta_T(p_1, p_2)$ .*

**Demostración.** 1) Por el teorema anterior, el supremo es un máximo y se alcanza en  $\partial\Theta_0$ , falta demostrar que se alcanza en  $\Theta_1^*$ .

Primero se analiza el comportamiento en  $\mathcal{F}_1$ .

$$\begin{aligned} \beta_T(p_1, p_2) &= \sum_{T(x, y) \leq t_0} f(x, n_1, p_1) f(y, n_2, 0) = \sum_{T(x, 0) \leq t_0} f(x, n_1, p_1) \\ &= \sum_{x \in \mathcal{S}_0} f(x, n_1, p_1) \end{aligned}$$

Si  $S_0 = \emptyset$ , entonces  $\beta_T(p_1, p_2) = 0$ .

Si  $S_0 \neq \emptyset$ , entonces por la condición de (SC1)  $0 \in S_0$ .

Para determinar donde se alcanza el máximo basta evaluar  $\beta_T(a, 0)$  y  $\beta_T(1, 0)$ .

Si  $n_1 \in S_0$ , entonces por la propiedad (SC1a)  $S_0 = \{x : 0 \leq x \leq n_1\}$  así  $\beta_T(p_1, 0) = \sum_{x=0}^{n_1} f(x, n_1, p_1) = 1$  por lo tanto  $\beta_T(a, 0) = \beta_T(1, 0) = 1$ .

Si  $n_1 \notin S_0$ , entonces  $\beta_T(p_1, 0) = \sum_{x=0}^{n'_1} f(x, n_1, p_1)$  con  $n'_1 < n_1$  por lo tanto se tiene que  $\beta_T(1, 0) = \sum_{x=0}^{n'_1} f(x, n_1, 1) = \sum_{x=0}^{n'_1} \binom{n_1}{x} 1^x 0^{n_1-x}$  como  $n'_1 < n_1$  entonces  $\beta_T(1, 0) = 0$  y  $\beta_T(a, 0) > 0$ .

Así en cualquiera de los dos casos anteriores, el máximo se alcanza en  $(a, 0) \in \mathcal{F}_3 \cup \Theta_0^*$ .

A continuación se analiza el comportamiento en  $\mathcal{F}_4$ , si  $\mathcal{F}_4 \neq \emptyset$ .

$$\begin{aligned} \beta_T(p_1, p_2) &= \sum_{T(x,y) \leq t_0} f(x, n_1, p_1) f(y, n_2, 1) = \sum_{T(x, n_2) \leq t_0} f(x, n_1, p_1) \\ &= \sum_{x \in S_{n_2}} f(x, n_1, p_1) \end{aligned}$$

Si  $S_{n_2} = \emptyset$ , entonces  $R_T = \emptyset$ .

Si  $S_{n_2} \neq \emptyset$ , entonces por la condición de (SC1)  $0 \in S_{n_2}$ .

Para determinar donde se alcanza el máximo basta evaluar  $\beta_T(d, 1)$  y  $\beta_T(1, 1)$ .

Si  $n_1 \in S_{n_2}$ , entonces por la propiedad (SC1a)  $S_{n_2} = \{x : 0 \leq x \leq n_1\}$  así  $\beta_T(p_1, 1) = \sum_{x=0}^{n_1} f(x, n_1, 1) = 1$  por lo tanto  $\beta_T(d, 1) = \beta_T(1, 1) = 1$ .

Si  $n_1 \notin S_{n_2}$ , entonces  $\beta_T(p_1, 1) = \sum_{x=0}^{n'_1} f(x, n_1, p_1)$  con  $n'_1 < n_1$  por lo tanto se tiene

que  $\beta_T(1, 1) = \sum_{x=0}^{n'_1} f(x, n_1, 1) = \sum_{x=0}^{n'_1} \binom{n_1}{x} 1^x 0^{n_1-x}$  como  $n'_1 < n_1$  entonces  $\beta_T(1, 1) = 0$  y  $\beta_T(d, 1) > 0$ .

Así en cualquiera de los dos casos anteriores, el máximo se alcanza en  $(d, 1) \in \Theta_0^*$ .

Se ha probado entonces que el máximo buscado se alcanza en  $\Theta_1^*$ .

2) Se prueba de manera similar. ■

Desde esta óptica, el teorema de Röhmel y Mansmann generalizado a pruebas exactas y asintóticas resulta ser una consecuencia directa del teorema anterior como se ve enseguida.

**Corolario 22** Si  $R_T \neq \emptyset$  y satisface la condición (C), entonces el supremo

$$\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2)$$

es un máximo y se alcanza en un punto frontera sobre  $\Theta_0^*$  es decir,

$$\sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2) = \max_{(p_1, p_2) \in \Theta_0^*} \beta_T(p_1, p_2)$$

**Proposición 23** Sea  $T$  una estadística de prueba con distribución asintótica  $A_T$ .

Sean  $R_T$  y  $R_T^A$  las regiones críticas para las pruebas exacta y asintótica, respectivamente.

Si  $R_T^A \neq \emptyset$  entonces  $R_T^A$  es la región crítica para una prueba exacta.

**Demostración.** Sean  $R_T = R_T(x_0, y_0) = \{(x, y) : T(x, y) \leq T(x_0, y_0)\}$  y  $R_T^A = \{(x, y) : T(x, y) \leq c_\alpha\}$ .

Si  $R_T^A \neq \emptyset$  entonces existe  $(x_0^*, y_0^*) \in \chi$  tal que

$$T(x_0^*, y_0^*) = \max \{T(x, y) : (x, y) \in R_T^A\} = \max \{T(x, y) : T(x, y) \leq c_\alpha\}$$

por lo tanto

$$R_T^A = \{(x, y) : T(x, y) \leq c_\alpha\} = \{(x, y) : T(x, y) \leq T(x_0^*, y_0^*)\} = R_T(x_0^*, y_0^*) \text{ así } R_T^A$$

es la región crítica correspondiente a una prueba exacta. ■

En la proposición anterior se demostró que la región crítica de toda prueba asintótica puede representarse como la región crítica de una prueba exacta, en otras palabras, la clase de las pruebas asintóticas está contenida en la clase de las pruebas exactas. Nótese que dada una prueba exacta su correspondiente prueba asintótica no necesariamente tiene el mismo tamaño.

**Teorema 24** *Sea  $T$  una estadística de prueba con distribución asintótica  $A_T$ . Si las regiones críticas para las pruebas exacta y asintótica son respectivamente  $R_T$  y  $R_T^A$ . Si  $R_T(x_0, y_0)$  cumple la condición de convexidad de Barnard  $\forall (x_0, y_0) \in \chi$  y si  $R_T^A \neq \emptyset$ , entonces  $R_T^A$  también cumple la condición de convexidad de Barnard.*

**Demostración.** Inmediata de la proposición anterior. ■

Röhmel y Mannsman en [26] demostraron que la región de rechazo para la prueba estadística exacta correspondiente a  $T_2$  cumple la condición de convexidad de Barnard, es decir,

**Teorema 25** *La región  $R_{T_2}$  cumple la condición de convexidad de Barnard.*

**Proposición 26** *Se tiene que  $\hat{\sigma}_i(n_1 - x_1, n_2 - x_2) = \hat{\sigma}_i(x_1, x_2)$  para  $i = 1, 3, 4, 6$ .*



**Demostración.** Se probará el resultado para  $i = 3$ .

$$\begin{aligned}
& \widehat{\sigma}_3(n_1 - x_1, n_2 - x_2) \\
&= \sqrt{\frac{1}{n_1} \frac{n_1 - x_1 + b}{n_1 + 2b} \left(1 - \frac{n_1 - x_1 + b}{n_1 + 2b}\right) + \frac{1}{n_2} \frac{n_2 - x_2 + b}{n_2 + 2b} \left(1 - \frac{n_2 - x_2 + b}{n_2 + 2b}\right)} \\
&= \sqrt{\frac{1}{n_1} \frac{x_1 + b}{n_1 + 2b} \left(1 - \frac{x_1 + b}{n_1 + 2b}\right) + \frac{1}{n_2} \frac{x_2 + b}{n_2 + 2b} \left(1 - \frac{x_2 + b}{n_2 + 2b}\right)} \\
&= \widehat{\sigma}_3(x_1, x_2)
\end{aligned}$$

El resultado para  $i = 1$  se obtiene tomando  $b=0$ .

Los resultados para  $i = 4, 6$  se obtienen en forma análoga. ■

Desde un punto de vista computacional es mucho más conveniente poder verificar que una estadística dada cumple la condición de convexidad de Barnard y como se mencionó antes, entonces el cálculo del nivel de significancia se simplifica mucho pues se efectúa sobre una parte de la frontera del espacio de parámetros y no sobre todo el espacio de parámetros.

Sin embargo, como se ha observado, el verificar que una región cumple la condición de convexidad de Barnard generalmente es bastante complicado debido a dificultades numéricas intrínsecas y es una propiedad que hay que verificar cuidadosamente ya que por ejemplo una prueba tan natural como  $T_1$  no la cumple.

Típicamente se verificará numéricamente que se cumple ya sea la condición (C) o alguna de las condiciones (SC1) o (SC2). Si se cumple la condición (C) entonces la maximización numérica puede realizarse sobre la parte de la frontera dada por la curva solamente; si se cumple alguna de las condiciones (SC1) o (SC2), entonces la

maximización numérica puede realizarse sobre la frontera solamente, en caso contrario, es decir, si no se cumple ninguna de las condiciones (SC1) y (SC2) ( y en consecuencia no se cumple la condición (C)), entonces la maximización deberá realizarse sobre todo el espacio de parámetros.

**Definición 27** *Se dice que una región crítica  $R$  satisface la **propiedad de simetría en la misma cola (SMC)** si  $(x, y) \in R \Rightarrow (n - y, n - x) \in R$*

**Proposición 28** *Si  $n = n_1 = n_2$ , entonces  $R_{iC}$  cumple la propiedad de simetría en la misma cola para  $i = 1, 3, 4, 6$ .*

**Demostración.** *Inmediata de que  $T_{iC}(n - x_2, n - x_1) = T_{iC}(x_1, x_2)$ . ■*

**Proposición 29** *Sean  $n_1 = n_2 = n$  y  $R(\alpha)$  una región crítica para el problema de prueba de hipótesis*

$$[H_0 : d \geq d_0] \text{ vs } [H_a : d < d_0]$$

*para el nivel de significancia nominal  $\alpha$ , si  $R(\alpha)$  cumple la condición (C) y la condición (SMC), entonces el nivel de significancia está dado por*

$$\alpha^* = \max_{\substack{p_2 = p_1 - d_0 \\ p_1 \in [d_0, \frac{1+d_0}{2}]}} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} \binom{n_1}{x_1} p_1^{x_1} (1 - p_1)^{n_1 - x_1} \binom{n_2}{x_2} p_2^{x_2} (1 - p_2)^{n_2 - x_2} I_{[(x_1, x_2) \in R(\alpha)]} \quad (3.3)$$

**Demostración.** Sea

$$h(p) = \sum_{x_1=0}^n \sum_{x_2=0}^n f(x_1, n, p) f(x_2, n, p - d_0) I_{[(x_1, x_2) \in R(\alpha)]}$$

$x'_1 = n - x_2$ ,  $x'_2 = n - x_1$ , por la propiedad de simetría en la misma cola y como  $f(n - x, n, p) = f(x, n, 1 - p)$  se tiene que

$$\begin{aligned} h(1 - p + d_0) &= \sum_{x_1=0}^n \sum_{x_2=0}^n f(x_1, n, 1 - p + d_0) f(x_2, n, 1 - p) I_{[(n-x_2, n-x_1) \in R(\alpha)]} \\ &= \sum_{x'_1=0}^n \sum_{x'_2=0}^n f(x'_1, n, p) f(x'_2, n, p - d_0) I_{[(x'_1, x'_2) \in R(\alpha)]} \\ &= h(p) \end{aligned}$$

así

$$h(1 - p + d_0) = h(p) \quad \forall p \in [d_0, 1] \quad (3.4)$$

$$\begin{aligned} \alpha^* &= \sup_{(p_1, p_2) \in \Theta_0} \beta_T(p_1, p_2) = \sup_{(p_1, p_2) \in \Theta_0} \sum_{(x_1, x_2) \in R(\alpha)} L(p_1, p_2; x_1, x_2) \\ &= \sup_{p_2 \leq p_1 - d_0} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} f(x_1, n, p_1) f(x_2, n, p_2) I_{[(x_1, x_2) \in R(\alpha)]} \end{aligned}$$

de donde por el corolario 22

$$\begin{aligned} \alpha^* &= \max_{p_1 \in [d_0, 1]} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} f(x_1, n, p_1) f(x_2, n, p_1 - d_0) I_{[(x_1, x_2) \in R(\alpha)]} \\ &= \max_{p_1 \in [d_0, 1]} h(p_1) I_{[(x_1, x_2) \in R(\alpha)]} \end{aligned}$$

además de que  $1 - p_1 + d_0 \in [d_0, \frac{1+d_0}{2}]$  si y sólo si  $p_1 \in [\frac{1+d_0}{2}, 1]$  y de la ecuación (3.4)

$$\begin{aligned} \alpha^* &= \max \left\{ \max_{p_1 \in [d_0, \frac{1+d_0}{2}]} h(p_1) I_{[(x_1, x_2) \in R(\alpha)]}, \max_{p_1 \in [\frac{1+d_0}{2}, 1]} h(p_1) I_{[(x_1, x_2) \in R(\alpha)]} \right\} \\ &= \max \left\{ \max_{p_1 \in [d_0, \frac{1+d_0}{2}]} h(p_1) I_{[(x_1, x_2) \in R(\alpha)]}, \max_{p_1 \in [\frac{1+d_0}{2}, 1]} h(1 - p_1 + d_0) I_{[(x_1, x_2) \in R(\alpha)]} \right\} \\ &= \max_{p_1 \in [d_0, \frac{1+d_0}{2}]} h(p_1) \end{aligned}$$

como se quería demostrar. ■

### 3.3. Procedimiento utilizado para calcular los niveles de significancia reales

Se usó el resultado previo para simplificar el cálculo de los niveles de significancia reales, para lo cual se dividió al intervalo  $[0,1]$  en  $k$  subintervalos de igual longitud ( $\Delta$ ) y se evaluó en cada uno de los valores así determinados, es decir, en  $p_1^{(i)} = d_0 + i\Delta = d_0 + i\frac{1+d_0}{2}$  para  $i = 0, \dots, k$ , en otras palabras, se evaluó la función de potencia en los puntos  $(d_0 + i\Delta, i\Delta)$  para  $i = 0, \dots, k$ .

El procedimiento utilizado para calcular el nivel de significancia real es el siguiente:

1. Verificar que se cumple la condición de convexidad de Barnard.
2. Verificar que se cumple la condición de simetría en la misma cola.
3. Aplicar la fórmula (3.3) utilizando los puntos descritos en el párrafo anterior con un incremento  $\Delta = .001$ .

Las verificaciones de las condiciones de convexidad de Barnard y de simetría en la misma cola se realizaron en forma analítica en los siguientes casos:

La condición de convexidad de Barnard para  $T_2$  (teorema 25).

La condición de simetría en la misma cola  $T_{iCj}$  para  $i = 1, 3, 4, 6$  y  $j = 0, 1, 2, 3, 4, 5$  (proposición 28).

En los casos restantes, la verificación se efectuó en forma numérica mediante programas de cómputo escritos en el lenguaje de programación S-PLUS.

### 3.4. Propiedades asintóticas de los niveles de significancia

A continuación se proporciona un resultado acerca del comportamiento asintótico del nivel de significancia.

**Proposición 30** Sea  $T = \frac{\hat{d}-d_0}{\hat{\sigma}}$  tiene distribución asintótica normal estándar

a) Entonces  $\lim_{n_1, n_2 \rightarrow \infty} \alpha^* \leq \alpha$

b) Si  $R$  cumple la condición de convexidad de Barnard (C) entonces  $\lim_{n_1, n_2 \rightarrow \infty} \alpha^* =$

$\alpha$ .

**Demostración.** a) Sean  $A = \left\{ (x, y) : \frac{\hat{d}}{\hat{\sigma}} < -z_\alpha + \frac{d_0}{\hat{\sigma}} \right\}$  y  $B = \left\{ (x, y) : \frac{\hat{d}}{\hat{\sigma}} < -z_\alpha + \frac{d}{\hat{\sigma}} \right\}$ , entonces bajo la hipótesis nula  $H_0 : d \geq d_0$  se tiene que  $A \subseteq B$  por lo tanto

$$\begin{aligned} P\{\text{Rechazar } H_0 \mid H_0\} &= P(T < -z_\alpha \mid H_0) = P\left\{ \frac{\hat{d}-d_0}{\hat{\sigma}} < -z_\alpha \mid d \geq d_0 \right\} \\ &= P\left\{ \frac{\hat{d}}{\hat{\sigma}} < -z_\alpha + \frac{d_0}{\hat{\sigma}} \mid d \geq d_0 \right\} = P(A) \leq P(B) \\ &= P\left\{ \frac{\hat{d}}{\hat{\sigma}} < -z_\alpha + \frac{d}{\hat{\sigma}} \right\} = P\left\{ \frac{\hat{d}-d}{\hat{\sigma}} < -z_\alpha \right\} \xrightarrow{n_1, n_2 \rightarrow \infty} \alpha \end{aligned}$$

en consecuencia

$$\lim_{n_1, n_2 \rightarrow \infty} \alpha^* = \lim_{n_1, n_2 \rightarrow \infty} \sup_{d \geq d_0} P\{\text{Rechazar } H_0 \mid H_0\} \leq \alpha$$

b) Del corolario 22 se tiene que

$$\begin{aligned}\alpha^* &= \sup_{d \geq d_0} P \{T < -z_\alpha \mid d \geq d_0\} = \max_{d=d_0} P \{T < -z_\alpha \mid d = d_0\} \\ &= \max_{d=d_0} P \left\{ \frac{\widehat{d} - d_0}{\widehat{\sigma}} < -z_\alpha \mid d = d_0 \right\} = \max_{d=d_0} P \left\{ \frac{\widehat{d} - d}{\widehat{\sigma}} < -z_\alpha \right\} \xrightarrow[n_1, n_2 \rightarrow \infty]{} \alpha\end{aligned}$$

■

A continuación se prueban resultados correspondientes a los anteriores, para la prueba de razón de verosimilitudes generalizada.

**Proposición 31** Para  $T_7$

a) Se tiene que  $\lim_{n_1, n_2 \rightarrow \infty} \alpha^* \leq \alpha$

b) Si  $R$  cumple la condición de convexidad de Barnard (C) entonces  $\lim_{n_1, n_2 \rightarrow \infty} \alpha^* = \alpha$

**Demostración.** a) Bajo la hipótesis nula  $H_0 : d \geq d_0$  se tiene

$$\begin{aligned}P \{ \text{Rechazar } H_0 \mid H_0 \} &= P \{ -2 \ln T_7 > \chi_{1-2\alpha}^2(1) \mid H_0 \} \\ &= P \{ -2 \ln T_7 > \chi_{1-2\alpha}^2(1) \mid d \geq d_0 \} \\ &\leq P \{ -2 \ln T_7 > \chi_{1-\alpha}^2(1) \} \xrightarrow[n_1, n_2 \rightarrow \infty]{} \alpha\end{aligned}$$

b) Del corolario 22 se tiene que

$$\begin{aligned}\alpha^* &= \sup_{d \geq d_0} P \{ -2 \ln T_7 > \chi_{1-2\alpha}^2(1) \mid d \geq d_0 \} = \max_{d=d_0} P \{ -2 \ln T_7 > \chi_{1-2\alpha}^2(1) \mid d = d_0 \} \\ &= \max_{d=d_0} P \{ -2 \ln T_7 > \chi_{1-2\alpha}^2(1) \mid \} \xrightarrow[n_1, n_2 \rightarrow \infty]{} \alpha\end{aligned}$$

■

### 3.5. Clasificación de las pruebas con base en su nivel de significancia

En la figura 3.3 se exhibe el comportamiento errático del nivel de significancia de la prueba de Blackwelder ( $T_{1C0}$ ) así como de la prueba de razón de verosimilitudes ( $T_{7C0}$ ) para el límite de no inferioridad  $d_0 = 0.1$  y  $\alpha = 0.05$ .

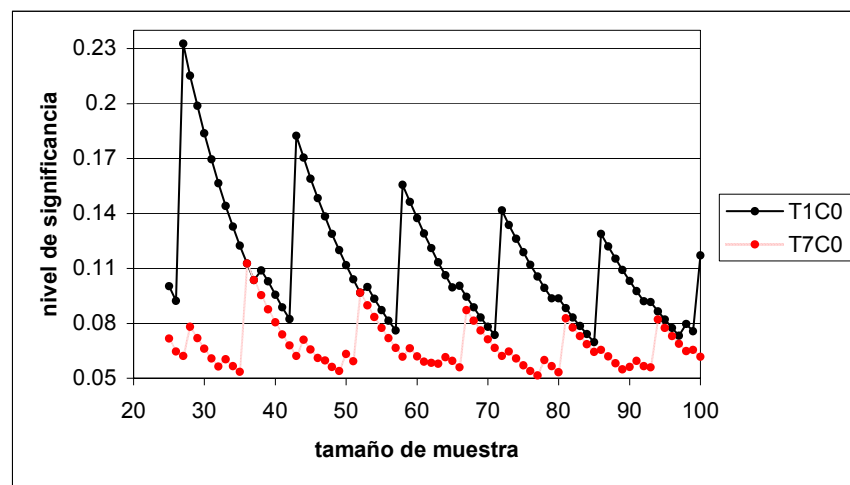


Figura 3.3: Niveles de significancia para las pruebas  $T_{1C0}$  y  $T_{7C0}$  para el límite de no-inferioridad  $d_0 = 0.1$  y  $\alpha = 0.05$ .

Las 42 pruebas estadísticas consideradas en este trabajo se clasificaron en dos tipos de acuerdo a los siguientes criterios (ver cuadros 3.4 y 3.5). La clasificación se hace con el objeto de tener en un grupo (las pruebas tipo  $\mathcal{A}$ ) a las pruebas razonablemente liberales y en el otro grupo (las pruebas tipo  $\mathcal{B}$ ) a las razonablemente conservadoras.

Una **prueba es del tipo  $\mathcal{A}$**  si para los tamaños de muestra en el rango  $25 \leq n \leq 100$ , al menos 90 % de los correspondientes niveles de significancia reales de la prueba

caen en el intervalo  $[0.045, 0.055]$ .

$d_0$	Prueba	Porcentaje de niveles de significancia reales que caen dentro del intervalo $[.045, .055]$
0.1	T2C1	97.37
	T5C1	93.42
	T7C4	93.42
0.15	T2C1	97.37
	T5C1	90.79
0.2	T2C1	94.74
	T5C1	94.74

Cuadro 3.4. Pruebas tipo  $\mathcal{A}$ .

Una **prueba es del tipo  $\mathcal{B}$**  si para los tamaños de muestra en el rango  $30 \leq n \leq 100$ , al menos 90 % de los correspondientes niveles de significancia reales de la prueba caen en el intervalo  $[0.04, 0.05]$ .

$d_0$	Prueba	Porcentaje de niveles de significancia reales que caen dentro del 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

Cuadro 3.5. Pruebas tipo  $\mathcal{B}$ .



## Capítulo 4

# Comparación de las pruebas

La comparación de las pruebas se llevó a cabo con base en sus niveles de significancia reales y en sus potencias, de la siguiente forma:

1. Se clasificaron las pruebas en dos grupos: de acuerdo a los dos tipos ( $\mathcal{A}$  y  $\mathcal{B}$ ) y se calcularon los porcentajes respectivos (Cuadros 3.4 y 3.5).

2. Para cada tipo de prueba ( $\mathcal{A}$  y  $\mathcal{B}$ ) y para cada valor de  $d_0$  (.1,.15,.2) se compararon las potencias de las pruebas únicamente para aquéllos tamaños de muestra donde la diferencia máxima entre los niveles de significancia reales resultó menor o igual que 0.0001 y donde al menos una de las potencias a comparar fuera mayor o igual que 0.7. La comparación de las potencias se realizó en los puntos  $(p_1, p_2) \in \Theta_{AD}$ , donde

$$\begin{aligned}\Theta_{AD} &= \left\{ (p_1, p_2) : p_1 = 0.05i, p_2 = \text{máx}(p_1 - d_0, 0) + 0.05j; \right. \\ &\quad \left. i = 0, 1, \dots, 20; j \text{ tal que } p_2 \in [\text{máx}(p_1 - d_0, 0), 1] \right\} \\ &= \left\{ (p_1, p_2) : p_1 = 0.05i, p_2 = \text{máx}(p_1 - d_0, 0) + 0.05j; \right. \\ &\quad \left. i = 0, 1, \dots, 20; j = 0, 1, \dots, 20(1 - \text{máx}(p_1 - d_0, 0)) \right\}\end{aligned}$$

## 4.1. Pruebas tipo $\mathcal{A}$

En el cuadro 4.1 se presentan los niveles de significancia reales para los tamaños de muestra  $25 \leq n \leq 100$  para los cuales la diferencia máxima fue menor o igual que 0.0001, para las tres pruebas del tipo  $\mathcal{A}$  cuando el límite de no-inferioridad se tomó  $d_0 = 0.1$ .

<b>n</b>	<b>T2C1</b>	<b>T5C1</b>	<b>T7C4</b>	<b>diferencia máxima</b>
<b>30</b>	0.04618	0.04618	0.04618	0.00000
<b>35</b>	0.04667	0.04667	0.04666	0.00001
<b>43</b>	0.04580	0.04580	0.04580	0.00000
<b>72</b>	0.05242	0.05242	0.05242	0.00000
<b>73</b>	0.05184	0.05184	0.05184	0.00000
<b>74</b>	0.05127	0.05127	0.05127	0.00000
<b>75</b>	0.05069	0.05069	0.05069	0.00000
<b>77</b>	0.04957	0.04957	0.04957	0.00000
<b>98</b>	0.05265	0.05265	0.05265	0.00000
<b>99</b>	0.05201	0.05201	0.05201	0.00000
<b>100</b>	0.05137	0.05137	0.05137	0.00000

Cuadro 4.1. Niveles de significancia reales para los tamaños de muestra  $25 \leq n \leq 100$  para las tres pruebas del tipo  $\mathcal{A}$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad  $d_0 = 0.1$ .

En el cuadro 4.2 se presenta un resumen de los valores obtenidos para comparar las potencias en el caso de las tres pruebas del tipo  $\mathcal{A}$ , para el límite de no inferioridad

$d_0 = 0.1$ .

Tamaños de muestra	30	35	43	72	73	74	75	77	98	99	100
Porcentaje de puntos en $\Theta_{AD}$ para los cuales $\beta_{T2C1} = \beta_{T5C1} = \beta_{T7C4}$	100	90	97	94	90	96	94	92	94	94	96
Max $\begin{bmatrix}  \beta_{T2C1} - \beta_{T5C1} , \\  \beta_{T2C1} - \beta_{T7C4} , \\  \beta_{T5C1} - \beta_{T7C4}  \end{bmatrix} (\times 10^2)$	0	3	3	3	4	6	5	3	3	3	2

Cuadro 4.2. Porcentajes de puntos donde se satisface la relación especificada entre las funciones de potencia. El porcentaje se calculó con relación a los puntos en  $\Theta_{AD}$ .

En el cuadro 4.3 se presentan los niveles de significancia reales para los tamaños de muestra  $25 \leq n \leq 100$  para los cuales la diferencia fue menor o igual que 0.0001, para las dos pruebas del tipo  $\mathcal{A}$  cuando el límite de no-inferioridad es  $d_0 = 0.15$ .

n	T2C1	T5C1	diferencia	n	T2C1	T5C1	diferencia
34	0.05343	0.05343	0.00000	67	0.04815	0.04815	0.00000
35	0.05205	0.05205	0.00000	68	0.04681	0.04681	0.00000
36	0.05069	0.05069	0.00000	76	0.05270	0.05270	0.00000
37	0.04937	0.04937	0.00000	77	0.05124	0.05124	0.00000
38	0.04809	0.04808	0.00000	78	0.04982	0.04982	0.00000
39	0.04683	0.04683	0.00000	79	0.04844	0.04844	0.00000
50	0.05344	0.05344	0.00000	80	0.04710	0.04710	0.00000
51	0.05193	0.05193	0.00000	82	0.04770	0.04770	0.00000
52	0.05047	0.05047	0.00000	87	0.05337	0.05337	0.00000
53	0.04906	0.04906	0.00000	88	0.05191	0.05191	0.00000
54	0.04769	0.04769	0.00000	89	0.05050	0.05050	0.00000
57	0.04675	0.04675	0.00000	90	0.04913	0.04913	0.00000
63	0.05397	0.05397	0.00000	93	0.04963	0.04963	0.00000
64	0.05245	0.05245	0.00000	98	0.05330	0.05330	0.00000
65	0.05097	0.05097	0.00000	99	0.05188	0.05188	0.00000
66	0.04954	0.04954	0.00000	100	0.05049	0.05049	0.00000

Cuadro 4.3. Niveles de significancia reales para los tamaños de muestra  $25 \leq n \leq 100$  para las dos pruebas del tipo  $\mathcal{A}$  para el nivel de significancia nominal

$\alpha = 0.05$  y el límite de no-inferioridad  $d_0 = 0.15$ .

En este caso (para las dos pruebas del tipo  $\mathcal{A}$  para el límite de no-inferioridad  $d_0 = 0.15$ ), para todos los tamaños de muestra donde se realizó la comparación, las potencias de las pruebas fueron iguales en el 100 % de los puntos.

En el cuadro 4.4 se presentan los niveles de significancia reales para los tamaños de muestra  $25 \leq n \leq 100$  para los cuales la diferencia fue menor o igual que 0.0001, para las dos pruebas del tipo  $\mathcal{A}$  cuando el límite de no-inferioridad es  $d_0 = 0.2$ .

<b>n</b>	<b>T2C1</b>	<b>T5C1</b>	<b>diferencia</b>	<b>n</b>	<b>T2C1</b>	<b>T5C1</b>	<b>diferencia</b>
<b>26</b>	0.0545	0.0545	0.0000	<b>62</b>	0.0523	0.0523	0.0000
<b>27</b>	0.0518	0.0518	0.0000	<b>63</b>	0.0498	0.0498	0.0000
<b>28</b>	0.0492	0.0492	0.0000	<b>64</b>	0.0475	0.0475	0.0000
<b>30</b>	0.0445	0.0445	0.0000	<b>69</b>	0.0539	0.0539	0.0000
<b>31</b>	0.0452	0.0452	0.0000	<b>70</b>	0.0514	0.0514	0.0000
<b>36</b>	0.0546	0.0546	0.0000	<b>71</b>	0.0491	0.0491	0.0000
<b>37</b>	0.0518	0.0518	0.0000	<b>75</b>	0.0501	0.0501	0.0000
<b>38</b>	0.0492	0.0492	0.0000	<b>77</b>	0.0525	0.0525	0.0000
<b>39</b>	0.0468	0.0468	0.0000	<b>78</b>	0.0502	0.0502	0.0000
<b>42</b>	0.0469	0.0469	0.0000	<b>79</b>	0.0480	0.0480	0.0000
<b>45</b>	0.0543	0.0543	0.0000	<b>84</b>	0.0532	0.0532	0.0000
<b>46</b>	0.0516	0.0516	0.0000	<b>85</b>	0.0509	0.0509	0.0000
<b>47</b>	0.0491	0.0491	0.0000	<b>86</b>	0.0487	0.0487	0.0000
<b>48</b>	0.0467	0.0467	0.0000	<b>91</b>	0.0535	0.0535	0.0000
<b>50</b>	0.0480	0.0480	0.0000	<b>92</b>	0.0512	0.0512	0.0000
<b>54</b>	0.0524	0.0524	0.0000	<b>98</b>	0.0535	0.0535	0.0000
<b>55</b>	0.0499	0.0499	0.0000	<b>100</b>	0.0492	0.0492	0.0000

Cuadro 4.4. Niveles de significancia reales para los tamaños de muestra  $25 \leq n \leq 100$  para las dos pruebas del tipo  $\mathcal{A}$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad  $d_0 = 0.2$ .

En este caso (para las dos pruebas del tipo  $\mathcal{A}$  para el límite de no-inferioridad  $d_0 = 0.2$ ), para todos los tamaños de muestra donde se realizó la comparación, las potencias de las pruebas fueron iguales en el 100 % de los puntos.

#### 4.1.1. Recomendaciones.

Para  $d_0 = 0.1$ , las potencias de las tres pruebas son prácticamente iguales, para  $d_0 = 0.15$ , las dos pruebas son iguales, también para  $d_0 = 0.2$ , las dos pruebas son iguales. Además como del cuadro 3.4 se tiene que la prueba  $T_{2C1}$  es la que tiene el mayor porcentaje de niveles de significancia reales en el intervalo  $[0.045, 0.055]$ , debido a lo anterior se considera a  $T_{2C1}$  como la mejor prueba del tipo  $\mathcal{A}$ .

En la figura 4.5 se presentan los niveles de significancia reales para la prueba  $T_{2C1}$  vs tamaños de muestra para los tres diferentes límites de no-inferioridad analizados.

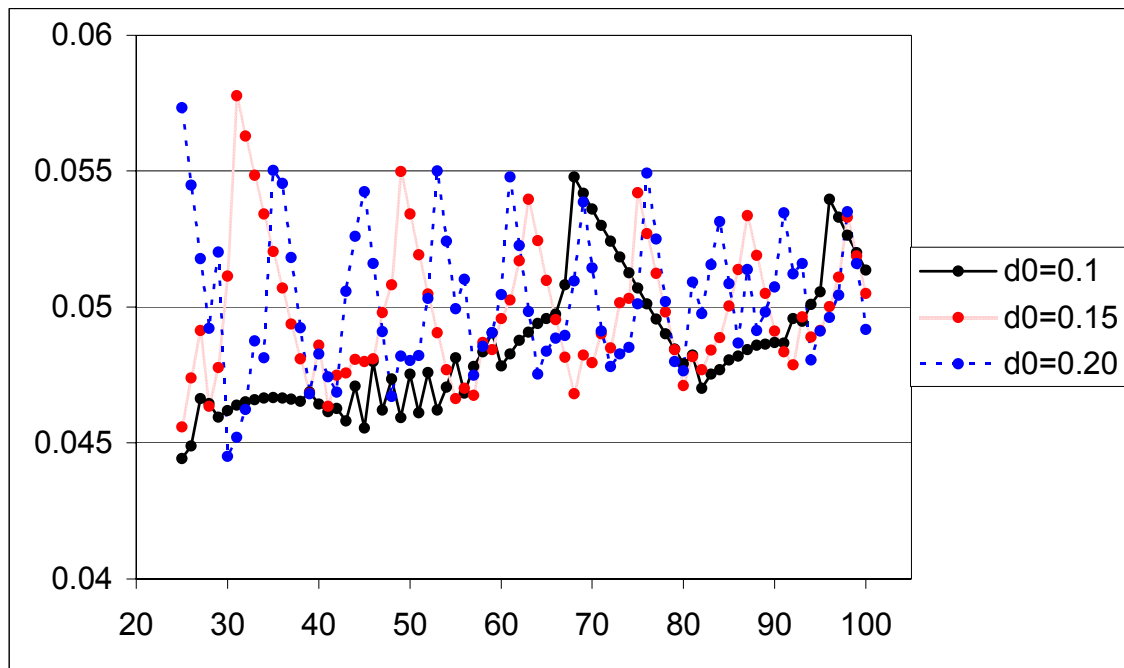


Figura 4.5: Niveles de significancia reales para la prueba  $T_{2C1}$  vs. tamaños de muestra para los diferentes límites de no inferioridad.

## 4.2. Pruebas tipo $\mathcal{B}$

Para las pruebas tipo  $\mathcal{B}$  se procedió en forma similar.

En los cuadros 4.6 a 4.8 se presentan los niveles de significancia reales para los tamaños de muestra  $30 \leq n \leq 100$  para los cuales la diferencia fue menor o igual que 0.0001, para las dos pruebas del tipo  $\mathcal{B}$  para los límites de no-inferioridad  $d_0 = 0.1$ , 0.15 y 0.2, respectivamente.

n	T2C2	T5C2	diferencia	n	T2C2	T5C2	diferencia
30	0.04618	0.04618	0.00000	50	0.04411	0.04411	0.00000
31	0.04636	0.04636	0.00000	51	0.04382	0.04382	0.00000
32	0.04650	0.04650	0.00000	52	0.04353	0.04353	0.00000
33	0.04659	0.04659	0.00000	53	0.04323	0.04323	0.00000
34	0.04665	0.04665	0.00000	54	0.04292	0.04292	0.00000
35	0.04666	0.04666	0.00000	55	0.04261	0.04261	0.00000
36	0.04665	0.04665	0.00000	56	0.04229	0.04229	0.00000
37	0.04660	0.04660	0.00000	57	0.04197	0.04197	0.00000
38	0.04652	0.04652	0.00000	66	0.04274	0.04274	0.00000
39	0.04642	0.04642	0.00000	78	0.04901	0.04901	0.00000
40	0.04630	0.04630	0.00000	79	0.04846	0.04846	0.00000
41	0.04615	0.04615	0.00000	80	0.04792	0.04792	0.00000
42	0.04598	0.04598	0.00000	81	0.04738	0.04738	0.00000
43	0.04580	0.04580	0.00000	82	0.04684	0.04684	0.00000
44	0.04560	0.04560	0.00000	83	0.04631	0.04631	0.00000
45	0.04538	0.04538	0.00000	84	0.04579	0.04579	0.00000
46	0.04515	0.04515	0.00000	85	0.04527	0.04527	0.00000
47	0.04491	0.04491	0.00000	86	0.04475	0.04475	0.00000
48	0.04465	0.04465	0.00000	87	0.04424	0.04424	0.00000
49	0.04438	0.04438	0.00000	88	0.04374	0.04374	0.00000

Cuadro 4.6: Niveles de significancia reales para los tamaños de muestra  $30 \leq n \leq 100$  para las dos pruebas del tipo  $\mathcal{B}$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad  $d_0 = 0.1$ .

n	T2C2	T5C2	diferencia	n	T2C2	T5C2	diferencia
38	0.04808	0.04808	0.00000	67	0.04815	0.04815	0.00000
39	0.04683	0.04683	0.00000	68	0.04681	0.04681	0.00000
40	0.04560	0.04560	0.00000	69	0.04550	0.04550	0.00000
41	0.04441	0.04441	0.00000	70	0.04424	0.04424	0.00000
42	0.04324	0.04324	0.00000	71	0.04301	0.04301	0.00000
43	0.04211	0.04211	0.00000	79	0.04844	0.04844	0.00000
45	0.03993	0.03993	0.00000	80	0.04710	0.04710	0.00000
47	0.04100	0.04100	0.00000	81	0.04581	0.04581	0.00000
54	0.04769	0.04769	0.00000	82	0.04455	0.04455	0.00000
55	0.04636	0.04636	0.00000	90	0.04913	0.04913	0.00000
56	0.04507	0.04507	0.00000	91	0.04779	0.04779	0.00000
57	0.04382	0.04382	0.00000	92	0.04650	0.04650	0.00000
58	0.04261	0.04261	0.00000	93	0.04524	0.04524	0.00000

Cuadro 4.7: Niveles de significancia para los tamaños de muestra  $30 \leq n \leq 100$  para las dos pruebas del tipo  $\mathcal{B}$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad  $d_0 = 0.15$ .

n	T2C2	T5C2	diferencia	n	T2C2	T5C2	diferencia
30	0.04448	0.04448	0.00000	71	0.04912	0.04912	0.00000
31	0.04231	0.04230	0.00001	72	0.04691	0.04691	0.00000
39	0.04681	0.04681	0.00000	73	0.04480	0.04480	0.00000
40	0.04450	0.04450	0.00000	79	0.04799	0.04799	0.00000
41	0.04231	0.04231	0.00000	80	0.04588	0.04588	0.00000
47	0.04909	0.04909	0.00000	81	0.04386	0.04386	0.00000
48	0.04671	0.04671	0.00000	86	0.04868	0.04868	0.00000
49	0.04445	0.04445	0.00000	87	0.04659	0.04659	0.00000
50	0.04230	0.04230	0.00000	88	0.04458	0.04458	0.00000
56	0.04757	0.04757	0.00000	93	0.04906	0.04906	0.00000
57	0.04532	0.04532	0.00000	94	0.04699	0.04699	0.00000
58	0.04318	0.04318	0.00000	95	0.04502	0.04502	0.00000
64	0.04754	0.04754	0.00000	96	0.04417	0.04417	0.00000
65	0.04534	0.04534	0.00000	100	0.04918	0.04918	0.00000

Cuadro 4.8: Niveles de significancia reales para los tamaños de muestra  $30 \leq n \leq 100$  para las dos pruebas del tipo  $\mathcal{B}$  para el nivel de significancia nominal  $\alpha = 0.05$  y el límite de no-inferioridad  $d_0 = 0.2$ .

Para los tres límites de no inferioridad para las pruebas del tipo  $\mathcal{B}$ , para todos los tamaños de muestra donde se realizó la comparación, las pruebas fueron iguales en el 100 % de los puntos.

#### 4.2.1. Recomendaciones.

Como las potencias de las dos pruebas fueron iguales para cada límite de no inferioridad ambas son igualmente buenas, sin embargo debido a que para las pruebas tipo  $\mathcal{A}$  la prueba recomendada fue de la familia  $T_2$ , solamente para hacer la recomendación más fácil de aplicar (es decir realizar menos programas de cómputo), se prefiere también en este caso una prueba de la familia  $T_2$ , así la prueba recomendada en este caso es  $T_{2C2}$ .

En la figura 4.26 se presentan los niveles de significancia reales para la prueba  $T_{2C2}$  vs tamaños de muestra para los tres diferentes límites de no-inferioridad analizados.



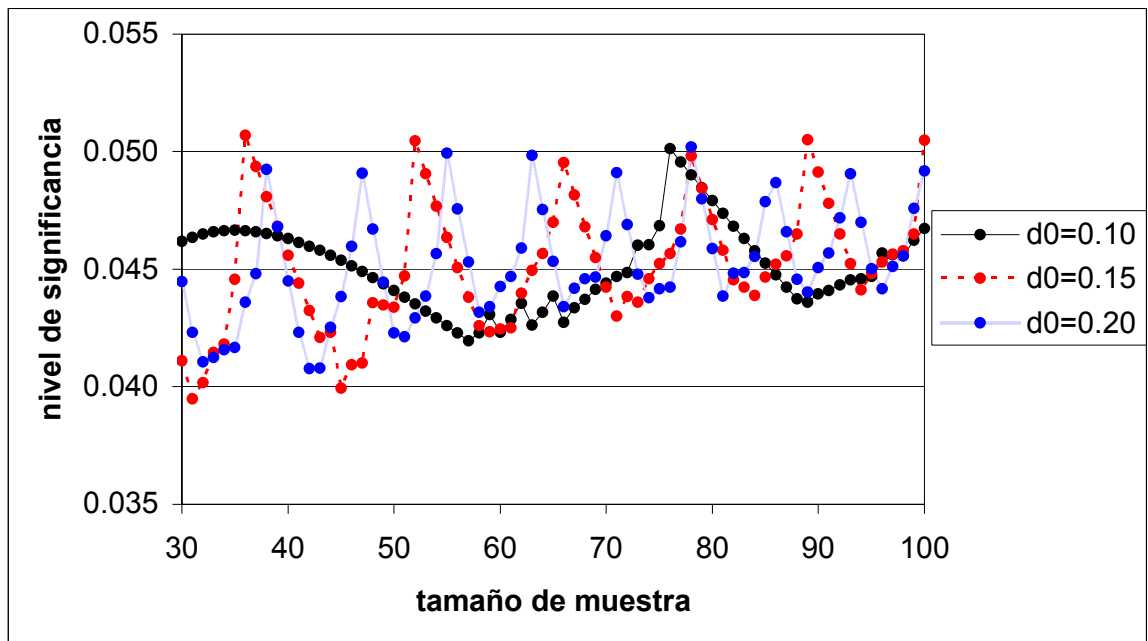


Figura 4.26: Niveles de significancia reales para la prueba  $T_{2C2}$  vs. tamaños de muestra para los diferentes límites de no inferioridad.

## Capítulo 5

# Conclusiones y posibles líneas de investigación futuras

De todo lo expuesto a lo largo de esta tesis, se resaltan a continuación las principales conclusiones obtenidas.

### 5.1. Conclusiones

- A partir de la clasificación de las pruebas en tipos  $\mathcal{A}$  y  $\mathcal{B}$  se puede recomendar de manera clara y teniendo en cuenta a las 42 pruebas analizadas que las mejores pruebas son, para pruebas tipo  $\mathcal{A}$ ,  $T_{2C1}$  mientras que para pruebas tipo  $\mathcal{B}$  es  $T_{2C2}$ .
- En términos generales, las pruebas para las cuales se utilizan como estimadores de proporciones a los estimadores de máxima verosimilitud restringida bajo la hipótesis nula ( $T_2$ ,  $T_5$  y  $T_7$ ) son mejores que aquellas que usan a los estimadores de máxima

verosimilitud ( $T_1$  y  $T_4$ ) y que las que usan estimadores de máxima verosimilitud generalizados, es decir los estimadores de Böhning-Viwatwongkassen ( $T_3$  y  $T_6$ ).

- La prueba de  $T_2$  es muy utilizada y recomendada en la mayoría de los artículos sobre no-inferioridad, aunque no realizan un análisis detallado y metodológicamente correcto como el realizado aquí; sin embargo de acuerdo a los resultados de esta tesis, no siempre  $T_2$  es la mejor, de hecho  $T_{2C0}$  nunca es la mejor.

- El presente trabajo contribuye a la realización de un análisis detallado y metodológicamente correcto para la comparación de pruebas de no inferioridad, mediante un procedimiento claro, el cual podría aplicarse a otras pruebas.

- Chen et al. [9] comparan pruebas de no-inferioridad asintóticas para la diferencia de dos proporciones, en una forma bastante elemental pues analizan el comportamiento de los niveles actuales (evaluaciones de la función potencia en algunos puntos) y no de los niveles de significancia reales.

- Se generaliza el muy útil teorema de Röhmel J. y Mansmann [25] en dos aspectos, esto contribuye a llenar un hueco existente en la metodología para el cálculo de niveles de significancia reales de pruebas de no-inferioridad, sentando con ello las bases para poder realizar comparaciones de pruebas, incluidas las de Blackwelder y Hauck-Anderson, para el caso de diseños desbalanceados (dichas comparaciones no se realizaron en este trabajo).

## 5.2. Posibles líneas de investigación futuras

Las posibles líneas de investigación que se desprenden de este trabajo.

Comparación de pruebas asintóticas de no-inferioridad para otras medidas de disimilaridad.

Comparación de pruebas asintóticas para diseños desbalanceados.

Comparación de pruebas exactas de no-inferioridad para la diferencia.

Comparación de pruebas exactas de no-inferioridad considerando otras medidas de disimilaridad.

Estudiar condiciones duales a (C); (SC1) y (SC2).

Obtención de resultados teóricos que permitan simplificar el cálculo de los niveles de significancia reales para pruebas de equivalencia o en su defecto, construcción de contraejemplos.

Comparación de pruebas de equivalencia exactas y asintóticas.

# Apéndice A

## Validación de los cálculos

En este apéndice se presentan algunas comparaciones de los cálculos realizados en este trabajo con los realizados en algunos artículos, esta comparación se realizó con el objetivo de corroborar la validación de los programas de cómputo empleados en este trabajo para calcular niveles de significancia reales y potencias de las pruebas evaluadas.

Es conveniente aclarar que Chan [6] evalúa la función de potencia en algunos puntos  $(p_1, p_2)$  para las versiones asintótica y exacta de la prueba  $T_{2C0}$  y también calcula los niveles de significancia reales de la prueba asintótica  $T_{2C0}$ , en el cuadro 5.1 se cotejan algunos de los resultados de esta tesis con los resultados presentados por Chan.

p1	d0	n	potencia en (p1,p2)			pruebas asintóticas					
			Chan	Tesis	Diferencia	potencia en (p1,p2)			nivel de significancia		
						Chan	Tesis	Diferencia	Chan	Tesis	Diferencia
0.95	0.1	35	0.0485	0.048500	0.0000	0.0500	0.049977	0.0000	0.0536	0.053646	0.0000
0.95	0.1	70	0.0498	0.049800	0.0000	0.0448	0.044812	0.0000	0.0536	0.053600	0.0000
0.95	0.2	35	0.0496	0.049600	0.0000	0.0464	0.046425	0.0000	0.0574	0.060524	0.0031
0.8	0.1	100	0.0491	0.050000	0.0009	0.0509	0.050937	0.0000	0.0516	0.057577	0.0060
0.8	0.15	25	0.0456	0.045600	0.0000	0.0520	0.052014	0.0000	0.0673	0.067301	0.0000
0.8	0.15	85	0.0491	0.049100	0.0000	0.0508	0.050826	0.0000	0.0564	0.056414	0.0000
0.6	0.2	90	0.0499	0.049900	0.0000	0.0558	0.055810	0.0000	0.0558	0.055810	0.0000
0.6	0.25	20	0.0484	0.048400	0.0000	0.0505	0.050500	0.0000	0.0591	0.059070	0.0000
0.6	0.25	50	0.0480	0.048000	0.0000	0.0502	0.050249	0.0000	0.0527	0.052714	0.0000
<b>Diferencia máxima</b>					<b>0.0009</b>			<b>0.0000</b>			<b>0.0060</b>

Cuadro 5.1. Niveles de significancia y potencias de la prueba  $T_{2C0}$  calculados por Chan y los calculados en esta tesis.

Es muy importante hacer notar que de acuerdo con el cuadro anterior, en dos ocasiones el nivel de significancia real obtenido en este trabajo para la prueba asintótica es mayor que el de Chan, esto se debe de manera específica al hecho de que como se mencionó arriba, en este trabajo se evaluó la función de potencia en los puntos  $(d_0 + i\Delta, i\Delta)$  para  $i = 0, \dots, k$ , donde  $k$  es el número de subintervalos de igual longitud ( $\Delta$ ) en que dividimos al intervalo  $[d_0, \frac{1+d_0}{2}]$ , en este trabajo se tomó  $\Delta = 0.001$ . Nótese que de esta forma el primer punto donde se evalúa la función potencia es  $(d_0, 0)$  sin embargo se ha podido corroborar que Chan quita este punto (si en este trabajo se hiciera lo mismo se obtendría exactamente el mismo valor) y empieza a evaluar en el que para esta tesis es el segundo punto, es decir, en  $(d_0 + \Delta, \Delta)$ , esta práctica no está justificada de forma alguna y mucho menos a la luz del teorema 14.

Chen et al. [9] calcularon los valores de la función potencia en algunos puntos  $(p_1, p_2)$  para los tamaños de muestra  $n = 75, 100, 150$  para la prueba  $T_{2C0}$ ; dichos valores son estimaciones basadas en simulación, por el contrario, en este trabajo cuando se calcula la función de potencia en algún punto  $(p_1, p_2)$  no se estima por simulación sino que se calcula su valor *exacto*, en el cuadro 5.2 se comparan algunos de los valores presentados por Chen con los obtenidos en esta tesis. En ninguna parte de su trabajo, Chen et al. [9] calcularon los niveles de significancia reales.

d	p1	p2	75			100			150		
			CTK	Tesis	Diferencia	CTK	Tesis	Diferencia	CTK	Tesis	Diferencia
0.1	0.5	0.4	0.024	0.024362	0.0004	0.0270	0.027152	0.0002	0.0270	0.027371	0.0004
		0.5	0.231	0.231328	0.0003	0.3100	0.310365	0.0004	0.4310	0.431264	0.0003
		0.6	0.691	0.691426	0.0004	0.8240	0.823532	0.0005	0.9420	0.942087	0.0001
	0.7	0.6	0.026	0.026412	0.0004	0.0250	0.025033	0.0000	0.0250	0.025338	0.0003
		0.7	0.271	0.271044	0.0000	0.3430	0.343053	0.0001	0.4760	0.476038	0.0000
		0.8	0.809	0.808994	0.0000	0.9060	0.905702	0.0003	0.9800	0.980062	0.0001
	0.9	0.8	0.023	0.023331	0.0003	0.0250	0.024830	0.0002	0.0250	0.024687	0.0003
		0.9	0.478	0.477799	0.0002	0.6230	0.623155	0.0002	0.8030	0.802848	0.0002
		0.95	0.909	0.909058	0.0001	0.9740	0.974401	0.0004	0.9980	0.998165	0.0002
0.2	0.5	0.3	0.026	0.026015	0.0000	0.0250	0.024727	0.0003	0.0250	0.024595	0.0004
		0.5	0.716	0.716112	0.0001	0.8210	0.820999	0.0000	0.9410	0.940557	0.0004
		0.7	0.999	0.9991	0.0001	1.0000	0.999947	0.0001	1.0000	1.000000	0.0000
	0.7	0.5	0.026	0.026015	0.0000	0.0250	0.024727	0.0003	0.0250	0.024595	0.0004
		0.7	0.771	0.770565	0.0004	0.8770	0.876614	0.0004	0.9680	0.967768	0.0002
		0.9	1.000	0.999993	0.0000	1.0000	1.000000	0.0000	1.0000	1.000000	0.0000
	0.9	0.7	0.024	0.024279	0.0003	0.0250	0.024580	0.0004	0.0240	0.024372	0.0004
		0.9	0.970	0.970331	0.0003	0.9940	0.994022	0.0000	1.0000	0.999816	0.0002
		0.95	1.000	0.999788	0.0002	1.0000	0.999995	0.0000	1	1.000000	0.0000
<b>Diferencia máxima</b>			<b>0.0004</b>			<b>0.0005</b>			<b>0.0004</b>		

Cuadro 5.2. Potencia de la prueba  $T_{2C0}$  calculados por Chan y los calculados en esta tesis.

Munk et al. [21] calcularon los valores de la función potencia en algunos puntos  $(p_1, p_2)$  para la prueba  $T_{7C0}$  para  $d_0 = 0.1$ , los valores obtenidos por Munk no son estimaciones mediante simulación sino que son los valores exactos al igual que los que los obtenidos en esta tesis. A continuación se presenta una comparación de los resultados de este trabajo con los de Munk. En ninguna parte de su trabajo, Munk et al. [21] calcularon los niveles de significancia reales.

p1	p2	n	Skipka	Tesis	Diferencia
0.2	0.1	10	0.0893	0.089301	0.0000
0.2	0.1	25	0.0533	0.053054	0.0002
0.2	0.1	50	0.0545	0.054456	0.0000
0.2	0.1	100	0.0522	0.052200	0.0000
0.2	0.1	500	0.0505	0.050454	0.0000
0.5	0.4	10	0.0595	0.059509	0.0000
0.5	0.4	25	0.0446	0.044593	0.0000
0.5	0.4	50	0.0451	0.045076	0.0000
0.5	0.4	100	0.0505	0.050506	0.0000
0.5	0.4	500	0.0518	0.051800	0.0000
<b>Diferencia máxima</b>					<b>0.0002</b>

Cuadro 5.3. Potencia de la prueba  $T_{2C0}$  calculados por Munk y los calculados en esta tesis.

*Las comparaciones exhibidas en los cuadros 5.1 a 5.3 corroboran la validación de los programas de cómputo empleados en este trabajo.*



## Apéndice B

### Resultados básicos y simbología

A continuación se presentan sin demostración algunos resultados conocidos, las demostraciones de tales resultados se encuentran establecidas en [13].

**Teorema 32** *Si  $A_n$  y  $B_n$  son dos sucesiones de variables aleatorias las cuales satisfacen*

$$A_n \xrightarrow{P} a \text{ y } B_n \xrightarrow{P} b$$

*respectivamente, entonces*

$$A_n + B_n \xrightarrow{P} a + b$$

$$A_n - B_n \xrightarrow{P} a - b$$

$$A_n \cdot B_n \xrightarrow{P} ab$$

$$A_n/B_n \xrightarrow{P} a/b, b \neq 0$$

**Teorema 33** *Si  $Y_n$  es una sucesión de variables aleatorias tales que  $Y_n \xrightarrow{P} c$ , y si  $f$  es una función continua en  $c$ , entonces  $f(Y_n) \xrightarrow{P} f(c)$*

**Teorema 34** (*Slutsky*) Si  $Y_n \xrightarrow{L} Y$  y  $A_n \xrightarrow{P} a$ ,  $B_n \xrightarrow{P} b$  entonces  $A_n + B_n Y_n \xrightarrow{L} a + bY$

**Teorema 35** Una condición suficiente para que  $\delta_n$  sea un estimador consistente de  $g(\theta)$  es que tanto el sesgo como la varianza de  $\delta_n$  tiendan a cero cuando  $n \rightarrow \infty$ .

**Lema 36** Si  $Y_n \xrightarrow{L} Y$ ,  $a$  y  $b$  son constantes con  $b \neq 0$ , entonces  $bY_n + a \xrightarrow{L} bY + a$

**Proposición 37** Sean  $V_1, \dots, V_{n_1}$  y  $W_1, \dots, W_{n_2}$  independientes con distribuciones  $F_1$  y  $F_2$ , respectivamente; medias respectivas  $\mu_1$  y  $\mu_2$  y varianzas finitas respectivas  $\sigma^2 > 0$  y  $\tau^2 > 0$ . Sean  $N = n_1 + n_2$ ,  $\sigma_N = \sqrt{\frac{\sigma^2}{n_1} + \frac{\tau^2}{n_2}}$ . Si  $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{N} = \zeta \in (0, 1)$ , entonces

$$\frac{\bar{V} - \bar{W} - (\mu_1 - \mu_2)}{\sigma_N} \xrightarrow{L} N(0, 1)$$

A continuación se presenta la simbología utilizada en el trabajo

<b>Símbolo</b>	<b>Explicación</b>
$B(n, p)$	Distribución Binomial con parámetros $n$ y $p$ .
$f(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$	Función de probabilidad de la distribución binomial con parámetros $n$ y $p$
$H_0 : p_1 - p_2 \geq d_0$	Hipótesis nula
$d = p_1 - p_2$	Diferencia de proporciones
$d_0$	Límite de no-inferioridad
$EMV$	Estimador de máxima verosimilitud
$EMVR$	Estimador de máxima verosimilitud restringido
$\hat{d} = \hat{p}_1 - \hat{p}_2$	Estimador de la diferencia de proporciones
$\hat{\sigma}_1$	Estimador de Blackwelder para $\sqrt{Var(\hat{d})}$
$\hat{\sigma}_2$	Estimador de Farrington-Manning(FM) para $\sqrt{Var(\hat{d})}$
$\hat{\sigma}_3$	Estimador de Böhning-Viwatwongkasen(BV) para $\sqrt{Var(\hat{d})}$
$\hat{\sigma}_4$	Estimador de Hauck-Anderson(HA) para $\sqrt{Var(\hat{d})}$
$\hat{\sigma}_5$	Estimador de FM-HA para $\sqrt{Var(\hat{d})}$
$\hat{\sigma}_6$	Estimador de BV-HA para $\sqrt{Var(\hat{d})}$
$\hat{p}_i = \frac{x_i}{n_i}$	EMV de $p_i$ .
$\check{p}_1$	EMVR de $p_1$ bajo la hipótesis $p_1 - p_2 \geq d_0$
$\check{p}_2 = \check{p}_1 - d_0$	EMVR de $p_2$ bajo la hipótesis $p_1 - p_2 \geq d_0$
$\tilde{p}_{i,b} = \frac{x_i + b}{n_i + 2b}$	Estimador de BV de $p_i$ .

# Apéndice C

## Programas de cómputo

En este apéndice se presentan dos programas representativos de los realizados para los cálculos, todos ellos fueron escritos en el lenguaje de programación S-PLUS. El lector interesado podrá obtener los demás programas solicitándolos al autor en la siguiente dirección electrónica: [falmendra@colpos.mx](mailto:falmendra@colpos.mx).

### C.1. Programa para verificar si una región crítica para la prueba $T_7$ cumple o no la condición

(C)

```
#  
# Verificación de la CONDICIÓN DE CONVEXIDAD DE BARNARD  
#  
CCB_T7U <- function(n1,n2,alfa,d)  
{  
  T7U <- matrix(0,nrow=n1+1,ncol=n2+1)  
  zalfa <- qnorm(1-alfa,0,1)
```

```

A <- 0
i <- 0
while (i<n1+1)
{
  j <- 0
  while (j<n2+1)
  {
    p1g <- i/n1
    p2g <- j/n2
    vc <- dbinom(i,n1,p1g)*dbinom(j,n2,p2g)
    #print(c(i,j,vc))
    rU <- -(i+j+n1*(1+2*d)+n2*(1+d))/(n1+n2)
    sU <- (j+i*(1+2*d)+d*(n2+n1*(1+d)))/(n1+n2)
    tU <- -(i*d*(1+d))/(n1+n2)
    DU <- sqrt(rU^2-3*sU)/3
    EU <- 2*rU^3/27-rU*sU/3+tU
    coU <- -EU/(2*DU^3)
    if ( coU < -1) coU <- -1
    if ( coU > 1) coU <- 1
    p1r <- 2*DU*cos(1/3*acos(coU)+4*pi/3)-rU/3
    p2r <- p1r - d
    if ( p1r < 0 ) p1r <- 0
    if (p1r > 1) p1r <- 1
    if ( p2r < 0 ) p2r <- 0
    if (p2r > 1) p2r <- 1
    #print(c(p1r,p2r))
    vr <- dbinom(i,n1,p1r)*dbinom(j,n2,p2r)
    T7U[i+1,j+1] <- (vr+A)/vc
    j <- j+1
  }
  i <- i+1
}
#print(round(Z,4))
# verificación de la condición (a)
j <- 0
while (j<n2+1)
{
  i <- 1
  while (i<n1+1)
  {
    if (T7U[i+1,j+1] < -zalfa && s7 != 0)
    {
      if (T7U[i,j+1] >= -zalfa)
      {
        print("no se cumple (a)")
        print(T7U[i+1,j+1])
      }
    }
  }
}

```

```

        i <- i+1
    }
    j <- j+1
}
# verificación de la condición (b)
i <- 0
while (i<n1+1)
{
    j <- 0
    while (j<n2)
    {
        if (T7U[i+1,j+1] < -zalfa && s7 != 0)
        {
            if (T7U[i+1,j+2] >= -zalfa)
            {
                print("no se cumple (b)")
                print(T7U[i+1,j+1])
            }
        }
        j <- j+1
    }
    i <- i+1
}
}
for (m in 2:20)
{
    for (k in 1:4)
    {
        CCB_T7U(5*m,5*m,.05,.05*k)
        print(c(5*m,.05,.05*k))
    }
}
}

```

## C.2. Programa para determinar los niveles de significancia reales para la prueba $T_7$ .

```

#
# CALCULO DE NIVELES DE SIGNIFICANCIA PARA LA ESTADISTICA T7U
#
#
# ..... INICIALIZACION .....
# Maximo nivel de No Inferioridad permitido")
d <- .1
alfa <- .05
# Maximo porcentaje permitido de NS fuera del intervalo (.1
recomendado)
pct <- .3
# tamaño de muestra inicial
m1 <- 30
# tamaño de muestra final
m2 <- 30
# con tamaños de muestra multiples de
mtm <- 1
# numero de particiones del intervalo [0,1]
npar <- 1000
# inicializar indns en 3 para calcular unicamente los NS
indns <- 3
# NOTAS:
# Para calcular en una CC especifica (v. gr. 3), en lugar de
icc in 1:6 poner icc in 3:3
# ..... TERMINA INICIALIZACION .....
nns <- ceiling((m2-m1+1)*pct)
#Vector de limites inferiores de los intervalos para NS
lii <- c(.045,.04)
# Vector de limites superiores de los intervalos para NS
lsi <- c(.055,.05)
for (icc in 1:1)
{
contf <- c(0,0,0)
NivSigT7U <- 1:m2
for (w in m1:m2)
{
n1 <- mtm*w

```

```

# Vector de correcciones por continuidad
CC <- c(0,1/(4*n1),1/(2*n1),1/n1,3/(2*n1),2/n1)
# Especificacion de diseño Balanceado
n2 <- n1
nr <- 0
SUMA7U <- 0
i <- 0
#
# obtención del numero de renglones para la matriz
# de la region de rechazo
#
NumReg <- function(n1,n2,alfa,d)
{  while (i < n1+1)
    {  j <- 0
        while (j < n2+1)
        {
            p1g <- i/n1
            p2g <- j/n2
            vc <- dbinom(i,n1,p1g)*dbinom(j,n2,p2g)
            rU <- -(i+j+n1*(1+2*d)+n2*(1+d))/(n1+n2)
            sU <- (j+i*(1+2*d)+d*(n2+n1*(1+d)))/(n1+n2)
            tU <- -(i*d*(1+d))/(n1+n2)
            DU <- sqrt(rU^2-3*sU)/3
            EU <- 2*rU^3/27-rU*sU/3+tU
            coU <- -EU/(2*DU^3)
            if ( coU < -1) coU <- -1
            if ( coU > 1) coU <- 1
            p1rU <- 2*DU*cos(1/3*acos(coU)+4*pi/3)-rU/3
            p2rU <- p1rU - d
            if ( p1rU < 0 ) p1rU <- 0
            if (p1rU > 1) p1rU <- 1
            if ( p2rU < 0 ) p2rU <- 0
            if (p2rU > 1) p2rU <- 1
            vrU <- dbinom(i,n1,p1rU)*dbinom(j,n2,p2rU)
            T7U <- (vrU)/vc+CC[icc]
            #print(c(i,j,p1rU,p2rU,vrU,vc,lam))
            if ( p1g-p2g<d && -2*log(T7U) > qchisq(1-2*alfa,1))
            {  nr <- nr + 1
                }
            j <- j+1
        }
        i <- i+1
    }
}

```



```

    }
    print(nr)
}
# La siguiente instruccion es indispensable
nr <- NumReg(n1,n2,alfa,d)
#
# Obtencion de la matriz de la region de rechazo
#
Matriz <- function(n1,n2,alfa,d)
{
  I <- matrix(0,nrow=nr,ncol=2)
  cont <- 1
  i <- 0
  while (i<n1+1)
  {
    j <- 0
    while (j<n2+1)
    {
      p1g <- i/n1
      p2g <- j/n2
      vc <- dbinom(i,n1,p1g)*dbinom(j,n2,p2g)
      rU <- -(i+j+n1*(1+2*d)+n2*(1+d))/(n1+n2)
      sU <- (j+i*(1+2*d)+d*(n2+n1*(1+d)))/(n1+n2)
      tU <- -(i*d*(1+d))/(n1+n2)
      DU <- sqrt(rU^2-3*sU)/3
      EU <- 2*rU^3/27-rU*sU/3+tU
      coU <- -EU/(2*DU^3)
      if ( coU < -1) coU <- -1
      if ( coU > 1) coU <- 1
      p1rU <- 2*DU*cos(1/3*acos(coU)+4*pi/3)-rU/3
      p2rU <- p1rU - d
      if ( p1rU < 0 ) p1rU <- 0
      if (p1rU > 1) p1rU <- 1
      if ( p2rU < 0 ) p2rU <- 0
      if (p2rU > 1) p2rU <- 1
      vrU <- dbinom(i,n1,p1rU)*dbinom(j,n2,p2rU)
      T7U <- vrU/vc+CC[icc]
      if ( p1g-p2g<d && -2*log(T7U) > qchisq(1-2*alfa,1))
      {
        I[cont,1] <- i
        I[cont,2] <- j
        cont <- cont + 1
      }
      j <- j+1
    }
    i <- i+1
  }
}

```

```

    }
    write(I,file="matind1")
}
#
# Las siguientes 2 instrucciones son indispensables
#
Matriz(n1,n2,alfa,d)
mrr <- matrix( scan("matind1"), ncol=2)
print(mrr)
#
# Función de potencia
#
FPT7U <- function(p1,n1,p2,n2,alfa,d)
{
  SUMA7U <- 0
  i <- 1
  while (i < nr+1)
    {
      SUMA7U <- SUMA7U+ dbinom(mrr[i,1],n1,p1)*
                        dbinom(mrr[i,2],n2,p2)

      i <- i+1
    }
  print(SUMA7U)
}
#
# Determinacion del nivel de significancia reales
#
if ((npar*d)%2==0) rep <- (npar/2)*(1-d)+1
if ((npar*d)%2==1) rep <- (npar+1-npar*d)/2
# cantidad de NS a calcular
NST7U <- matrix(0,nrow=rep,ncol=rep )
NST7 <- 1:rep
for (i in 1:rep)
# Para NS para T7U poner j in i:i
# Para NS para T7U poner j in 1:i
  { for (j in i: i)
    {
      # Monitoreo de resultados
      print("# iteracion          muestra% fuera de intervalos
tipo CC")
      print(c(i,"          ", n1,
round(100*contf[1]/(m2-m1+1),2),
round(100*contf[2]/(m2-m1+1),2),,icc-1))
      NST7U[i,j] <- FPT7U(d+(i-1)/npar,n1,(j-1)/
npar,n2,alfa,d)
    }
  }
}

```

```

    }
    if ( NST7U[i,i] > lsi[1] && NST7U[i,i] > lsi[2] &&
indns != 3) break
  }
  print("Nivel de significancia reales Asintotico para la prueba
de No Inferioridad")
  NivSigT7U[w] <- max(NST7U)
  print(NivSigT7U[w])
  if (NivSigT7U[w] > lsi[1] || NivSigT7U[w] < lii[1])    contf[1]
<- contf[1] + 1
  if (NivSigT7U[w] > lsi[2] || NivSigT7U[w] < lii[2])    contf[2]
<- contf[2] + 1
  if (100-w+contf[1] < nns && 100-w+contf[2] < nns && indns
!= 3) break
  if (contf[1] >= nns && indns != 3)
  {   print("..... NO ACEPTABLE .....")
      print(Çorreccion por continuidad")
      switch(icc,print(ç0"),print(ç1"),print(ç2"),
print(ç3"),print(ç4"),print(ç5"))
      print(c("Intervalo", "[.045,.055]"))
  }
  if (contf[2] >= nns && indns != 3)
  {   print("..... NO ACEPTABLE .....")
      print(c("prueba", "T7"))
      print(Çorreccion por continuidad")
      switch(icc,print(ç0"),print(ç1"),print(ç2"),
print(ç3"),print(ç4"),print(ç5"))
      print(c("Intervalo", "[.04,.05]"))
  }
  if (contf[1] >= nns && contf[2] >= nns && indns != 3) break
}
print(NivSigT7U)
}

```

# Bibliografía

- [1] Agresti A., Caffo B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *American Statistician*. **54**, 280-288.
- [2] Barnard, G. A. (1947). Significance tests for 2x2 tables. *Biometrika*. **34**, 123-138.
- [3] Blackwelder W. C. (1982) “Proving the null hypothesis” in clinical trials. *Controlled Clinical Trials*. **3**, 345-353.
- [4] Böhning, D, Viwatwongkasen, C. (2005). Revisiting proportion estimators. *Statistical methods in medical research*. **14**, 1-23.
- [5] Casella, G. and Berger, R. L. (2002) *Statistical Inference*. Second Edition. Thompson.
- [6] 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.
- [7] Chan, I. S. F. (1999). *Author’s reply on* Exact tests of equivalence and efficacy

- with a non zero lower bound for comparative studies. *Statistics in Medicine*. **18**, 1735-1737.
- [8] Chan, I. S. F. (2003). Proving non-inferiority or equivalence of two treatments with dichotomous endpoints using exact methods. *Statistical methods in medical research*. **12**, 37-58.
- [9] Chen, J. Tsong, Y. and Kang, S. (2000). Test for equivalence or noninferiority between two proportions. *Drug Information Journal*. **34**, 569-578.
- [10] Dunnett, C. W. . Gent, M. (1977). Significance testing to establish equivalence between treatments with special reference to data in the form 2x2 tables. *Biometrics*. **33**, 593-602
- [11] Farrington, C. and Manning, G. (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
- [12] Hauck, W. and Anderson, S. (1986) comparison of large-sample confidence interval methods for the difference of two binomial probabilities. *Am Stat*. **40**, 318-322
- [13] Lehmann, E. (1999). *Elements of Large-sample theory*. Springer.
- [14] Li, Z. and Chuang-Stein, C. (2006). A note on comparing two binomial pro-

- portions in confirmatory noninferiority trials. *Drug Information Journal*. **40**, 203-208.
- [15] Liu, J.P. and Weng, C.S. (1993). Evaluation of parametric and nonparametric two one-sided tests procedures for assessing bioequivalence of average bioavailability. *Journal of Biopharmaceutical Statistics*. **3**(1), 85-102.
- [16] Martin, A. and Herranz, I. (2002). Equivalence testing for binomial random variables: which test to use?. *The American Statistician* **56**(3): 253-254, Letter to the editor.
- [17] Martin, A. and Herranz, I. (2004). Asymptotical test on the equivalence, substantial difference and non-inferiority problems with two proportions. *Biometrical Journal*. **46**, 305-319.
- [18] Martin, A. and Herranz, I. (2004). Exact unconditional non-classics tests on the difference of two proportions. *Computational Statistics & Data Analysis*. **45**, 373-388.
- [19] Mehrotra, D. V., Chan I. S. F. and Berger R. L. (2003) A cautionary note on exact unconditional inference for the difference between two independent binomial proportions. *Biometrics*. **59**, 441-450.
- [20] Miettinen, O. and Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine*. **4**, 213-226.

- [21] Munk, A., Skipka, G. and Stratmann, B. (2005). Testing general hypotheses under binomial sampling: The two sample case - asymptotic theory and exact procedures. *Computational Statistics and Data Analysis*. **49(3)** 723-739.
- [22] Robertson T., Wright F. T., Dykstra R. L. (1988). *Order Restricted Statistical Inference*. Jhon Wiley & Sons, Chichester.
- [23] Rodary C., Com-Nougue C. and Tournade M. (1989). How to establish equivalence between treatments: a one-sided clinical trial in paediatric oncology. *Stat Med*; **8**; 593-598
- [24] Röhmel J. and Mansmann, U. (1999). *Letter to the editor on* Exact test of equivalence and efficacy with a non-zero lower bound for comparative studies. *Statistics in Medicine*. **18**, 1734-1735.
- [25] Röhmel J. and Mansmann, U. (1999). Unconditional nonasymptotic onesided tests for independent binomial proportions when the interest lies in showing noninferiority and or superiority. *Biometrical Journal*. **2**, 149-170.
- [26] 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 Journal*. **47**, 1, 37-47.
- [27] Ruiz-Maya Pérez, L. y Martín Pliego, F. J. (2002). *Fundamentos de inferencia estadística*. Thompson. (España).

- [28] Tu, D. (1997). A comparative study of some statistical procedures in establishing therapeutic equivalence of nonsystemic drugs with binary endpoints. *Drug Information Journal*. **31**, 1291-1300.