



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN EN
CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

CORRECCIÓN POR CONTINUIDAD PARA LA PRUEBA
ASINTÓTICA DE NO-INFERIORIDAD DE LASTER-
JOHNSON-KOTLER PARA DOS PROPORCIONES
INDEPENDIENTES

YONI MIGUEL CASTILLO TZEC

T E S I S

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO
2013

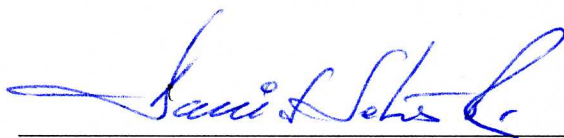
La presente tesis titulada: **CORRECCIÓN POR CONTINUIDAD PARA LA PRUEBA ASINTÓTICA DE NO-INFERIORIDAD DE LASTER-JOHNSON-KOTLER PARA DOS PROPORCIONES INDEPENDIENTES**, realizada por el alumno: **YONI MIGUEL CASTILLO TZEC**, bajo la dirección del consejo particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO



Dr. David Antonio Sotres Ramos

ASESOR



Dr. Félix Almendra Arao

ASESOR



Dr. Eduardo Gutiérrez González

Montecillo, Texcoco, Estado de México, Marzo de 2013

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el apoyo económico brindado para poder realizar mis estudios de Maestría.

Al Colegio de Postgraduados, por permitirme la realización de una etapa más en mi formación académica.

Al Dr. David A. Sotres Ramos por su paciencia, tiempo dedicado, valiosa aportación y ayuda en la realización de este trabajo.

A los miembros de mi Consejo Particular, Dr. Félix Almendra Arao y Dr. Eduardo Gutiérrez González, por su colaboración en la revisión del presente trabajo.

A mis padres y hermanos, por la confianza y apoyo brindados.

A mis familiares y amigos.

Dedicatoria

A mis padres:

Semey Eleazar Castillo Kú

Olga María Tzec Uh

A mis hermanos:

Darwin, Jaime, Lidia, Antonio, Felipe, Eleazar,

Candelario, Olga, Asunción, Edgar.

Índice general

Índice general	i
Índice de figuras	iii
Índice de cuadros	vi
Resumen	vii
Abstract	viii
1 Introducción	1
1.1 Introducción	1
1.2 Objetivos	4
1.3 Revisión bibliográfica	5
2 Pruebas de No-inferioridad	6
2.1 Introducción	6
2.2 Prueba de Blackwelder	7
2.3 Prueba de Laster-Johnson-Kotler (LJK)	8
2.4 Propiedades de la prueba de LJK	10
3 Nivel de significancia real de la prueba de LJK	14
3.1 Introducción	14

3.2	Obtención de los niveles de significancia reales de la prueba original de LJK	16
3.3	Corrección por continuidad para la prueba de LJK	23
4	Resultados	25
4.1	Factor de corrección C_1	25
4.2	Factor de corrección C_2	28
5	Conclusiones y posibles líneas de investigación futuras	34
5.1	Conclusiones	34
5.2	Posibles líneas de investigación futuras	35
	Apéndices	37
A	Programas de cómputo	37
A.1	Cálculo de los niveles de significancia reales	37
A.2	Programa para verificar si una región crítica cumple con la condición de convexidad de Barnard	40
	Bibliografía	44

Índice de figuras

3.2.1	Niveles de significancia real para la estadísticas T_1 y T_2 para el nivel de significancia nominal $\alpha = 0.025$ y el valor de $R_0 = 0.85$ para los tamaños de muestra $n = 30, 31, \dots, 100$	20
3.2.2	Niveles de significancia real para la estadísticas T_1 y T_2 para el nivel de significancia nominal $\alpha = 0.05$ y el valor de $R_0 = 0.85$ para los tamaños de muestra $n = 30, 31, \dots, 100$	20
3.2.3	Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.01$ y 4 valores de $R_0, 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	21
3.2.4	Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.025$ y 4 valores de $R_0, 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	22
3.2.5	Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.05$ y 4 valores de $R_0, 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	22
3.2.6	Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.10$ y 4 valores de $R_0, 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	23

4.1.1	Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.01$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	26
4.1.2	Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.025$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	27
4.1.3	Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.05$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	27
4.1.4	Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.10$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	28
4.2.1	Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.01$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	29
4.2.2	Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.025$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	29
4.2.3	Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.05$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	30

4.2.4	Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.10$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$	30
-------	---	----

Índice de cuadros

2.4.1	Comparación de tamaños de muestra para las pruebas de no-inferioridad de LJK y Blackwelder para varios valores de $R_V(p_2/p_1)$, R_0 y p_1 con $1-\beta=0.8$ y $\alpha = 0.05$. Los valores correspondientes a Blackwelder aparecen en ().	12
4.2.1	Porcentaje de valores que caen en el intervalo $[\alpha - 0.01, \alpha + 0.01]$ para los cuatro valores considerados de α , para la prueba de LJK, usando los estadísticos T_1 , T_{C_1} y T_{C_2}	32
4.2.2	Porcentaje de valores que caen en el intervalo $[\max\{0, \alpha - 0.02\}, \alpha]$ para los cuatro valores considerados de α , para la prueba de LJK, usando los estadísticos T_1 , T_{C_1} y T_{C_2}	33

Corrección por continuidad para la prueba asintótica de no-inferioridad de Laster-Johnson-Kotler para dos proporciones independientes

Yoni Miguel Castillo Tzec

Colegio de Postgraduados, 2013

Las pruebas asintóticas de no-inferioridad se usan frecuentemente en ensayos clínicos. El criterio del ‘al menos tan bueno como’ fue introducido por Laster *et al.* [17] para datos dicotómicos. En este enfoque (LJK), el margen de no-inferioridad se considera como un porcentaje del tratamiento activo, en lugar de una diferencia fija. Este procedimiento ha mostrado ser más eficiente que el enfoque de una diferencia fija produciendo tamaños de muestra más pequeños. También, el procedimiento presenta varias ventajas en su diseño, eficiencia estadística e interpretación de las pruebas de no-inferioridad. Sin embargo, el procedimiento de LJK tiene la desventaja de que el tamaño de la prueba es usualmente mayor que el nivel de significancia nominal especificado o requerido (α). En este trabajo, mediante el uso de factores de corrección por continuidad, se ha ajustado el procedimiento para evitar este inconveniente.

Palabras Clave No-inferioridad, datos dicotómicos, tamaño de prueba, prueba de Laster-Johnson-Kotler, corrección por continuidad.

**Continuity correction for the Laster-Johnson-Kotler non-inferiority
asymptotic statistical test for two independent proportions**

Yoni Miguel Castillo Tzec

Colegio de Postgraduados, 2013

Non-inferiority asymptotic statistical tests are frequently used in clinical trials. The ‘at least as good’ criterion was introduced by Laster, *et al.* [17] for dichotomous data. In this approach (LJK), the margin of non-inferiority is taken as a percentage of the control response, rather than a fixed difference. The procedure is seen to be more efficient than the fixed margin approach yielding smaller sample sizes. Also, the procedure offers several advantages in the design, statistical efficiency and interpretability of non-inferiority trials. However, the LJK procedure has the disadvantage that its test size is usually much greater than the specified or required nominal significance level (α). In this paper, by using a continuity correction factor, we have adjusted this procedure to solve this drawback.

Key words: Non-inferiority, dichotomous data, test size, Laster-Johnson-Kotler test, continuity correction.

Capítulo 1

Introducción

1.1. Introducción

El ensayo clínico aleatorizado (ECA) es uno de los avances más importantes en el siglo veinte [9],[13] y [21]. Su importancia creció a partir de que en Medicina se estableció como norma demostrar con evidencia la eficacia de los medicamentos, los biológicos y los dispositivos. Históricamente, utilizar al placebo como grupo control fue la estrategia óptima para demostrar la eficacia de un tratamiento experimental. En base a este diseño, por ejemplo, la Secretaría de Salud en México y la Food and Drug Administration (FDA) en Estados Unidos han aprobado la comercialización de muchos medicamentos. En el caso de Estados Unidos, para justificar la aprobación de un tratamiento, la FDA requiere ejecutar dos ECA-s en donde el tratamiento experimental resulta superior al placebo. El mismo rigor se cumple a nivel internacional. La “International Conference on Harmonization (ICH)” ha consolidado procedimientos para el registro de productos farmacéuticos en la Unión Europea, Japón y los Estados Unidos, por ejemplo, la guía ICH-E9 establece principios estadísticos para los ECA-s [15]. La guía ICH-E10 establece la selección de

controles apropiados para los ECA-s [16].

En la ausencia de tratamientos efectivos, los ensayos clínicos controlados por placebo son incontrovertibles. Sin embargo, cuando ya existe un tratamiento eficaz, es cuestionable la ética del ECA controlado por placebo. En este contexto, los ataques en contra del ECA con placebo son muchos y substanciales [5], [11], [12] y [20], siendo la Declaración de Helsinki el de mayor importancia [22]. En el artículo II.3 de esta declaración se menciona: “En cualquier estudio médico, cada paciente – incluyendo aquellos en el grupo control, debe asegurarse de que se le aplique el mejor tratamiento probado existente”.

En la situación de que el placebo sea considerado no ético, esto nos lleva a los ensayos controlados con un tratamiento activo, en los cuales el tratamiento experimental es comparado directamente con un tratamiento control activo probadamente eficaz. Si el patrocinador (propietario de la molécula nueva) piensa que el nuevo tratamiento es superior al tratamiento control activo, entonces el ensayo clínico estándar de superioridad es el indicado. Sin embargo, si de antemano el patrocinador no cree en la superioridad del tratamiento experimental, entonces un ensayo de no-inferioridad puede ser apropiado.

Los ensayos clínicos de no-inferioridad se han convertido en una de las mejores respuestas al ataque de los ECA-s controlados con placebo. En estos ensayos el nuevo tratamiento experimental es comparado con un tratamiento activo y probadamente eficaz.

El objetivo de un estudio de no-inferioridad es probar que un nuevo tratamiento, que es más seguro, más económico o más fácil de aplicar, no es substancialmente inferior en eficacia al tratamiento estándar. Por ejemplo en el caso de tratamientos para SIDA, nuevos productos pueden tener regímenes de administración mucho más sencillos y con mínimos efectos adversos con lo cual se asegura la adherencia al tratamiento, a pesar de que su

eficacia sea muy similar al tratamiento estándar.

Las pruebas de no-inferioridad han sido ampliamente estudiadas por diversos autores, por ejemplo Blackwelder [6], Böhning–Viwatwongkasen [7], Farrington y Maning [8], Al-mendra y Sotres [4], entre otros. La medida de disimilaridad más usada es la diferencia de proporciones. La hipótesis de no-inferioridad originalmente planteada por Blackwelder es la siguiente:

$$H_0 : p_1 - p_2 \geq \delta \quad \text{vs} \quad H_1 : p_1 - p_2 < \delta \quad (1.1.1)$$

donde p_1 y p_2 son las probabilidades de éxito para los tratamientos estándar y experimental, respectivamente; δ representa el llamado margen de no-inferioridad y es una constante positiva.

Sin embargo, aunque existen otros enfoques para plantear la hipótesis de no inferioridad, como la razón de proporciones o razón de momios, éstas han sido menos exploradas que la diferencia de proporciones. En este sentido resulta interesante el enfoque de no-inferioridad presentado por Laster *et al.* [17], que propone una forma alternativa de formular la hipótesis de no-inferioridad, haciéndolo en base a la razón de proporciones.

Específicamente, Laster *et al.* [17] formula la hipótesis de no-inferioridad de la siguiente manera:

$$H_0 : p_2 - R_0 p_1 \leq 0 \quad \text{vs} \quad H_1 : p_2 - R_0 p_1 > 0 \quad (1.1.2)$$

donde R_0 es una constante tal que $0 < R_0 < 1$. Usualmente, R_0 es cercano a 1.

Este enfoque ofrece importantes ventajas sobre la hipótesis clásica de no inferioridad propuesta por Blackwelder. Por ejemplo, para los diseñadores del estudio resulta más sencillo expresar el margen de no-inferioridad en términos de un porcentaje (R_0) del tra-

tamiento estándar que debe ser superado por el nuevo tratamiento, que seleccionar un valor δ arbitrario como diferencia clínica. Otra ventaja es cuando se tienen varias variables de eficacia y solo se requiere establecer un mismo porcentaje (R_0) como margen de no-inferioridad para todas las variables, en lugar de tener que seleccionar de manera un tanto arbitraria diferentes valores de δ para cada una de las variables de eficacia, lo cual apoya la claridad de las conclusiones del estudio, también la estadística de prueba de la estrategia de Laster, Johnson y Kotler (LJK) tiene menor varianza lo que se traduce en menores tamaños de muestra, éstas y otras ventajas de la prueba de LJK pueden verse en Laster *et al.* [17].

La prueba de Laster-Johnson-Kotler (LJK) es asintótica, por lo que el tamaño de la prueba se aleja del valor nominal (α). Así resulta de interés estudiar el comportamiento del nivel de significancia real de la prueba respecto al nivel nominal.

1.2. Objetivos

1. Calcular el nivel de significancia real para la prueba original de LJK y evaluar la diferencia entre este nivel real y el nivel nominal.
2. Proponer factores de corrección por continuidad para la prueba de LJK, para reducir substancialmente la diferencia entre los niveles de significancia real y nominal de la prueba.

1.3. Revisión bibliográfica

Existen diversos artículos que tratan sobre el tema de no inferioridad. Sin embargo, la gran mayoría de estos trabajos están basados en la formulación clásica de la hipótesis de no inferioridad originalmente planteada por Blackwelder [6]. Por ejemplo Almendra y Sotres [2], realizan un estudio donde se comparan diversas pruebas asintóticas de no inferioridad basado en el nivel de significancia real y en sus potencias, incluyendo las pruebas de Blackwelder, Böhning–Viwatwongkasen, Farrington–Maning entre otras. En este trabajo se concluye que para diseños balanceados los tamaños de muestra $30 \leq n \leq 100$, considerando un nivel de significancia nominal $\alpha = 0.05$ y los tres niveles de no inferioridad 0.01, 0.15, 0.20, se recomienda usar la prueba de Farrington–Maning con el factor de corrección $1/(2n)$.

Por otra parte en [1] y [3] se exhibe el mal comportamiento de la prueba de Blackwelder en términos de su tamaño de prueba.

El trabajo más reciente donde se realiza un estudio comparativo entre las pruebas: una basada en la medida de disimilaridad de la diferencia de proporciones y la otra basada en la razón de proporciones, es presentado por Laster *et al.* [17]. En este trabajo se reportan ventajas importantes de la prueba de LJK, basada en la razón de proporciones, sobre la prueba de Blackwelder, ver Laster *et al.* [17].

En la literatura consultada no se encontró ningún trabajo enfocado a estudiar el comportamiento del tamaño de la prueba de LJK respecto al nivel de significancia nominal, o la necesidad de plantear algún factor de corrección para esta prueba. Este es el objeto principal de la presente tesis.

Capítulo 2

Pruebas de No-inferioridad

2.1. Introducción

El modelo estadístico empleado para comparar dos proporciones considera a dos muestras aleatorias independientes de la distribución Bernoulli $\{X_{11}, X_{12}, \dots, X_{1n_1}\}$ y $\{X_{21}, X_{22}, \dots, X_{2n_2}\}$ correspondientes a los tratamientos estándar con probabilidad de éxito p_1 y al tratamiento experimental con probabilidad de éxito p_2 , respectivamente.

Sean $X_1 = \sum_{j=1}^{n_1} X_{1j}$ y $X_2 = \sum_{j=1}^{n_2} X_{2j}$, entonces, se tiene el espacio muestral con base a las estadísticas X_1 y X_2 definido como $\chi = \{(x_1, x_2) \in \{0, \dots, n_1\} \times \{0, \dots, n_2\}\}$. El espacio de parámetros es $\Theta = \{(p_1, p_2) : p_1, p_2 \in [0, 1]\}$.

Nótese que X_i tiene distribución binomial con parámetros dados por (n_i, p_i) para $i = 1, 2$, donde X_1 y X_2 son independientes.

En general la prueba de hipótesis de no inferioridad se formula de la siguiente manera

$$H_0 : p_2 \leq g(p_1) \quad \text{vs} \quad H_1 : p_2 > g(p_1) \quad (2.1.1)$$

donde $g : [0, 1] \rightarrow [0, 1]$ es una función continua y no decreciente, con $g \neq 1$. La hipótesis planteada en (2.1.1) incluye como casos particulares a funciones como la diferencia de proporciones, la razón de proporciones y la razón de momios.

En el presente trabajo se consideran pruebas de hipótesis con regiones críticas de la forma $R_T(\alpha) = \{(x_1, x_2) \in \chi \mid T(x_1, x_2) \leq t\}$ donde t es una constante que depende del nivel de significancia nominal α y $T(X_1, X_2)$ es la estadística de prueba para contrastar la hipótesis en (2.1.1).

Así, la función de verosimilitud conjunta está dada por

$$L(p_1, p_2; x_1, x_2) = \prod_{i=1}^2 \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i} \quad (2.1.2)$$

Sea $R_T = R_T(\alpha)$ la region crítica de una prueba de hipótesis para contrastar la hipótesis en (2.1.1). Entonces la función de potencia para esta prueba se define como:

$$\beta_T(p_1, p_2) = \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} L(p_1, p_2; x_1, x_2) I_{[(x_1, x_2) \in R_T]} \quad (2.1.3)$$

2.2. Prueba de Blackwelder

El estadístico de prueba de Blackwelder (1982) es:

$$T_{BW} = \frac{\hat{p}_1 - \hat{p}_2 - \delta}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

donde \hat{p}_1 y \hat{p}_2 denotan las proporciones observadas, es decir $\hat{p}_1 = \frac{X_1}{n_1}$ y $\hat{p}_2 = \frac{X_2}{n_2}$ con $\delta > 0$ y cercano a 0. La región crítica de Blackwelder es:

$$R_{BW}(\alpha) = \{(x_1, x_2) \in \chi : T_{BW}(x_1, x_2) < -z_\alpha\}$$

donde z_α es el cuantil superior α de la distribución normal estándar, es decir, $P(Z > z_\alpha) = \alpha$; Z es una variable aleatoria con distribución normal estándar.

2.3. Prueba de Laster-Johnson-Kotler (LJK)

A diferencia del enfoque propuesto por Blackwelder, Laster *et al.* proponen una nueva forma de plantear la prueba de hipótesis de no inferioridad como

$$H_0 : p_2 - R_0 p_1 \leq 0 \quad \text{vs} \quad H_1 : p_2 - R_0 p_1 > 0 \quad (2.3.1)$$

donde R_0 es una constante tal que $0 < R_0 < 1$. Usualmente, R_0 es cercano a 1.

La ventaja de escribir la hipótesis de no inferioridad como en (2.3.1), es que la eficacia clínica del tratamiento experimental puede verse como un porcentaje de la terapia estándar, esto es

$$R_V = \frac{p_2}{p_1}$$

es decir, rechazar H_0 implica que el tratamiento experimental es al menos tan bueno como la terapia estándar por más de $R_0 \times 100$ por ciento ($p_2 > R_0 p_1$).

La estadística de prueba de LJK es

$$T_{LJK} = \frac{\hat{p}_2 - R_0 \hat{p}_1}{\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)R_0^2}{n_1}}} \quad (2.3.2)$$

donde $R_0 < 1$. La región crítica de la prueba de la LJK es:

$$R_{LJK}(\alpha) = \{(x_1, x_2) \in \chi : T_{LJK}(x_1, x_2) > z_\alpha\}$$

donde χ y z_α fueron descritos anteriormente.

Existe una relación uno a uno entre las formulaciones de la hipótesis de no-inferioridad establecida por Blackwelder y LJK, como se muestra a continuación. Sea

$$\delta = (1 - R_0)p_1$$

entonces, considerando $H_0 : p_1 - p_2 \geq \delta$, donde $p_1 > 0$, se tiene

$$p_1 - (1 - R_0)p_1 \geq p_2$$

$$p_1 - p_1 + R_0 p_1 \geq p_2$$

$$R_0 p_1 \geq p_2$$

que es la hipótesis nula H_0 planteada en (2.3.1). Análogamente se establece la equivalencia entre las hipótesis alternativas de Blackwelder y LJK.

2.4. Propiedades de la prueba de LJK

Para una amplia gama de valores de p_1 , p_2 , R_0 y α , Laster *et al.* [17], prueban que la prueba de hipótesis propuesta por LJK es más eficiente que la prueba de Blackwelder, esto debido a la desviación estándar más pequeña definida en la prueba de LJK comparada con la respectiva desviación estándar de Blackwelder, como se muestra a continuación. Para el caso de Blackwelder se tiene

$$\begin{aligned} \text{Var}(\hat{p}_1 - \hat{p}_2 - \delta) &= \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2) \\ &= \frac{1}{n_1^2} \text{Var}(X_1) + \frac{1}{n_2^2} \text{Var}(X_2) \\ &= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \end{aligned}$$

por independencia. Para LJK

$$\begin{aligned} \text{Var}(\hat{p}_2 - R_0\hat{p}_1) &= \text{Var}(\hat{p}_2) + R_0^2 \text{Var}(\hat{p}_1) \\ &= \frac{p_2(1-p_2)}{n_2} + R_0^2 \frac{p_1(1-p_1)}{n_1} \end{aligned}$$

Luego, cuando $R_0 < 1$, se obtiene que

$$\text{Var}(\hat{p}_1 - \hat{p}_2 - \delta) > \text{Var}(\hat{p}_2 - R_0\hat{p}_1)$$

Por esta razón, los tamaños de muestra requeridas por grupo para la prueba de LJK son menores respecto a la prueba de Blackwelder. Esto puede verse a continuación. Para la

prueba de Blackwelder los tamaños de muestra por grupo está dado por

$$n_{BW} = \frac{(z_{1-\alpha} - z_\beta)^2 [p_1(1 - p_1) + p_2(1 - p_2)]}{(p_1 - p_2 - \delta)^2} \quad (2.4.1)$$

y para LJK se tiene

$$n_{LJK} = \frac{(z_{1-\alpha} - z_\beta)^2 [p_2(1 - p_2) + p_1(1 - p_1)R_0^2]}{(p_2 - p_1R_0)^2} \quad (2.4.2)$$

Ahora, considerando a δ definido en términos de R_0 y p_1 como

$$\delta = (1 - R_0)p_1$$

el denominador de la expresión (2.4.2) puede ser reescrito como sigue

$$\begin{aligned} [p_2 - p_1R_0]^2 &= [p_2 - p_1(1 - \delta/p_1)]^2 \\ &= [p_2 - p_1 + \delta]^2 \end{aligned}$$

Nótese que $(p_2 - p_1 + \delta)^2 = (p_1 - p_2 - \delta)^2$, por lo que los denominadores de n_{BW} y n_{LJK}

son equivalentes. Así, haciendo la razón n_{LJK}/n_{BW} se tiene que

$$\begin{aligned} \frac{n_{LJK}}{n_{BW}} &= \frac{(z_{1-\alpha} - z_\beta)^2 [p_2(1 - p_2) + p_1(1 - p_1)R_0^2] (p_1 - p_2 - \delta)^2}{(z_{1-\alpha} - z_\beta)^2 [p_1(1 - p_1) + p_2(1 - p_2)] (p_1 - p_2 - \delta)^2} \\ &= \frac{p_2(1 - p_2) + p_1(1 - p_1)R_0^2}{p_1(1 - p_1) + p_2(1 - p_2)} \end{aligned}$$

y dado que $R_0 < 1$, entonces

$$\frac{n_{LJK}}{n_{BW}} < 1$$

es decir, n_{LJK} siempre será más pequeña que n_{BW} . En la Tabla 2.4.1 se muestran algunos ejemplos, comparando los tamaños de muestra por grupo para ambas pruebas.

	R_0	$R_V = p_2/p_1$							
		0.80		0.85		0.90		0.95	
$p_1 = 0.10$	0.50	660	(1124)	506	(847)	403	(664)	331	(537)
	0.75	30721	(40459)	7938	(10373)	3642	(4723)	2111	(2720)
	0.80			33479	(41491)	8625	(10628)	3945	(4835)
	0.85					36335	(42511)	9336	(10880)
	0.90							39290	(43519)
$p_1 = 0.40$	0.50	119	(196)	90	(146)	70	(114)	56	(91)
	0.75	5450	(7073)	1389	(1794)	628	(808)	358	(459)
	0.80			5843	(7178)	1484	(1818)	668	(817)
	0.85					6241	(7271)	1580	(1838)
	0.90							6646	(7351)
$p_1 = 0.80$	0.50	29	(42)	20	(30)	15	(22)	11	(16)
	0.75	1238	(1509)	297	(365)	125	(155)	66	(83)
	0.80			1237	(1459)	294	(349)	122	(147)
	0.85					1226	(1397)	288	(331)
	0.90							1206	(1323)

Tabla 2.4.1. Comparación de tamaños de muestra para las pruebas de no-inferioridad de LJK y Blackwelder para varios valores de $R_V(p_2/p_1)$, R_0 y p_1 con $1-\beta=0.8$ y $\alpha = 0.05$. Los valores correspondientes a Blackwelder aparecen en ().

Además de estas características positivas de la prueba de LJK, también deben recordarse las otras ventajas que tiene esta prueba enumeradas en la sección 1.1.

Por otra parte, la prueba de LJK tiene algunas desventajas:

- La estadística de prueba no está definida en los pares de puntos $(0, 0)$, $(0, n_2)$, $(n_1, 0)$ y (n_1, n_2) .
- Como veremos más adelante el tamaño de la prueba de LJK está muy por arriba del nivel nominal de significancia (α) .

Capítulo 3

Nivel de significancia real de la prueba de LJK

Para realizar el cálculo del nivel de significancia (tamaño) real de las pruebas de no-inferioridad de una forma eficiente, en la literatura se han desarrollado algunos resultados que han logrado reducir significativamente el tiempo de cómputo para la obtención de estos niveles. Estos resultados se presentan para las pruebas de no-inferioridad en general en la sección 3.1, mientras que la aplicación de estos resultados para la prueba de LJK se presenta en la sección 3.2.

3.1. Introducción

Uno de los conceptos necesarios para facilitar la obtención de los niveles de significancia reales, es la llamada condición de convexidad de Barnard.

Definición 3.1.1 Una prueba estadística de no-inferioridad T , para contrastar las hipótesis en (2.1.1), con región de rechazo R_T cumple la condición de convexidad de Barnard si

$$1. (x_1, x_2) \in R_T \Rightarrow (x_1 - 1, x_2) \in R_T \quad \forall \quad 1 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2$$

$$2. (x_1, x_2) \in R_T \Rightarrow (x_1, x_2 + 1) \in R_T \quad \forall \quad 0 \leq x_1 \leq n_1, 0 \leq x_2 \leq n_2 - 1$$

La definición anterior es importante, ya que es pieza clave para el teorema dado por Röhmel y Mansmann en [19]. Este resultado permite calcular el nivel de significancia real de pruebas de no-inferioridad para contrastar las hipótesis en (2.1.1), como un máximo que se alcanza en un punto frontera sobre la curva $g(p_1) = p_2$. Sin embargo, el resultado mencionado es solo aplicable a pruebas exactas. En virtud de que la prueba de LJK es de naturaleza asintótica, es importante considerar la generalización del Teorema de Röhmel y Mansmann que aplica a pruebas tanto exactas como asintóticas. Dicha generalización fue realizada por Almendra y Sotres en [4] y se presenta a continuación.

Teorema 3.1.1 Sea T la estadística de prueba de una prueba de no-inferioridad para el problema en (2.1.1) y con región crítica $R_T(\alpha) = \{(x_1, x_2) : T(x_1, x_2) \leq c_\alpha\}$ y sean $\Theta_0 = \{(p_1, p_2) \in [0, 1]^2 : p_2 \leq g(p_1)\}$, $a \leq p_1 \leq 1$, $0 \leq a$ y $g(a) = 0$ cuando $a > 0$ con g creciente y no intersecta $p_1 = p_2$, supóngase que la región crítica es no vacía y satisface la condición de convexidad de Barnard, 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 la curva $p_2 = g(p_1)$.

La demostración del teorema anterior se puede ver en [4]. Y con el resultado se tiene que si $\Theta_0^* = \{(p_1, p_2) \in [0, 1]^2 : p_2 = g(p_1)\}$ entonces

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

Así por ejemplo, para la prueba de LJK se utiliza la función $g(p_1) = R_0 p_1$, y el nivel de significancia real α^* se obtiene como

$$\alpha^* = \max_{\substack{p_2=R_0 p_1 \\ p_1 \in [0,1]}} \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} \binom{n_1}{x_1} p_1^{x_1} (1-p_1)^{n_1-x_1} \binom{n_2}{x_2} p_2^{x_2} (1-p_2)^{n_2-x_2} I_{[(x_1, x_2) \in R(\alpha)]} \quad (3.1.1)$$

donde $R(\alpha)$ es la región crítica de la prueba de LJK.

3.2. Obtención de los niveles de significancia reales de la prueba original de LJK

Haciendo uso de la expresión (3.1.1) se obtuvieron los niveles de significancia reales de la prueba de LJK. En este trabajo se define nivel de significancia real de la prueba de LJK como la siguiente aproximación del nivel exacto de LJK. La aproximación del nivel de significancia exacto α^* , en la fórmula (3.1.1), se hizo reemplazando el intervalo continuo $[0, 1]$ por el conjunto discreto de puntos $\{0, 0.001, 0.002, \dots, 0.999, 1\}$. Los tamaños de muestra utilizados son $n = 30, \dots, 200$, en este caso se considera un diseño balanceado $n = n_1 = n_2$. Se consideran cuatro distintos valores para R_0 , 0.80, 0.85, 0.90 y 0.95; y además $\alpha = 0.01, 0.025, 0.05$ y 0.10.

La verificación de la condición de convexidad de Barnard para la prueba de LJK se realizó de manera numérica mediante un programa realizado en R, ver anexo A.2.

En la evaluación del estadístico de prueba de LJK dado en (2.3.2) surgen algunos problemas de indeterminación al usar el estimador de máxima verosimilitud para estimar a p_i , $i = 1, 2$, esto es usando $\hat{p}_i = \frac{x_i}{n_i}$. Este conflicto se da específicamente en los puntos

$(x_1 = 0, x_2 = 0)$, $(x_1 = 0, x_2 = n_2)$, $(x_1 = n_1, x_2 = 0)$ y $(x_1 = n_1, x_2 = n_2)$, ya que el denominador del estadístico T_{LJK} se hace cero, por lo que éste queda indeterminado. Para superar esta dificultad, en la literatura se ha sugerido alguna forma intuitiva de redefinir los puntos mencionados de tal manera que se evita el problema de la indeterminación, ver [2].

Siguiendo la idea usada por Almendra y Sotres en [2], el problema de la indeterminación se resuelve recorriendo un valor de 0.01 hacia adelante al valor de 0 y en caso de n_i retrasándolo la misma cantidad, es decir $n_i - 0.01$. Así, aplicando este método intuitivo se define la siguiente función y algunas de sus propiedades.

Observación 3.2.1 Sean $\phi(n, x) = \frac{\frac{x}{n}(1-\frac{x}{n})}{n} = \frac{x(n-x)}{n^3}$, entonces

1. $\phi(n, n - x) = \phi(n, x)$
2. $\hat{\sigma}(x_1, x_2) = \sqrt{\phi(n_2, x_2) + \phi(n_1, x_1)R_0^2}$
3. ϕ es creciente en $n \Leftrightarrow x > \frac{2}{3}n$
4. ϕ es decreciente en $n \Leftrightarrow x < \frac{2}{3}n$
5. ϕ es creciente en $x \Leftrightarrow x < \frac{n}{2}$
6. ϕ es decreciente en $x \Leftrightarrow x > \frac{n}{2}$

Entonces, se redefine el denominador de la estadística T_{LJK} , de la manera siguiente

$$\hat{\sigma}(x_1, x_2) = \begin{cases} \sqrt{\phi(n_2, 0.01) + \phi(n_1, 0.01)R_0^2} & x_1 = 0, x_2 = 0 \\ \sqrt{\phi(n_2, n_2 - 0.01) + \phi(n_1, 0.01)R_0^2} & x_1 = 0, x_2 = n_2 \\ \sqrt{\phi(n_2, 0.01) + \phi(n_1, n_1 - 0.01)R_0^2} & x_1 = n_1, x_2 = 0 \\ \sqrt{\phi(n_2, n_2 - 0.01) + \phi(n_1, n_1 - 0.01)R_0^2} & x_1 = n_1, x_2 = n_2 \end{cases}$$

Con lo anterior, se tiene que $\hat{\sigma}(x_1, x_2) \neq 0$ en los pares de puntos considerados. Además, debido a la observación 3.2.1(1) se tiene que $\hat{\sigma}(0, 0) = \hat{\sigma}(0, n_2) = \hat{\sigma}(n_1, 0) = \hat{\sigma}(n_1, n_2)$.

Así, con los puntos redefinidos, el estadístico de prueba T_{LJK} queda definido como sigue

$$T_1 = \frac{\hat{p}_2 - \hat{p}_1 R_0}{\hat{\sigma}(x_1, x_2)} \quad (3.2.1)$$

donde p_i , $i = 1, 2$, son los estimadores de MV dados por $\frac{X_i}{n_i}$.

Otra forma de sortear este problema de indeterminación, es utilizar otro tipo de estimadores distintos al estimador de máxima verosimilitud las cuales no presentan el problema de indeterminación en el estadístico T_{LJK} , tales como son el estimador propuesto por Böhning y Viwatwongkasen en [7] y un tipo de estimadores obtenidos bajo la hipótesis nula propuesto por Miettinen y Nurminen en [18].

El estimador sugerido por Böhning y Viwatwongkasen está dado por $\tilde{p}_b = \frac{X+b}{n+2b}$, con $b \geq 0$. Aunque han sido sugeridas varias elecciones para b , $1/6$, $1/2$, 1 o 2 , para este trabajo ha sido considerado $b = 1$, que tiene una justificación bayesiana. El estimador $\tilde{p} = \frac{X+1}{n+2}$, es el estimador de Bayes de p cuando se toma como distribución *a priori* a la distribución

Beta(1, 1). Nótese que con este estimador no se tiene el problema de $\hat{\sigma} = 0$, por lo que el estadístico T_{LJK} no tiene problemas de indeterminación. Así, la estadística de prueba para la prueba de hipótesis planteada en (2.3.1) queda como sigue

$$T_2 = \frac{\hat{p}_2 - \hat{p}_1 R_0}{\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)R_0^2}{n_1}}} \quad (3.2.2)$$

Dado los dos estadísticos definidos en (3.2.1) y (3.2.2), se tiene que para un nivel de significancia nominal α , se rechaza $H_0 : p_2 - R_0 p_1 \leq 0$ a favor de $H_1 : p_2 - R_0 p_1 > 0$ si y sólo si $T_i > z_\alpha$. Por lo que la región de rechazo quedaría de la forma $R_T(\alpha) = \{(x_1, x_2) \in \chi : T_i > z_\alpha\}$ donde z_α es el percentil superior α de la distribución normal estándar, esto es $P(Z > z_\alpha) = \alpha$, para $i = 1, 2$; Z es una variable aleatoria con distribución normal estándar.

Se sabe que la las pruebas asintóticas son aproximaciones que van mejorando en su precisión a medida que el tamaño de la muestra va aumentando. Por lo que, debido a la naturaleza asintótica de la prueba de LJK considerada en este trabajo, el comportamiento de la prueba para muestras no tan grandes, por ejemplo tamaños de muestra ≤ 100 , son bastante liberales ya que exceden considerablemente el nivel nominal de la prueba. Como ejemplo, en las Figuras 3.2.1 y 3.2.2 se presentan los niveles de significancia reales para las estadísticas de prueba T_1 y T_2 considerando a $R = 0.85$ para el nivel de significancia nominal $\alpha = 0.025$ y 0.05 se consideran diseños balanceados, $n_1 = n_2 = n$. Donde T_1 corresponde a la estadística de prueba de LJK utilizando la redefinición en la observación 3.2.1, mientras que T_2 corresponde a la estadística de prueba de LJK utilizando la redefinición con los estimadores de Bayes: $\tilde{p} = (X + 1)/(n + 2)$.

3.2 Obtención de los niveles de significancia reales de la prueba original de LJK

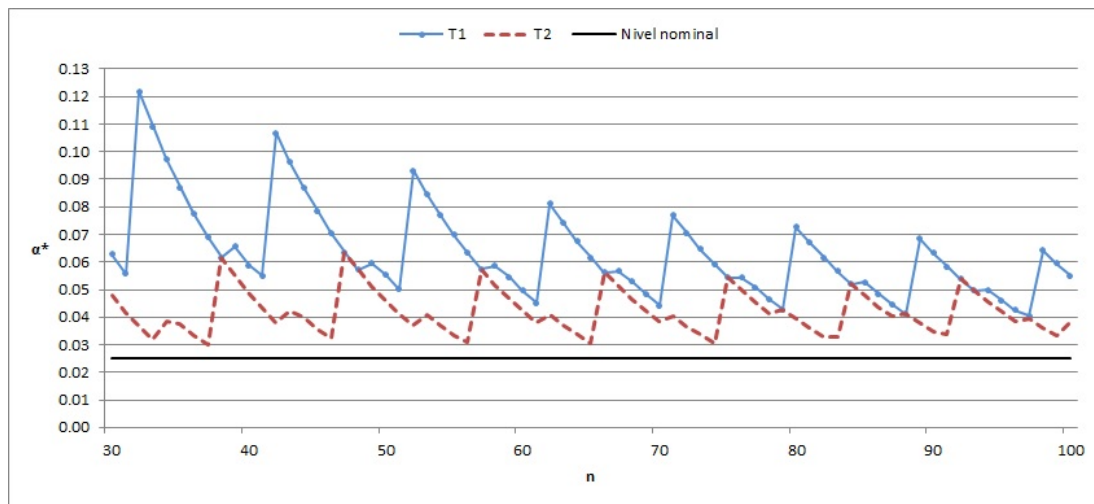


Figura 3.2.1. Niveles de significancia real para la estadísticas T_1 y T_2 para el nivel de significancia nominal $\alpha = 0.025$ y el valor de $R_0 = 0.85$ para los tamaños de muestra $n = 30, 31, \dots, 100$.

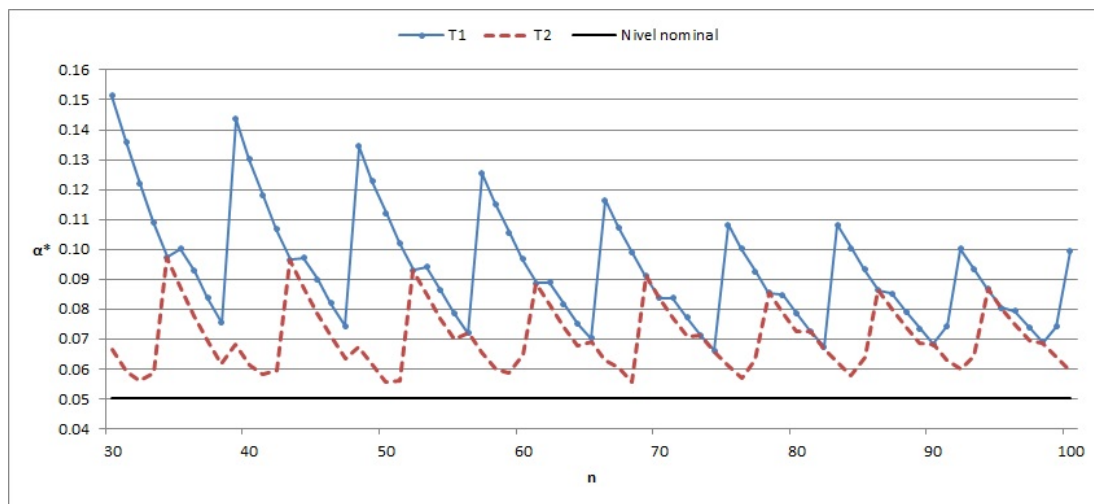


Figura 3.2.2. Niveles de significancia real para la estadísticas T_1 y T_2 para el nivel de significancia nominal $\alpha = 0.05$ y el valor de $R_0 = 0.85$ para los tamaños de muestra $n = 30, 31, \dots, 100$.

Como se observa en las figuras 3.2.1 y 3.2.2 las pruebas T_1 y T_2 (prueba de LJK con dos diferentes redefiniciones de su denominador) son bastantes liberales, es decir, exceden considerablemente al nivel nominal α .

En las figuras 3.2.3-3.2.6 se tienen los niveles de significancia reales obtenidos para la estadística T_1 , considerando 4 valores de R_0 , 0.80, 0.85, 0.90 y 0.95. Se considera un diseño balanceado con $n_1 = n_2 = n$ y los niveles de significancia nominal $\alpha = 0.01, 0.025, 0.05$ y 0.10. En ellas se pueden observar el comportamiento errático del nivel de significancia de la prueba de LJK (T_1) para los cuatros distintos valores de R_0 , ya que la totalidad de los puntos obtenidos están arriba de los respectivos niveles nominales α , 0.01, 0.025, 0.05 y 0.10. Esto crea la necesidad del uso de algún factor de corrección por continuidad para ayudar a solucionar esta irregularidad.

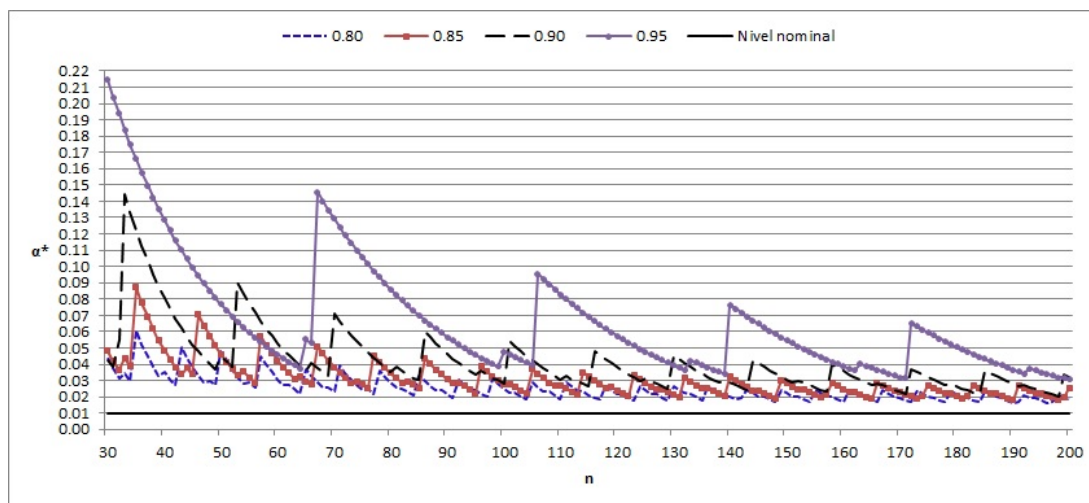


Figura 3.2.3. Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.01$ y 4 valores de R_0 , 0.80, 0.85, 0.90 y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

3.2 Obtención de los niveles de significancia reales de la prueba original de LJK

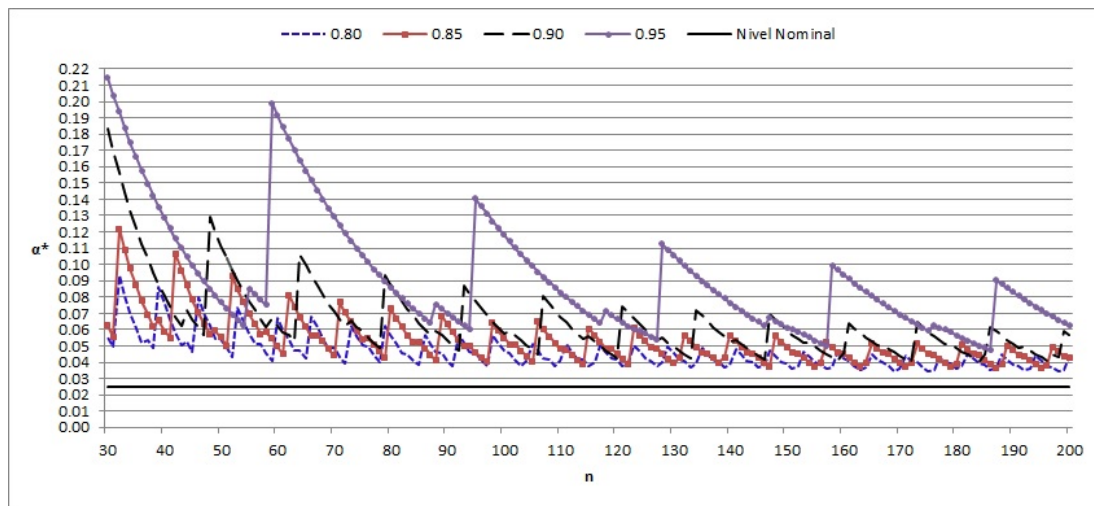


Figura 3.2.4. Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.025$ y 4 valores de R_0 , 0.80, 0.85, 0.90 y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

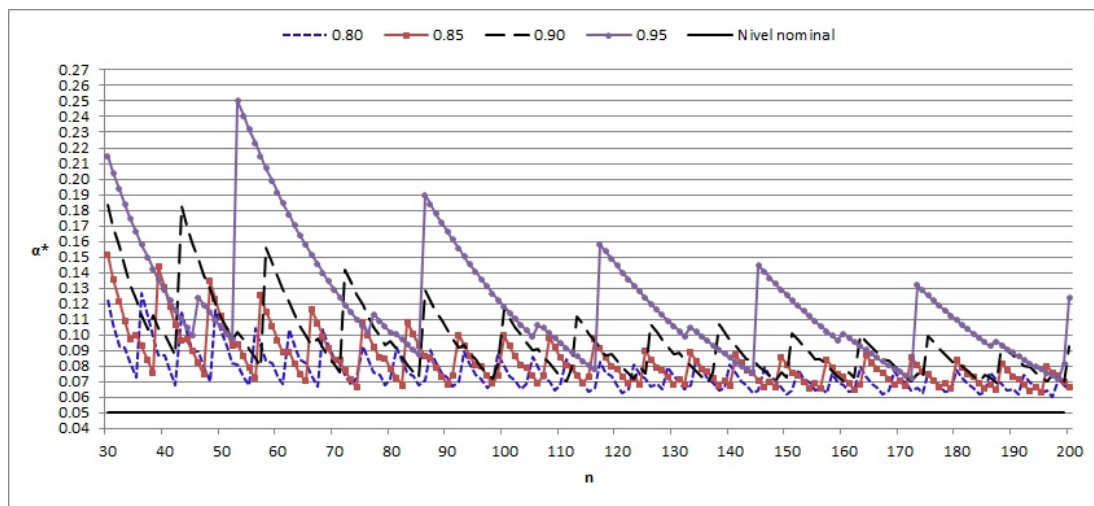


Figura 3.2.5. Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.05$ y 4 valores de R_0 , 0.80, 0.85, 0.90 y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

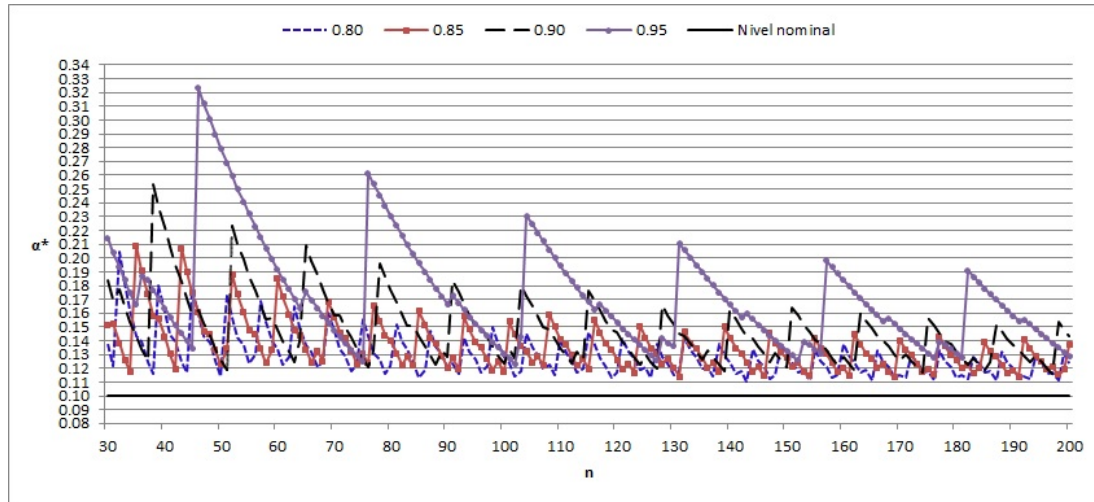


Figura 3.2.6. Niveles de significancia real para la estadística T_1 para el nivel de significancia nominal $\alpha = 0.10$ y 4 valores de R_0 , 0.80, 0.85, 0.90 y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

3.3. Corrección por continuidad para la prueba de LJK

En vista del comportamiento errático del nivel de significancia de la prueba de LJK, se decidió considerar algún factor de corrección por continuidad para la solución de esa irregularidad, ya que frecuentemente los procedimientos aproximados que usan factor de corrección por continuidad logran mejorarse considerablemente, en el sentido de que mantienen el nivel de significancia real muy cercano al nivel de significancia nominal, ver [2]. Debido a lo anterior, en este trabajo se propone el uso de los siguientes dos factores de corrección por continuidad:

$$C_1 = \frac{2}{3n} \quad \text{y} \quad C_2 = \frac{1}{n}$$

El factor de corrección C_1 se consideró en este trabajo, dado que ha mostrado ser un elemento sustancial en la mejora de los niveles de significancia reales de varias pruebas asintóticas, ver por ejemplo Almendra y Sotres [2]. El factor de corrección C_2 , es el factor de corrección por continuidad de Yates (ver Fleiss *et. al.* [10], sección 3.3). Dicho factor está dado por

$$C_Y = \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

que para el caso de diseños balanceados, es decir, para $n_1 = n_2 = n$, se tiene que $C_Y = \frac{1}{n}$. En este trabajo, a C_Y se ha denotado por C_2 . El factor de corrección de Yates, además de ser una elección común, Hauck y Anderson [14] recomiendan el uso de C_2 ya que ha probado ser adecuado en la construcción de intervalos de confianza.

Debido a lo anterior, se han considerado dos tipos de correcciones por continuidad a la prueba de LJK, cuyas estadísticas de prueba se denotarán de la siguiente forma:

$$T_{C_i} = \frac{\hat{p}_2 - \hat{p}_1 R_0 - C_i}{\tilde{\sigma}} \quad (3.3.1)$$

para $i = 1, 2$, y donde C_1 y C_2 se definieron arriba y el denominador $\tilde{\sigma}$ queda como el denominador de la estadística definida en (3.2.2). Nótese que este denominador en (3.2.2) siempre es diferente de cero y por lo tanto la estadística en (3.3.1) no presenta problemas de indeterminación.

Capítulo 4

Resultados

En este capítulo se presentan los niveles de significancia reales obtenidos para la prueba estadística de LJK utilizando los dos factores de corrección definidos en (3.3.1).

Se dirá que los niveles de significancia reales de la prueba de LJK son “aceptables” cuando un alto porcentaje, del total de los valores obtenidos para los diferentes tamaños de muestra considerados, pertenezcan a un intervalo razonable alrededor del valor nominal α , por ejemplo, el intervalo cerrado $[0.04, 0.06]$ para el caso $\alpha = 0.05$. Esto es, de manera general se considera el intervalo cerrado $[\alpha - 0.01, \alpha + 0.01]$ para los cuatro valores considerados de α . De esta manera se han evaluado los dos factores de corrección C_1 y C_2 .

4.1. Factor de corrección C_1

Con base en la fórmula (3.1.1) derivada del teorema 3.1.1, se calcularon los niveles de significancia reales de la prueba de LJK usando el estadístico T_{C_i} dado en la ecuación (3.3.1), con $i = 1$, esto es

$$T_{C_1} = \frac{\hat{p}_2 - \hat{p}_1 R_0 - \frac{2}{3n}}{\tilde{\sigma}} \quad (4.1.1)$$

donde $\tilde{\sigma}$ es el denominador de la expresión (3.2.2). Los resultados se muestran a continuación en las figuras 4.1.1 a 4.1.4.

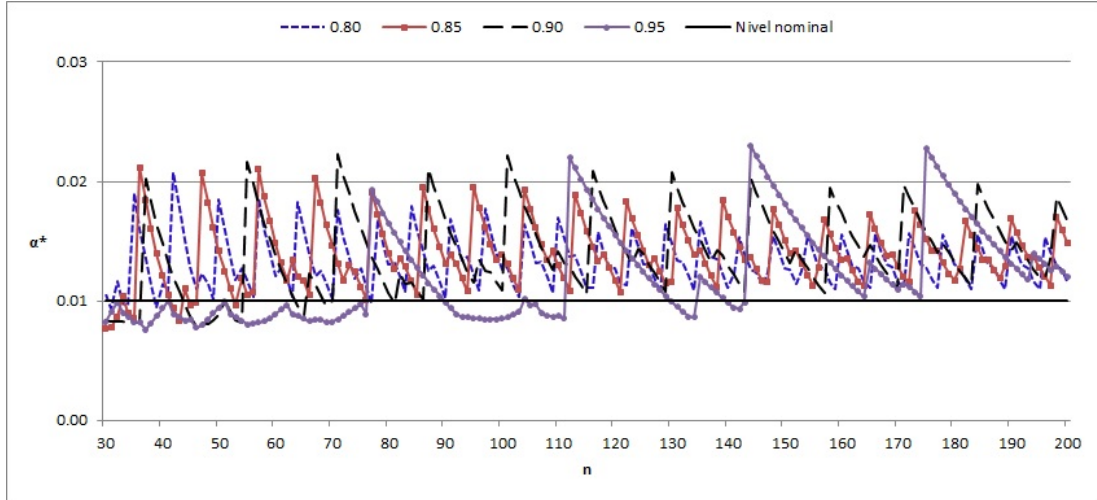


Figura 4.1.1. Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.01$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

El uso del factor de corrección C_1 muestra una mejora sustancial en los niveles de significancia real de la prueba de LJK. Como se observan en las figuras 4.1.1 a 4.1.4, se tiene que la mayoría de los valores obtenidos están en el intervalo cerrado $[\alpha - 0.01, \alpha + 0.01]$ para los cuatro valores de α , 0.01, 0.025, 0.05 y 0.10; y con los cuatro valores considerados de R_0 . De esta manera se ha logrado un mejor comportamiento de la prueba de LJK con respecto a los niveles de significancia reales dado que se ha logrado llevar a dichos niveles en un intervalo estrecho que contiene al nivel nominal α .

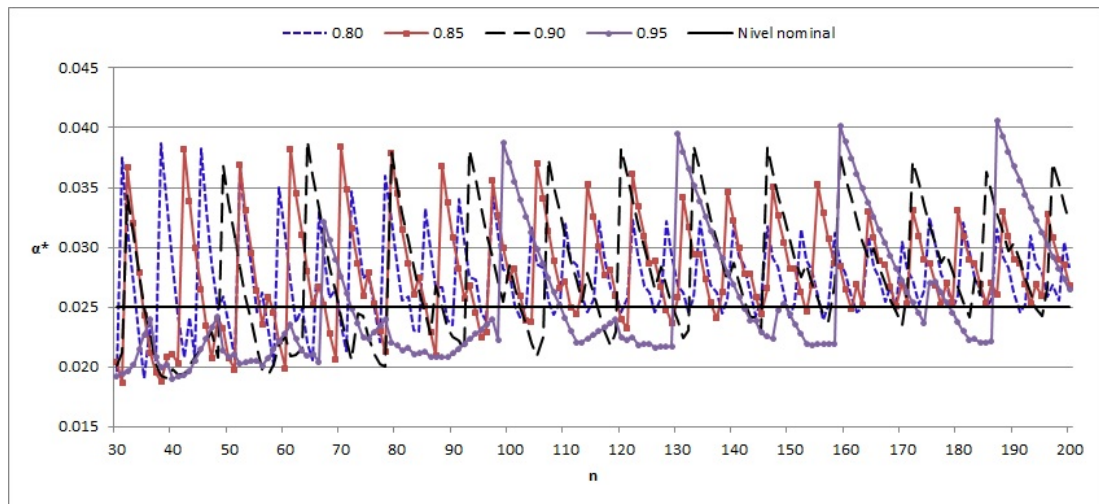


Figura 4.1.2. Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.025$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

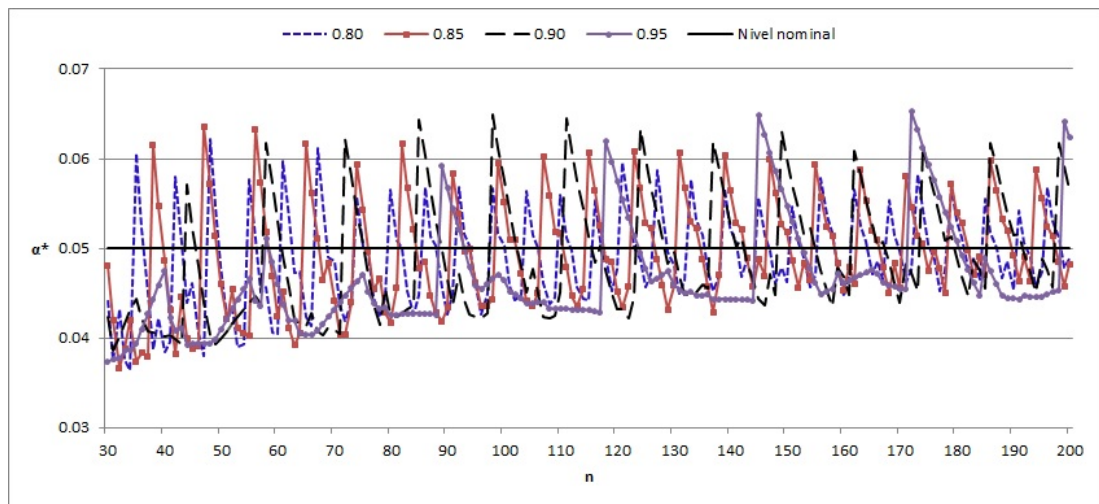


Figura 4.1.3. Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.05$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

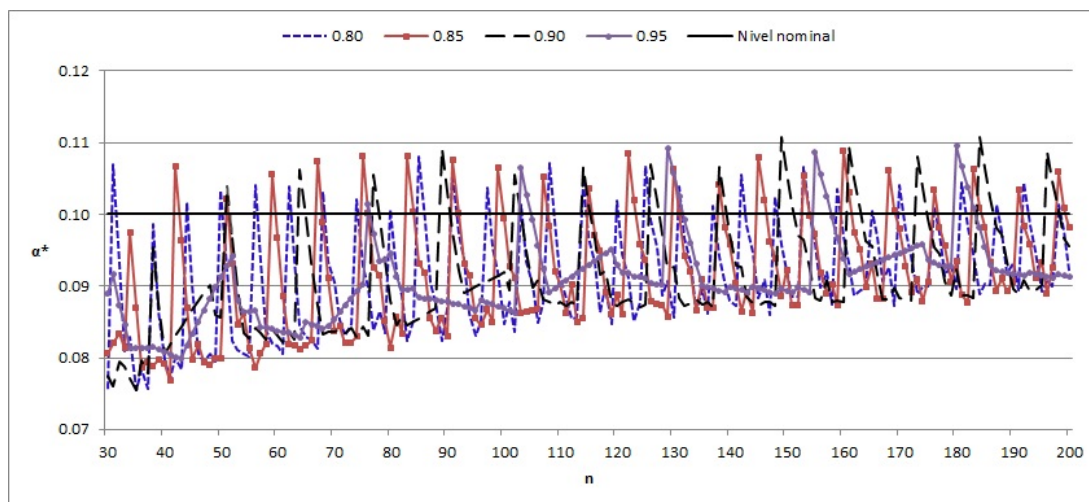


Figura 4.1.4. Niveles de significancia real para la estadísticas T_{C_1} para el nivel de significancia nominal $\alpha = 0.10$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

4.2. Factor de corrección C_2

En forma análoga al análisis realizado en la sección anterior, en esta sección se analiza el comportamiento del tamaño real de la prueba de LJK considerando al factor de corrección C_2 en la expresión (3.3.1), es decir,

$$T_{C_2} = \frac{\hat{p}_2 - \hat{p}_1 R_0 - \frac{1}{n}}{\tilde{\sigma}} \quad (4.2.1)$$

los resultados obtenidos se muestran en las figuras 4.2.1 a 4.2.4.

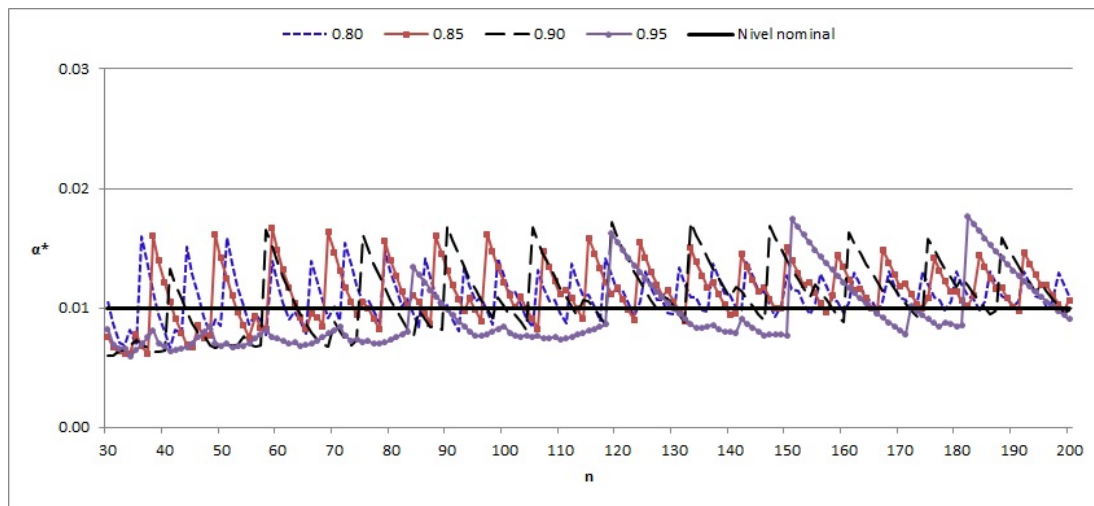


Figura 4.2.1. Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.01$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

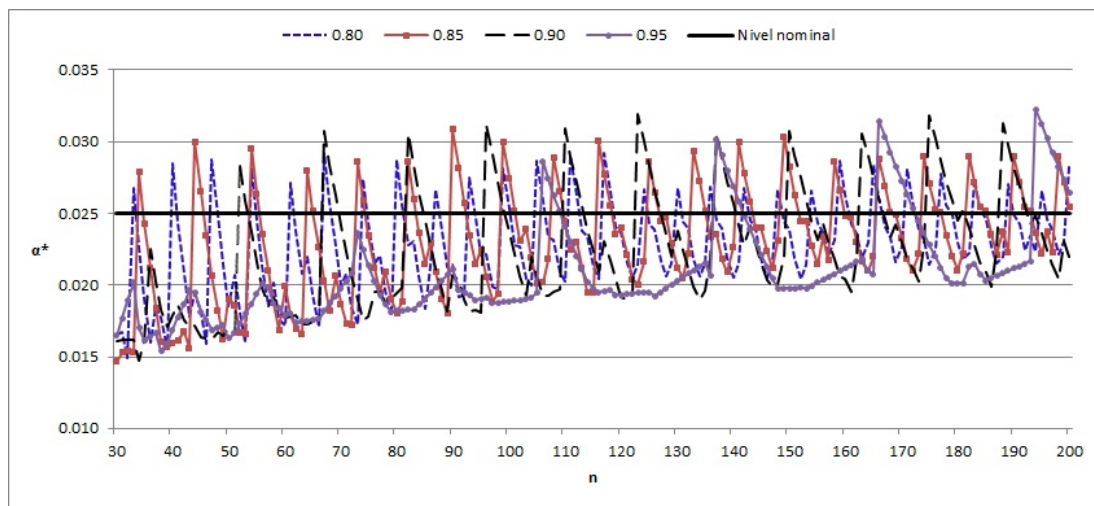


Figura 4.2.2. Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.025$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

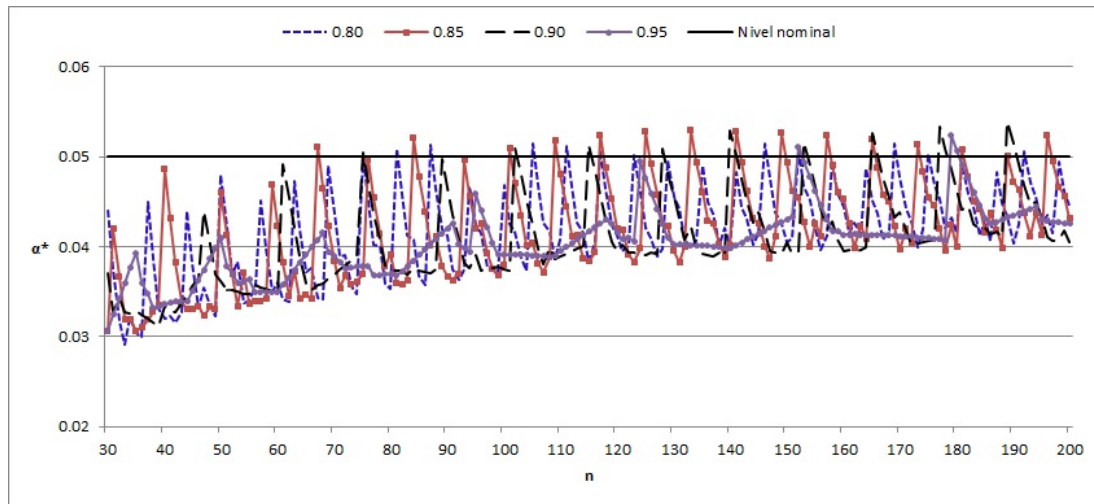


Figura 4.2.3. Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.05$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

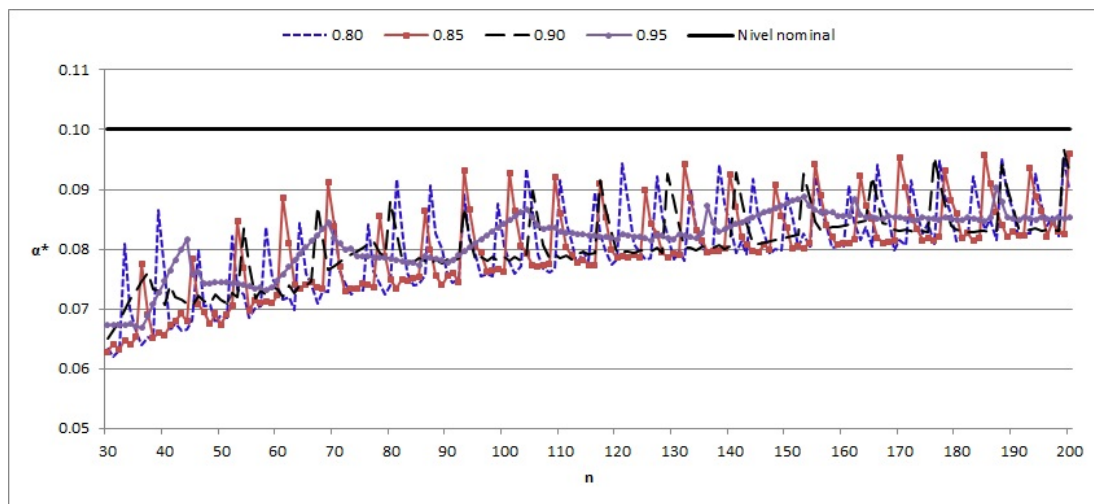


Figura 4.2.4. Niveles de significancia real para la estadísticas T_{C_2} para el nivel de significancia nominal $\alpha = 0.10$ y los valores de $R_0 = 0.80, 0.85, 0.90$ y 0.95 para los tamaños de muestra $n = 30, 31, \dots, 200$.

Las figuras 4.2.1 a 4.2.4 nos muestran que con el uso de C_2 se obtiene una mejora sustancial en los niveles de significancia real de la prueba ya que considerando un intervalo razonable por ejemplo $[\alpha - 0.01, \alpha + 0.01]$ para los valores considerados de α (a excepción de $\alpha = 0.10$) se tiene que la mayoría de los valores obtenidos pertenecen a dicho intervalo, para los cuatro valores de R_0 . Ahora, en el caso de $\alpha = 0.10$ los niveles de significancia reales obtenidos que pertenecen al intervalo $[0.09, 0.11]$ son menos del 10% para los cuatro valores de R_0 . Sin embargo, es de notarse en la figura 4.2.4 que todos los valores de α^* obtenidos están por debajo del nivel nominal $\alpha = 0.10$ lo que convierte a la prueba de LJK en conservadora para este caso. Esto último, también sucede con los resultados obtenidos para el nivel nominal $\alpha = 0.05$, como se puede observar en la figura 4.2.3, la mayoría de los valores obtenidos están por debajo del nivel nominal 0.05.

Los resultados obtenidos para la prueba de LJK sin factor de corrección y con el uso de los factores de corrección por continuidad C_1 y C_2 se resumen en la tabla 4.2.1.

Los porcentajes reportados en la tabla 4.2.1 se obtuvieron calculando el número de tamaños reales de LJK que pertenecen al intervalo $[\alpha - 0.01, \alpha + 0.01]$ dividido por 171, ya que el número total de niveles de significancia reales calculados para los tamaños de muestra considerados $n = 30, 31, \dots, 200$, es igual a 171.

Para los niveles de significancia nominal $\alpha = 0.01$ y 0.025 los factores de corrección C_1 y C_2 reportan en general buenos resultados para los valores de R_0 considerados, ya que como se observa en la tabla 4.2.1, los niveles de significancia reales obtenidos que pertenecen al intervalo considerado están por arriba del 90%, a excepción del porcentaje obtenido para $\alpha = 0.025$ con $R_0 = 0.90$, igual a 87%. El factor de corrección C_2 reporta una ligera ventaja respecto al uso de C_1 para estos dos niveles de significancia.

α	Intervalo	Estadístico	R_0			
			0.80	0.85	0.90	0.95
0.01	[0,0.02]	T_1	28.07	7.02	0.00	0.00
		T_{C_1}	99.42	97.66	94.15	93.57
		T_{C_2}	100.0	100.0	100.0	100.0
0.025	[0.015,0.035]	T_1	2.34	0.00	0.00	0.00
		T_{C_1}	96.49	92.40	87.13	90.64
		T_{C_2}	99.42	99.42	99.42	100.0
0.05	[0.04,0.06]	T_1	0.00	0.00	0.00	0.00
		T_{C_1}	92.98	88.30	87.13	87.72
		T_{C_2}	67.84	60.23	47.37	57.89
0.10	[0.09,0.11]	T_1	0.00	0.00	0.00	0.00
		T_{C_1}	53.80	51.46	40.35	51.46
		T_{C_2}	9.94	9.94	6.43	0.58

Tabla 4.2.1. Porcentaje de valores que caen en el intervalo $[\alpha - 0.01, \alpha + 0.01]$ para los cuatro valores considerados de α , para la prueba de LJK, usando los estadísticos T_1 , T_{C_1} y T_{C_2} .

Por otro lado, para los niveles nominal $\alpha = 0.05$ y 0.10 el factor de corrección con el que mejores resultados se obtiene es el factor C_1 , ya que para los valores 0.80 , 0.85 , 0.90 y 0.95 correspondientes a R_0 se tiene un alto porcentaje de valores que caen en el intervalo considerado, entre 87% y 92% para $\alpha = 0.05$ y entre 40% y 53% para $\alpha = 0.10$.

Como se mencionó anteriormente, los resultados obtenidos para el caso del factor de corrección C_2 con los niveles de significancia nominal 0.05 y 0.10 convierten a la prueba de LJK a tipo conservadora. Considerando este enfoque de lograr que la prueba de LJK sea conservadora, ahora se toma el intervalo $[\max\{0, \alpha - 0.02\}, \alpha]$. En este contexto el factor de corrección C_2 es con el que se obtiene mejores resultados para los valores de $\alpha = 0.01$, 0.025 y 0.05 , siendo el nivel nominal 0.05 el que reporta mejores resultados ya que entre 91% y 98% de los valores obtenidos caen en el intervalo $[0.03, 0.05]$. Los resultados de esta

nueva clasificación se muestra en la tabla 4.2.2.

Los porcentajes reportados en la tabla 4.2.2 se obtuvieron calculando el número de tamaños reales de LJK que pertenecen al intervalo $[max\{0, \alpha - 0.02\}, \alpha]$ dividido por 171, ya que el número total de niveles de significancia reales calculados para los tamaños de muestra considerados $n = 30, 31, \dots, 200$, es igual a 171.

α	Intervalo	Estadístico	R_0			
			0.80	0.85	0.90	0.95
0.01	[0,0.01]	T_1	0.00	0.00	0.00	0.00
		T_{C_1}	2.92	5.85	12.28	44.44
		T_{C_2}	30.41	30.99	40.35	71.93
0.025	[0.005,0.025]	T_1	0.00	0.00	0.00	0.00
		T_{C_1}	26.32	22.81	36.26	64.33
		T_{C_2}	77.19	69.01	74.85	86.55
0.05	[0.03,0.05]	T_1	0.00	0.00	0.00	0.00
		T_{C_1}	59.06	60.23	63.74	82.46
		T_{C_2}	91.81	91.23	92.40	98.25
0.10	[0.08,0.10]	T_1	0.00	0.00	0.00	0.00
		T_{C_1}	77.19	75.44	79.53	92.40
		T_{C_2}	51.46	49.12	50.88	69.01

Tabla 4.2.2. Porcentaje de valores que caen en el intervalo $[max\{0, \alpha - 0.02\}, \alpha]$ para los cuatro valores considerados de α , para la prueba de LJK, usando los estadísticos T_1 , T_{C_1} y T_{C_2} .

Capítulo 5

Conclusiones y posibles líneas de investigación futuras

A partir de los resultados obtenidos en este trabajo, a continuación se presentan las siguientes conclusiones.

5.1. Conclusiones

- Para la realización de un estudio, la prueba de LJK requiere menores tamaños de muestra que la prueba clásica de no-inferioridad de Blackwelder. Además, para los planeadores de los estudios de no-inferioridad y para la interpretación de las conclusiones del estudio, la formulación de hipótesis de LJK resulta mucho más conveniente y práctica que el enfoque clásico de Blackwelder, ver [17].
- La prueba original de LJK es demasiado liberal, es decir, el tamaño real de la prueba excede considerablemente al nivel nominal α .
- El estadístico de la prueba original de LJK dado en [17] presenta el inconveniente de que para algunos puntos la prueba queda indeterminada. La redefinición considerada en este trabajo da solución al problema, ver (3.2.2).

- El uso del estimador propuesto en (3.2.2) no genera problemas de indeterminación, lo cual hace posible el cálculo de los niveles de significancia reales de prueba de LJK.
- La prueba de LJK corregida por el factor de corrección C_2 produce los mejores resultados si nos apegamos al criterio de que los niveles de significancia reales estén cercanos (en el intervalo $[\alpha - 0.01, \alpha + 0.01]$) al valor nominal seleccionado, para los casos de $\alpha = 0.01$ y $\alpha = 0.025$. El factor de corrección C_1 produce los mejores resultados, bajo el mismo contexto, para los niveles nominal 0.05 y 0.10.
- Con la finalidad de que la prueba de LJK sea una prueba de tipo conservadora, el factor de corrección de Yates (C_2) es la que reporta mejores resultados al mantener la gran mayoría de los niveles de significancia reales por debajo en el intervalo $[\max\{0, \alpha - 0.02\}, \alpha]$ del nivel nominal seleccionado, 0.01, 0.025 y 0.05. El factor de corrección C_1 tiene mejores resultados al seleccionar el nivel nominal $\alpha = 0.10$.

5.2. Posibles líneas de investigación futuras

Algunas posibles líneas de investigación que se pueden desprender de este trabajo se enlistan a continuación.

- Factor de corrección para la prueba asintótica de no-inferioridad usando como medida de disimilaridad a la razón de momios.
- Factor de corrección para las pruebas asintóticas de no-inferioridad para diseños desbalanceados.

- Estudiar la prueba de LJK usando los estimadores de máxima verosimilitud restringida a la hipótesis nula.

Apéndice A

Programas de cómputo

A.1. Cálculo de los niveles de significancia reales

```
#-----  
# Función que obtiene el estadístico de LJK MV  
LJ.MV <- function(x1, x2, n1, n2, R)  
{  
  p1 <-x1/n1  
  p2 <-x2/n2  
  
  if ((x1 == 0 && x2 == 0 ) || (x1 == 0 && x2 == n2) ||  
      (x1 == n1 && x2 == 0) || (x1 == n1 && x2 == n2))  
    s <- sqrt((n1-1)*0.01/n1 + (n2-1)*0.01/n2)  
  
  else s <- sqrt(p2*(1-p2)/n2 + p1*(1-p1)*(R^2)/n1)
```

```
t <- (p2-p1*R)/s
return(t)
}

#-----
# Función que obtiene la matriz de estadísticos.
m.est<- function(n1, n2, R, prueba)
{
  matriz <- matrix(0, ncol=3, nrow=(n1+1)*(n2+1))
  k <- 1
  for(i in 0:n1)
  {
    for(j in 0:n2)
    {
      matriz[k,3] <- prueba(i, j, n1, n2, R)
      matriz[k,1] <- i
      matriz[k,2] <- j
      k <- k + 1
    }
  }
  o.matriz <- matriz[order(matriz[,3],decreasing = TRUE),]
  return(o.matriz)
}
```

```
#-----  
# Función que obtiene el alfa real (nivel nominal 0.05)  
sig.real <- function(n1, n2, R, matriz, prob)  
{  
  z <- qnorm(1-0.05)  
  suma <- rep(0, length(prob))  
  for(i in 1:length(prob))  
  {  
    j <- 1  
    while(matriz[j,3] > z)  
    {  
      suma[i] <- suma[i] + dbinom(matriz[j,1],n1,prob[i])  
      *dbinom(matriz[j,2],n2,R*prob[i])  
      j <- j + 1  
    }  
  }  
  return(max(suma))  
}  
  
#-----  
R <- 0.95
```

```
p <- seq(0,1,0.001)
res <-c()
j <- 1

for(i in 30:200)
{
  MT <- m.est(i, i, R, LJ.MV)
  res[j] <- sig.real(i, i, R, MT, p)
  j <- j + 1
}
```

A.2. Programa para verificar si una región crítica cumple con la condición de convexidad de Barnard

```
#-----
#Función que obtiene el estadístico de LJK
LJ.B <- function(x1, x2, n1, n2, R)
{
  p1 <-x1/n1
  p2 <-x2/n2
  p1.b <- (x1+1)/(n1+2)
  p2.b <- (x2+1)/(n2+2)
  s <- sqrt(p2.b*(1-p2.b)/n2 + p1.b*(1-p1.b)*(R^2)/n1)
```

```
t <- (p2-p1*R-(1/n1))/s
return(t)
}

#-----
# Función que obtiene la matriz de estadísticos.
m.est<- function(n1, n2, R, prueba)
{
  matriz <- matrix(0, ncol=n1+1, nrow=n2+1)
  for(i in 0:n1)
  {
    for(j in 0:n2)
    {
      matriz[i+1,j+1] <- prueba(i, j, n1, n2, R)
    }
  }
  return(matriz)
}

#-----
#Función que verifica las condiciones de Barnard
Barnard <- function(matriz, n1, n2, alfa)
{
```

```
z <- qnorm(1-alfa,0,1)
# VERIFICACIÓN DE LA CONDICIÓN (a)
j <- 0
while (j < n2 + 1)
{
  i <- 1
  while(i < n1 + 1)
  {
    if( matriz[i+1,j+1] > z )
    {
      if( matriz[i, j+1] <= z)
      {
        print("no se cumple (a)")
      }
    }
    i <- i+1
  }
  j <- j+1
}
# VERIFICACIÓN DE LA CONDICIÓN (b)
i <- 0
while(i < n1+1)
{
```

```
j <- 0
while(j < n2)
{
  if( matriz[i+1,j+1] > z)
  {
    if( matriz[i+1,j+2] <= z)
    {
      print("no se cumple (b)")
      print(i+1)
    }
  }
  j <- j+1
}
i <- i+1
}

#-----
for(m in 10:200)
{
  MT <- m.est(m, m, 0.8, LJ.B)
  Barnard(MT, m, m, 0.05)
}
```

Bibliografía

- [1] Almendra, A. F. (2009). A study of the classical asymptotic noninferiority test for two binomial proportions. *Drug Information Journal*. **43(5)**: 567-572.
- [2] Almendra, A. F. y Sotres, R. D. (2009). Comparación de algunas pruebas estadísticas asintóticas de no-inferioridad para dos proporciones independientes. *Agrociencia*. **46**: 163-172.
- [3] Almendra, A. F., Castro A. J. J. and Reyes, C. H. (2013). Convergence of test sizes for the Blackwelder's non-inferiority test. *Advances and applications in statistics*. **32(1)**: 15-31.
- [4] Almendra, A. F. and Sotres, R. D. (2012). Some properties of non inferiority tests for two independent probabilities. *Communications in Statistics - Theory and Methods*. **41**:1636-1646.
- [5] Angell, M. (1997). The ethics of clinical research in the third world (editorial). *New England Journal of medicine*. **337**: 847-849.

- [6] Blackwelder, W. C. (1982). "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials*. **3**: 345-353.
- [7] Böhning, D. and C. Viwatwongkasen (2005). Revisiting proportion estimators. *Statistical Methods in Medical Research*. **14**: 1-23.
- [8] 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. *Controlled Clinical Trials*. **9**:1447-1454.
- [9] Fisher, L.D. (1999). Advances in clinical trials in the twentieth century. *Annual Review of Public Health*. **20**: 109-124.
- [10] Fleiss, J. L., Levin, B. and Paik M. C. (2003). *Statistical Methods for rates and proportions*. Third Edition. John Wiley & Sons, Inc.
- [11] Freedman, B., Weijer, C., Glass, K.C.(1996). Placebo orthodoxy in clinical research I: Empirical and methodological myths. *Journal of Law and Medical Ethics*. **24**: **243-251**.
- [12] Freedman, B., Weijer, C., Glass, K.C.(1996). Placebo orthodoxy in clinical research II: Ethical, legal, and regulatory myths. *Journal of Law and Medical Ethics*. **24**: 252-259
- [13] Harrington, D.P. (2000). The randomized clinical trial. *Journal of the American Statistical Association*. **95**: 312-315.

-
- [14] Hauck, W. and Anderson, S. (1986). A comparison of large-sample confidence interval methods for the difference of two binomial probabilities. *American Statistical Association*. **40**:318-322.
- [15] International Conference on Harmonization. (1998). Statistical principles for clinical trials (ICH-E9). Food and Drug Administration, DHHS.
- [16] International Conference on Harmonization. (2000). Guidance on choice of control group and related design and conduct issues in clinical trials (ICH-E10). Food and Drug Administration, DHHS.
- [17] Laster, L. L., Johnson, M. F. and Kotler, M. L. (2006). Non-inferiority trials: the ‘at least as good as’ criterion with dichotomous data. *Statistics in Medicine*. **25**:1115-1130.
- [18] Miettinen, O. and Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine*. **4**:213-226.
- [19] Römel, 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.
- [20] Rothman, K.L., Michels, K.B (1994). The continued unethical use of placebo controls. *New England Journal of medicine*. **331**: 393-398.
- [21] Smith, R. (1998). Fifty years of randomized controlled trials. *British Medical Journal*. **317**:1166.

- [22] World Medical Association Declaration of Helsinki. (1997). Recommendations guiding physicians in biomedical research involving human subject. *Journal of the American Medical Association*. **277**: 925-926.