

COMPARACIÓN DE LOS MÉTODOS PERMUTACIONAL Y ASINTÓTICO EN
PRUEBAS DE HIPÓTESIS APLICADAS A TABLAS DE CONTINGENCIA DE
DIMENSIÓN 2×2 Y $R \times C$ NO ORDENADAS

Ing. Lourdes A. Sánchez A.
Tutor: Dra. Harú Martínez de Cordero

Maracay, Octubre de 2.005

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE AGRONOMÍA
COMISIÓN DE ESTUDIOS DE POSTGRADO
POSTGRADO DE ESTADÍSTICA

COMPARACIÓN DE LOS MÉTODOS PERMUTACIONAL Y ASINTÓTICO EN PRUEBAS
DE HIPÓTESIS APLICADAS A TABLAS DE CONTINGENCIA DE DIMENSIÓN 2×2 Y
 $R \times C$ NO ORDENADAS

Ing. Lourdes A. Sánchez A.

Tutor: Dra. Harú Martínez de Cordero

Maracay, Octubre de 2.005

Trabajo especial de grado presentado ante los honorables miembros del Comité Académico del Postgrado de Estadística de la Facultad de Agronomía de la Universidad Central de Venezuela como último requisito exigido para optar al título de **Magister Scientiarum en Estadística**.

UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE AGRONOMÍA
COMISION DE ESTUDIOS DE POSTGRADO
POSTGRADO EN ESTADISTICA
MARACAY

VEREDICTO

Quienes suscriben, Miembros del Jurado designado por el Consejo de la Facultad de Agronomía de la Universidad Central de Venezuela, para examinar el Trabajo de Grado presentado por la ciudadana SANCHEZ AGUILERA LOURDES A., C.I. No. 7.190.805, titulado "COMPARACIÓN DE LOS MÉTODOS PERMUTACIONAL Y ASINTÓTICO EN PRUEBAS DE HIPÓTESIS APLICADAS A TABLAS DE CONTINGENCIA DE DIMENSIÓN 2×2 Y $R \times C$ NO ORDENADAS", a los fines de cumplir con el requisito académico exigido para optar al Título de *MAGISTER SCIENTIARUM EN ESTADÍSTICA*, dejan constancia de lo siguiente:

1. Leído como fue dicho Trabajo de Grado, por cada uno de los Miembros del Jurado, éste fijó el día viernes 21 de octubre de 2005, a las 08:00 a.m., para que la autora lo defendiera en forma pública, lo que ésta hizo, en el Salón Principal del Postgrado en Estadística, mediante un resumen oral de su contenido, luego de lo cual respondió satisfactoriamente a las preguntas que le fueron formuladas por el Jurado, todo ello conforme a lo dispuesto en los artículos 45, 47, 50, 51, 52 y 53 del Reglamento de Estudios de Postgrado de la Universidad Central de Venezuela y a las Normas para la Elaboración de Tesis de Grado de la Facultad de Agronomía de la misma Universidad.
2. Finalizada la defensa pública del Trabajo de Grado, el Jurado decidió aprobarlo, por considerar sin hacerse solidario de las ideas expuestas por la autora, que se ajusta a lo dispuesto y exigido por el Reglamento de Estudios de Postgrado. Para dar este Veredicto el jurado estimó que se estudiaron el comportamiento de varios estadísticos de pruebas aplicadas al análisis de datos categorizados bajo diferentes esquemas experimentales y se cumplieron los objetivos del trabajo lo cual representa una contribución al estudio de pruebas de permutación aplicadas a tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas.

[Handwritten signature]

[Handwritten signature]

[Handwritten signature]

Página 2 del Veredicto del Trabajo de Grado de la ciudadana SANCHEZ AGUILERA LOURDES A.

En fe de lo cual se levanta la presente acta, a los veintiún días del mes de octubre de dos mil cinco, dejándose también constancia de que, conforme a lo dispuesto en la normativa jurídica vigente actuó como Coordinador del Jurado la tutora, Dra. Harú Martínez de Cordero.

Firma del Jurado Examinador:


Dr. Miguel Balza
C.I. No. 4.800.337




Dr. Lusbi Herrera
C.I. No. 3.255.060


Dra. Harú M. de Cordero
Coordinador
C.I. No. 4.391.240

21-10-05
Veredicto

Miriam

“Jamás lo grandioso se logró sin entusiasmo”

Ralph Waldo Emerson

DEDICATORIA

A Dios Padre Todopoderoso.
A mi madre, quien con paciencia inagotable esperaba el momento de entrega de
este trabajo para compartir el triunfo.
¡Gracias, madre mía!... ¡Que Dios te tenga en la Gloria!

AGRADECIMIENTOS

A Nino, Veronica e Irene por su alto grado de tolerancia durante el desarrollo de este trabajo.

A mi tutora, Dra. Harú Martínez de Cordero, por su estímulo y sabios consejos.

A Luis Manuel, por sus críticas, comentarios y contagioso entusiasmo.

TABLA DE CONTENIDO

	Página
1. INTRODUCCIÓN	1
1.1. Planteamiento del problema	1
1.2. Justificación de la investigación	4
1.3. Alcance	5
1.4. Limitaciones	6
2. OBJETIVOS	7
2.1. Objetivo general	7
2.2. Objetivos específicos	7
3. HIPÓTESIS DE TRABAJO	8
4. MATERIALES Y MÉTODOS	9
4.1. Materiales	9
4.2. Métodos	10
5. MARCO TEÓRICO	22
5.1. Antecedentes	22
5.1.1. Historia y trabajos previos con pruebas de permutación	22
5.1.2. Trabajos previos en datos categorizados	25
5.1.2.1. Alan Agresti y colaboradores	25
5.1.2.2. Cyrus Mehta, Nitin Patel y colaboradores	26
5.1.2.3. Otros aportes importantes	27

5.2. Generalidades sobre pruebas de permutación	28
5.2.1. Prueba de permutación	30
5.2.2. Prueba no paramétrica de rangos	31
5.2.3. Prueba paramétrica	32
5.3. Pruebas de permutación aplicadas a tablas de contingencia de dimensión 2×2	33
5.3.1. Pruebas de hipótesis exactas condicionadas	35
5.3.2. Pruebas de hipótesis exactas no condicionadas	43
5.3.3. Comparación entre las pruebas de permutación condicionadas y no condicionadas en tablas de contingencia de dimensión 2×2	49
5.4. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$ no ordenadas	55
5.5. Aspectos computacionales en pruebas de permutación para datos categorizados	61
5.5.1. Enumeración completa	61
5.5.2. Simulación	65
6. RESULTADOS Y DISCUSIÓN	69
6.1. Estudio comparativo del comportamiento de las metodologías exactas y asintóticas sobre tablas de contingencia tomadas de algunos trabajos publicados en la literatura científica internacional	69
6.1.1. Tablas de contingencia de dimensión 2×2	69

6.1.2. Tablas de contingencia de dimensión $R \times C$ no ordenadas	82
6.2. Estudio del comportamiento de las metodologías exactas y asintóticas a partir de la simulación de tablas de contingencia	88
6.2.1. Simulación en tablas de contingencia de dimensión 2×2	90
6.2.2. Simulación en tablas de contingencia de dimensión $R \times C$ no ordenadas	94
6.2.3. Tamaños de muestra minimax	99
7. CONCLUSIONES Y RECOMENDACIONES	102
8. REFERENCIAS BIBLIOGRÁFICAS	105
9. ANEXOS	113

ÍNDICE DE TABLAS

	Página
Tabla 1. Estructuras de vectores de parámetros binomiales para las tablas de contingencia de dimensión 2×2	12
Tabla 2. Estructuras de vectores de parámetros multinomiales para las tablas de contingencia de dimensión $R \times C$	13
Tabla 3. Configuraciones poblacionales de las tablas de contingencia	13
Tabla 4. Totales marginales de las filas de las tablas de contingencia 2×2	14
Tabla 5. Totales marginales de las filas de las tablas de contingencia simuladas de dimensión $R \times C$	21
Tabla 6. Esquema de una tabla de contingencia observada \mathbf{x} de dimensión 2×2	34
Tabla 7. Tabla de contingencia genérica \mathbf{y} de dimensión 2×2	38
Tabla 8. Esquema de una tabla de contingencia observada \mathbf{x} de dimensión $R \times C$	55
Tabla 9. Esquema de una tabla de contingencia genérica \mathbf{y} de dimensión $R \times C$	57
Tabla 10. Clasificación de las respuestas a una pregunta dicotómica en dos grados distintos clasificados según el género	69
Tabla 11. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo 1	70

Tabla 12. Resultados de un estudio retrospectivo en el tratamiento de cáncer de laringe	72
Tabla 13. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente para el ejemplo 2	72
Tabla 14. Empleados despedidos y retenidos por cierta empresa para cuatro períodos de despidos clasificados según la edad	74
Tabla 15. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente para el ejemplo 3, primer período	75
Tabla 16. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente para el ejemplo 3, cuarto período	75
Tabla 17. Tabla de contingencia utilizada por Berger y Boos para ilustrar la modificación propuesta por ellos para el cálculo del p -valor exacto de la prueba no condicionada de Barnard	78
Tabla 18. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo de Berger y Boos	78
Tabla 19. Resultados del estudio sueco sobre uso de aspirina e infarto al miocardio	80
Tabla 20. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al estudio sueco sobre el uso de aspirina e infarto al miocardio	80
Tabla 21. Ejemplo utilizado por Agresti et al para estimar el p -valor exacto utilizando el método de ensayos de Monte Carlo	82
Tabla 22. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo utilizado por Agresti et al	83

Tabla 23. Tipo de tumor cerebral y ubicación del mismo	84
Tabla 24. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo utilizado por Davis	84
Tabla 25. Ejemplo utilizado por Senchaudhuri et al para comparar al ensayo de Monte Carlo clásico con el ensayo de Monte Carlo con control de variables	85
Tabla 26. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo utilizado por Senchaudhuri et al	86
Tabla 27. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en diferentes pruebas aplicadas para tablas de contingencia de dimensión 2×2	93
Tabla 28. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en la prueba de Freeman-Halton para tablas de contingencia no ordenadas de dimensión $R \times C$	96
Tabla 29. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en la prueba de χ^2 para tablas de contingencia no ordenadas de dimensión $R \times C$	97
Tabla 30. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en la prueba de razón de verosimilitud para tablas de contingencia no ordenadas de dimensión $R \times C$	98
Tabla 31. Tamaños de muestra minimax observados para el umbral $\Delta^* = 0,01$ para las distintas pruebas en tablas de contingencia de dimensión 2×2	100

Tabla 32. *Tamaños de muestra minimax observados para el umbral $\Delta^* = 0,01$ en la prueba de Freeman-Halton para tablas de contingencia de dimensión $R \times C$ no ordenadas*

101

ÍNDICE DE FIGURAS

	Página
<i>Figura 1. Valores de la medida de similitud Δ según el tamaño de muestra para las pruebas consideradas en tablas de contingencia 3×3 balanceadas y vectores de probabilidades multinomiales uniformes</i>	89

RESUMEN

En este trabajo se comparan los métodos permutacional y asintótico para pruebas de hipótesis aplicadas a tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas, utilizando como medida de comparación la diferencia entre el p -valor exacto y asintótico.

A fin de ilustrar las condiciones bajo las cuales los p -valores asintóticos asociados a las pruebas de hipótesis consideradas: prueba de Fisher, χ^2 y razón de verosimilitud; convergen a sus contrapartes exactos obtenidos por pruebas de permutación, se analizaron ejemplos publicados en la literatura científica internacional que usan el enfoque asintótico, y por otra parte, se realizó un estudio de tablas de contingencia simuladas, utilizando muestreo aleatorio multinomial, con la finalidad de determinar el tamaño de muestra a partir del cual ambos p -valores obtenidos son equivalentes.

Los resultados mostraron que las diferencias entre los p -valores asociados a las pruebas de hipótesis aplicadas variaron según la dimensión de la tabla de contingencia, tamaño de muestra, y balanceo de los datos; y que el tamaño de muestra a partir del cual ambas metodologías son equivalentes estuvo en la mayoría de los casos por encima de $N = 1000$, lo que permite recomendar el uso de pruebas de permutación para las tablas de contingencia 2×2 y $R \times C$ no ordenadas en las pruebas de hipótesis consideradas.

Palabras claves: pruebas asintóticas, pruebas de permutación, tablas de contingencia no ordenadas, p -valor, muestreo aleatorio multinomial.

ABSTRACT

In this work the methods permutational and asymptotic are compared for hypothesis tests applied to 2×2 and $R \times C$ dimension unordered contingency tables, by using the difference between the exact p -value and asymptotic one as a comparison measure.

In order to illustrate the conditions under which the asymptotic p -values associated to the tests of considered hypothesis: Fisher test, χ^2 , likelihood ratio; converge to their counterpart exacts obtained by permutation tests, examples published in the international scientific literature that use the asymptotic approach were analyzed, and on the other hand, a simulation study of contingency tables using multinomial random sampling was carried out, with the purpose of determining the sample size starting from which both obtained p -values are equivalent.

The results showed that the differences among the p -values associated to the hypothesis test applied ranked according to the contingency table dimension, sample size, and data balance; and the sample size starting from which both methodologies are equivalent was in most of the cases superior to $N = 1000$, which allows to recommend the use of permutation tests for the 2×2 and $R \times C$ dimension unordered contingency tables in the hypothesis test considered.

***Key words:* asymptotic methods, permutation tests, unordered contingency tables, p -value, multinomial random sampling.**

1. INTRODUCCIÓN

1.1. Planteamiento del problema

Las tablas de contingencia son ampliamente utilizadas en la investigación de un gran número de disciplinas; las cuales incluyen entre otras, a las ciencias médicas, económicas, sociales y biológicas; ya que constituyen una herramienta de análisis muy importante puesto que permiten establecer relaciones entre las variables causa y las variables respuesta, cualquiera que sea la situación considerada bajo estudio y resultan de mucha utilidad al investigador porque éstas pueden utilizarse en conjunto de datos medidos en cualquier escala. Tal es el caso cuando se desea determinar la relación entre el hábito de fumar (fumador, no fumador) e infartos al miocardio (presencia de infarto, ausencia de infarto), se mide la frecuencia directamente para todas las posibilidades dadas, agrupándose éstas en una tabla de contingencia. Otra situación ocurre cuando los datos se toman en base a unas clases preestablecidas dependiendo si la variable respuesta es cuantitativa; por ejemplo, se quiere determinar si el consumo de bebidas alcohólicas tiene relación con infartos al miocardio, las frecuencias se agrupan de acuerdo a intervalos o rangos preestablecidos que obedezcan a un orden en particular, construyéndose así de esta manera la tabla de contingencia respectiva.

Si bien es cierto que existen muchas pruebas diseñadas para tablas de contingencia, también lo es que la mayoría de éstas tienen su fundamento en el

comportamiento asintótico de los estadísticos utilizados para tales pruebas y cuya distribución teórica se encuentra tabulada, en este sentido vale la pena mencionar la conocida prueba de independencia basada en el estadístico de Pearson y su correspondiente distribución asintótica χ^2 . Estos métodos ofrecen un comportamiento adecuado en el caso de muestras muy grandes y balanceadas en cuanto a totales de filas y columnas de la tabla de contingencia considerada, pero que no necesariamente el comportamiento es el mismo si la muestra es grande y desbalanceada o si la muestra es pequeña, balanceada o no, o también si la tabla de contingencia presenta frecuencias iguales a cero o uno en algunas de sus celdas; es decir, que la muestra sea dispersa, (Agresti, 2002; Cytel Software, 2001 y 2003; Good, 2000). Además, muchos de los paquetes estadísticos que se utilizan no arrojan el resultado del estadístico para una prueba dada en particular si no están presentes un número mínimo para las frecuencias en las celdas o advierten que tal resultado es inválido.

Por otra parte, Agresti y Wackerly, (1977); y Baglivo, Oliver y Pagano, (1988); indican que en el caso de muestras pequeñas, cuando no se cumplen los supuestos para la aplicación del enfoque asintótico, el investigador puede realizar ajustes convenientes sobre los datos originales que lejos de obtener inferencias estadísticas confiables lo conduce a obtener conclusiones erróneas.

Una solución a estos problemas lo constituyen los métodos basados en pruebas de permutación aplicadas a datos categorizados, las cuales por lo general utilizan el mismo estadístico que las correspondientes aproximaciones asintóticas, pero que a diferencia de éstas, utilizan la distribución de las permutaciones de los datos, distribución discreta y libre de parámetros desconocidos, en vez de la distribución teórica correspondiente, lo cual permite que la prueba realizada a partir de éste enfoque presente niveles de significación exactos.

Hace algunas décadas, la aplicación de los métodos permutacionales estaba limitada debido a la complejidad de los cálculos computacionales requeridos en contraposición con el enfoque clásico en el que se utiliza una distribución teórica conocida, de tal forma que la exigencia en cuanto a requerimientos computacionales es mínima; pero gracias al avance tecnológico de las computadoras, los métodos basados en pruebas de permutación han comenzado a aplicarse en el campo de la investigación en diferentes disciplinas, (Good, 2000). En este sentido, cada vez que sea imposible o difícil obtener muestras de gran tamaño, dada las condiciones económicas o la naturaleza de la población, para lograr resultados confiables y conclusiones válidas mediante la aplicación de métodos asintóticos; la alternativa idónea es manejar muestras no necesariamente grandes, balanceadas o no, dispersas o no, con el uso de los métodos exactos, a pesar de los exigentes cálculos computacionales que las mismas requieren.

En este trabajo se estudia el comportamiento de los métodos permutacional y asintótico, en pruebas condicionadas y no condicionadas para tablas de contingencia de dimensión 2×2 , y en pruebas condicionadas para tablas de contingencias de dimensión $R \times C$ no ordenadas, haciendo énfasis en la comparación de los p -valores permutacional y asintótico a los efectos de establecer discrepancias y similitudes para ambas metodologías bajo diferentes escenarios; muestras grandes, balanceadas y desbalanceadas y, muestras pequeñas, balanceadas y desbalanceadas.

1.2. Justificación de la investigación

Como es sabido, muchas investigaciones toman en cuenta las relaciones entre las variables de estudio mediante el uso de tablas de contingencia; por lo que un análisis adecuado y conveniente de éstas aportaría información útil acerca de tales variables. No obstante, la literatura indica que los métodos basados en el enfoque clásico, tradicionalmente usados en la investigación, son sólo aproximaciones subyacentes de una distribución teórica conocida, Good, (2000); señala que si la muestra bajo estudio es pequeña y/o irregularmente distribuida, tales aproximaciones son inexactas y en consecuencia ofrecen poca utilidad para realizar inferencias estadísticas confiables y que, por el contrario, si se usan los métodos basados en pruebas de permutación se logran niveles de significación

exactos independientemente del tamaño de la muestra y/o de la forma de la distribución.

Por otra parte, Bishop, Fienberg y Holland, (1975); citados por Good, (2000); afirman que a medida que se incrementa el tamaño de la muestra, los métodos exactos convergen a las distribuciones asintóticas conocidas, y en consecuencia ambos enfoques son equivalentes para el caso de muestra grandes, por lo que la aplicación de uno u otro método dependerá del investigador.

De manera que un estudio comparativo entre ambos enfoques permite disponer de mayores elementos para una selección justa y conveniente acerca de cual método utilizar dado un determinado y específico escenario de la muestra bajo estudio.

1.3. Alcance

El estudio comprende la metodología para realizar inferencias estadísticas exactas sobre datos categorizados, específicamente para pruebas condicionadas y no condicionadas en tablas de contingencia de dimensión 2×2 y para pruebas condicionadas para tablas de contingencia de dimensión $R \times C$ no ordenadas, con la finalidad de ofrecer una alternativa a los métodos asintóticos en aquellas situaciones donde éstos fallan.

1.4. Limitaciones

La aplicación de las metodologías exactas depende del poder de procesamiento del computador utilizado, debido al elevado número de operaciones computacionales requerido por estos métodos, por lo que el rango de uso de estos métodos basados en pruebas de permutación estará en función del avance tecnológico de las computadoras; en este sentido Good, (2000); comenta que el desarrollo de estas metodologías exactas va a la mano de la evolución del poder de procesamiento tanto del hardware como del software y los algoritmos usados por tales métodos.

2. OBJETIVOS

2.1. Objetivo general

- ❖ Comparar los métodos permutacional y asintótico en pruebas de hipótesis aplicadas a tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas.

2.2. Objetivos específicos

- ❖ Establecer comparación desde el punto de vista teórico entre la metodología clásica y la correspondiente a pruebas de permutación aplicadas a tablas de contingencia.
- ❖ Determinar el p -valor obtenido por la metodología de pruebas de permutación para tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas.
- ❖ Determinar el p -valor obtenido por la metodología asintótica para tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas.
- ❖ Comparar los p -valores obtenidos por la metodología asintótica y la correspondiente a pruebas de permutación para tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas.
- ❖ Determinar la pertinencia de los métodos en función del tamaño de muestra y balanceo de los datos.

3. HIPÓTESIS DE TRABAJO

Los p -valores permutacional y asintótico obtenidos en pruebas de hipótesis aplicadas en tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas arrojarán resultados similares cuando el tamaño de muestra es grande y balanceada, mientras que para muestras pequeñas y grandes, pero desbalanceadas y dispersas, se obtendrán p -valores diferentes.

4. MATERIALES Y MÉTODOS

4.1. Materiales

Los materiales utilizados en este estudio fueron:

- Computadora con procesador Intel Pentium® 4 de 1.7 Gigahertz, tarjeta madre MSI®, 256 Megabytes en memoria RAM y 60 Gigabytes de capacidad en disco duro, como herramienta de procesamiento para calcular los p -valores exactos, asintóticos y aproximados por simulación de Monte Carlo.
- StatXact 6.0 with Cytel Studio® como herramienta de cálculo de los p -valores, ya que este software cuenta con algoritmos capaces de calcular los p -valores exactos y asintóticos además del estimado de p -valores exactos mediante simulación por ensayos de Monte Carlo.
- Ejemplos publicados en la literatura internacional acerca del uso y aplicación de tablas de contingencia en diferentes disciplinas.
- Dataplot® versión 11/2002, software utilizado como herramienta de simulación de muestras aleatorias multinomiales independientes.
- Microsoft Visual FoxPro® versión 7.0 para Windows, software utilizado para crear y manejar la base de datos constituida por las muestras aleatorias multinomiales independientes generadas con el software Dataplot®.
- Conjunto de tablas de contingencia simuladas utilizando tanto el software Dataplot® como el software Microsoft Visual FoxPro®.

4.2. Métodos

Se revisaron los aspectos teóricos correspondientes a ambos enfoques, exacto y asintótico, relativos a las pruebas de hipótesis condicionadas y no condicionadas aplicadas en tablas de contingencia de dimensión 2×2 y las pruebas de hipótesis condicionadas aplicadas a tablas de contingencia de dimensión $R \times C$.

Para el caso de las tablas de contingencia de dimensión 2×2 las pruebas aplicadas en este estudio fueron:

- Prueba χ^2 de Pearson.
- Prueba de razón de verosimilitud.
- Prueba exacta de Fisher.
- Prueba de Barnard.

Y para el caso de las tablas de contingencia de dimensión $R \times C$ no ordenadas se aplicaron:

- Prueba χ^2 de Pearson.
- Prueba de razón de verosimilitud.
- Prueba de Freeman-Halton.

Los ejemplos considerados se seleccionaron de tal forma que permitieran mostrar cuando el comportamiento es disímil entre ambos enfoques, en este sentido, las muestras presentaron las siguientes características:

- Muestras pequeñas balanceadas, desbalanceadas y dispersas.
- Muestras grandes balanceadas, desbalanceadas y dispersas.

Posteriormente se llevaron a cabo las pruebas de hipótesis pertinentes a cada caso y se determinaron los p -valores exactos y asintóticos.

Para aquellos casos en los que por limitaciones del computador no se puede calcular el p -valor exacto, se calcularon el p -valor asintótico y la estimación del p -valor exacto por el método de ensayos de Monte Carlo clásico.

El método de ensayos de Monte Carlo clásico utilizado estuvo basado en una muestra aleatoria de 20000 tablas de contingencia, siguiendo la recomendación de Agresti, Wackerly y Boyett, (1979); quienes afirman que con un máximo de 16589,44 tablas se garantiza la construcción de un intervalo para el p -valor exacto con una amplitud máxima de 0,01 unidades al 99% de confianza.

Por otra parte, las tablas de contingencia simuladas fueron diseñadas de tal manera que permitieran mostrar como la tasa de convergencia del enfoque asintótico al exacto se ve afectada por características intrínsecas a la tabla, tales como dimensión, tamaño de muestra, vectores de parámetros multinomiales, configuración poblacional y balanceo; para ello se consideraron las siguientes estructuras:

- Las dimensiones consideradas en este estudio para las tablas de contingencia no ordenadas se dividieron en dos grupos, a saber:

- Tablas de contingencia de dimensión 2×2 .
- Tablas de contingencia de dimensión $R \times C$, donde éstas variaron desde $2 \times c$ hasta $5 \times c$; con $c = 3, 4, 5$.
- Los tamaños de muestra considerados para cada una de las realizaciones en las dimensiones de las tablas de contingencia diseñadas fueron: $N = 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500, 1000$.
- Las estructuras de los parámetros considerados fueron:
 - Para las tablas de dimensión 2×2 se consideraron dos estructuras de parámetros binomiales: probabilidades uniformes y decrecientes, las cuales se muestran en la tabla 1.

Tabla 1. Estructuras de vectores de parámetros binomiales para las tablas de contingencia de dimensión 2×2 .

VECTORES DE PARÁMETROS BINOMIALES	
Uniforme (U)	Decreciente (D)
$\left[\frac{1}{2}, \frac{1}{2}\right]$	$\left[\frac{4}{5}, \frac{1}{5}\right]$

- Para las tablas de contingencia de dimensión $R \times C$ se consideraron tres estructuras para los vectores de parámetros multinomiales, a saber: vectores de probabilidades uniformes, simétricos y

decrecientes. Tales estructuras variaron según las c categorías de respuestas en las columnas como se muestra en la tabla 2.

Tabla 2. Estructuras de vectores de parámetros multinomiales para las tablas de contingencia de dimensión $R \times C$.

VECTORES DE PARÁMETROS MULTINOMIALES			
Categorías de Respuesta	Uniforme (U)	Simétrico (S)	Decreciente (D)
$c = 3$	$\left[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right]$	$\left[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right]$	$\left[\frac{3}{5}, \frac{3}{10}, \frac{1}{10}\right]$
$c = 4$	$\left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right]$	$\left[\frac{3}{20}, \frac{7}{20}, \frac{7}{20}, \frac{3}{20}\right]$	$\left[\frac{1}{2}, \frac{3}{10}, \frac{3}{20}, \frac{1}{20}\right]$
$c = 5$	$\left[\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right]$	$\left[\frac{1}{10}, \frac{1}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{10}\right]$	$\left[\frac{2}{5}, \frac{3}{10}, \frac{3}{20}, \frac{1}{10}, \frac{1}{20}\right]$

- A partir de estos vectores de probabilidad multinomial se crearon las configuraciones poblacionales para las tablas de contingencia según si la hipótesis nula de igualdad entre los vectores de probabilidad multinomiales sea cierta o falsa, estas configuraciones se presentan en la tabla 3.

Tabla 3. Configuraciones poblacionales de las tablas de contingencia.

CONFIGURACIONES POBLACIONALES DE LAS TABLAS DE CONTINGENCIA					
Dimensión		H ₀ Cierta		H ₀ falsa	
2×2		UU	-	DD	UD
$R \times C$	$r = 2$	UU	SS	DD	US, UD, SD
	$r = 3$	UUU	SSS	DDD	USD
	$r = 4$	UUUU	SSSS	DDDD	USSD
	$r = 5$	UUUUU	SSSSS	DDDDD	USSDD

- Para cada una de todas las posibles combinaciones en cuanto a dimensión, tamaño de muestra, parámetros y configuración poblacional, se simularon dos tablas de contingencia, una con totales marginales de filas balanceados y otra con totales marginales de filas desbalanceados según el esquema presentado en la tabla 4 para las tablas de contingencia de dimensión 2×2 y la tabla 5, mostrada al final de esta sección, para las tablas de contingencia de dimensión $R \times C$.

Tabla 4. Totales marginales de las filas de las tablas de contingencia 2×2 .

TOTALES MARGINALES DE LAS FILAS DE LAS TABLAS DE CONTINGENCIA 2×2		
<i>N</i>	Balanceado	Desbalanceado
5	3-2	4-1
10	5-5	8-2
15	8-7	12-3
20	10-10	16-4
25	13-12	20-5
30	15-15	24-6
40	20-20	32-8
50	25-25	40-10
75	38-37	60-15
100	50-50	80-20
250	125-125	200-50
500	250-250	400-100
1000	500-500	800-200

Todas estas posibles combinaciones de tablas de contingencia generadas hacen un total de 1404 tablas de contingencia de dimensión $R \times C$ no ordenadas y 78 tablas de contingencia de dimensión 2×2 .

Posteriormente, se generaron diez muestras aleatorias binomiales independientes para el caso de tablas de contingencia 2×2 y diez muestras aleatorias multinomiales independientes para el caso de tablas de contingencia $R \times C$; esto se llevó a cabo para cada uno de los totales marginales de filas y vectores de probabilidad considerados utilizando el Software Dataplot[®] versión 11/2002, (National Institute of Standards and Technology, 2002); todos estos resultados se almacenaron en un programa manejador de base de datos escrito en lenguaje Microsoft Visual FoxPro[®] Versión 7.0 y mediante el cual se crearon las tablas de contingencia diseñadas, extrayendo sin reemplazamiento las filas correspondientes a cada una de las configuraciones poblacionales consideradas; de tal forma que las tablas de contingencia simuladas presenten un esquema de muestreo de producto de poblaciones multinomiales independientes para las tablas de dimensión $R \times C$ y producto de dos poblaciones binomiales independientes para las tablas de dimensión 2×2 , tal como lo establece Agresti, (2002).

Las instrucciones para generar las muestras de poblaciones multinomiales y binomiales independientes utilizando el software Dataplot[®] se muestran en los anexos 1 y 2.

A cada una de estas tablas generadas por simulación se les aplicaron las pruebas exactas y asintóticas respectivas utilizando el software StatXact 6.0 with Cytel Studio[®]. Para el cálculo del p -valor exacto se configuró al software para un tiempo de procesamiento máximo de 15 minutos, si al cabo de este lapso no se hubiera obtenido el p -valor exacto, se procedió a su estimación utilizando ensayos de Monte Carlo.

Para la comparación de los p -valores, exacto y asintótico, se utilizó como medida de similitud a:

$$\Delta = |p_2 - \tilde{p}_2|$$

donde

p_2 : es el p -valor exacto para una hipótesis nula bilateral.

\tilde{p}_2 : es el p -valor asintótico para una hipótesis nula bilateral.

Es evidente que en la medida que el p -valor asintótico converge al p -valor exacto, Δ tenderá a cero; y se espera que esto ocurra para tamaños de muestra grandes, esto es:

$$\lim_{N \rightarrow \infty} \Delta = 0$$

Berger, (2000); propone una estrategia basada en esta medida de similitud Δ para decidir entre el uso de metodologías exactas o asintóticas:

$$\begin{cases} \text{Si } \Delta \leq \Delta^* & \text{Reportar el } p\text{-valor asintótico} \\ \text{Si } \Delta > \Delta^* & \text{Reportar el } p\text{-valor exacto} \end{cases}$$

donde Δ^* es un umbral de decisión entre ambas metodologías.

Si bien Berger, (2000); recomienda el uso de esta regla de decisión y muestra que el nivel de significación real de la prueba utilizada puede alcanzar un máximo de $\alpha + \Delta^*$, no sugiere un valor específico para Δ^* .

En este trabajo se utilizó como umbral de decisión a $\Delta^* = 0,01$, aunque el valor óptimo puede ser objeto de estudios posteriores y variar de una disciplina a otra; se considera que $\Delta^* = 0,01$ puede ser adecuado para este estudio preliminar, ya que un valor muy grande para éste inflará de manera desmesurada a la probabilidad de cometer el error tipo I y en cambio, un valor muy pequeño conllevará a utilizar pruebas asintóticas sólo para tamaños de muestra muy grandes.

A partir de este criterio de decisión se determinó un tamaño de muestra crítico para cada uno de los escenarios considerados en este estudio, entendiéndose por tamaño de muestra crítico a aquel tamaño de muestra N_c a partir del cual el Δ observado en la simulación será menor o igual al Δ^* establecido; y es partir de este tamaño que pudieran utilizarse métodos asintóticos en vez de métodos exactos, es decir:

$$N_c = \min \left[N_{d, b, t, h} : \Delta \leq \Delta^* \right]$$

donde

N_c : es el tamaño de muestra crítico para la tabla de contingencia de dimensión d , tipo de balanceo b , prueba t y configuración de los vectores de parámetros h .

d : es la dimensión de la tabla de contingencia considerada,

$$d = \{2 \times 2\}$$

para el caso de las tablas de contingencia de dimensión 2×2 , y

$$d = \left\{ \begin{array}{l} 2 \times 3, 2 \times 4, 2 \times 5, 3 \times 3, 3 \times 4, 3 \times 5 \\ 4 \times 3, 4 \times 4, 4 \times 5, 5 \times 3, 5 \times 4, 5 \times 5 \end{array} \right\}$$

para el caso de las tablas de contingencia de dimensión $R \times C$ no ordenadas.

b : es la clase de balanceo de los totales marginales de las filas de la tabla de contingencia considerada, tal que

$$b = \{\text{balanceado, desbalanceado}\}$$

t : es la prueba correspondiente aplicada a la tabla de contingencia considerada, tal que

$$t = \left\{ \begin{array}{l} \text{Fisher, Chi cuadrado} \\ \text{Verosimilitud, Barnard} \end{array} \right\}$$

para el caso de las tablas de contingencia de dimensión 2×2 , y

$$t = \left\{ \begin{array}{l} \text{Freeman-Halton, Chi cuadrado} \\ \text{Verosimilitud} \end{array} \right\}$$

para el caso de las tablas de contingencia de dimensión $R \times C$ no ordenadas.

h : es la configuración poblacional de los vectores de parámetros considerada, mostradas en la tabla 1 para las tablas de contingencia de dimensión 2×2 y en la tabla 2 para las tablas de contingencia de dimensión $R \times C$ no ordenadas.

En base a las tablas de contingencia analizadas se recomendó la pertinencia de una u otra metodología según los tamaños de muestra considerados. Esta recomendación se hizo a partir de los tamaños de muestras críticos determinados por dimensión, configuración poblacional, tipo de prueba aplicada y balanceo en los totales marginales de las filas.

Finalmente, el autor propone una recomendación minimax para el uso de una u otra metodología de acuerdo a los tamaños de muestra críticos obtenidos para todas las situaciones consideradas en este estudio. Así, para cada prueba, dimensión y balanceo de datos, el tamaño de muestra mínimo elegido para utilizar pruebas asintóticas está dado por:

$$N_{\minimax_{d, b, t}} = \max_{d, b, t} [N_c]$$

Esta elección de N _{minimax} ^{d, b, t} garantiza que a partir de este tamaño de muestra se cumple que $\Delta \leq \Delta^*$ para cualquiera de las configuraciones de los vectores de parámetros considerados.

Tales hallazgos se presentan en la sección correspondiente a resultados y discusión.

Tabla 5. Totales marginales de las filas de las tablas de contingencia simuladas de dimensión $R \times C$.

TOTALES MARGINALES DE LAS FILAS DE LAS TABLAS DE CONTINGENCIA $R \times C$								
N	$2 \times c$		$3 \times c$		$4 \times c$		$5 \times c$	
	Balanceado	Desbalanceado	Balanceado	Desbalanceado	Balanceado	Desbalanceado	Balanceado	Desbalanceado
5	3-2	4-1	2-2-1	3-1-1	2-1-1-1	2-1-1-1	1-1-1-1-1	1-1-1-1-1
10	6-4	8-2	3-3-4	7-2-1	2-3-3-2	5-3-1-1	2-2-2-2-2	4-2-2-1-1
15	8-7	12-3	5-5-5	10-3-2	4-4-4-3	7-5-2-1	3-3-3-3-3	7-4-2-1-1
20	10-10	16-4	7-7-6	14-4-2	5-5-5-5	10-6-3-1	4-4-4-4-4	9-6-3-1-1
25	13-12	20-5	8-8-9	17-5-3	6-6-6-7	12-7-4-2	5-5-5-5-5	11-8-4-1-1
30	15-15	24-6	10-10-10	21-6-3	7-8-8-7	15-9-4-2	6-6-6-6-6	13-9-4-2-2
40	20-20	32-8	13-13-14	28-8-4	10-10-10-10	20-12-6-2	8-8-8-8-8	18-12-6-2-2
50	25-25	40-10	16-17-17	35-10-5	12-13-13-12	25-15-7-3	10-10-10-10-10	23-15-8-2-2
75	38-37	60-15	25-25-25	52-15-8	18-19-19-19	37-23-11-4	15-15-15-15-15	34-22-11-4-4
100	50-50	80-20	33-33-34	70-20-10	25-25-25-25	50-30-15-5	20-20-20-20-20	45-30-15-5-5
250	125-125	200-50	83-83-84	175-50-25	62-63-63-62	125-75-37-13	50-50-50-50-50	113-75-38-12-12
500	250-250	400-100	166-167-167	350-100-50	125-125-125-125	250-150-75-25	100-100-100-100-100	225-150-75-25-25
1000	500-500	800-200	333-333-334	700-200-100	250-250-250-250	500-300-150-50	200-200-200-200-200	450-300-150-50-50

5. MARCO TEÓRICO

El presente capítulo trata de los aspectos teóricos relacionados con las pruebas de permutación aplicadas a tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas. En la sección 5.1 se hace una revisión bibliográfica de los trabajos previos en pruebas de permutación haciendo énfasis en su aplicación en datos categorizados; en la sección 5.2 se discuten aspectos generales sobre pruebas de permutación; luego, en las secciones 5.3 y 5.4 se estudian las pruebas de permutación aplicadas a tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas; finalmente, en la sección 5.5 se expone una reseña histórica sobre los aspectos computacionales relacionados con las pruebas de permutación en datos categorizados.

5.1. Antecedentes

5.1.1. Historia y trabajos previos con pruebas de permutación

Ludbrook y Dudley, (1998); comentan que R. A. Fisher, introdujo en 1935, en su libro *"Experimental Designs"* :

- La prueba exacta de Fisher para tablas de contingencia 2×2 .
- Prueba de permutación para diferencia de medias.

Fisher analizó los resultados obtenidos por Charles Darwin referentes a la comparación de la altura entre plantas de maíz autopolinizadas y con polinización

cruzada; para ello se vio en la necesidad de desarrollar a mano las 32768 posibles permutaciones de los datos y encontró un p -valor exacto dado por:

$$p\text{-valor exacto: } 1726/32768 = 0,05267$$

donde el numerador corresponde a las 1726 permutaciones cuyo estadístico de prueba fue tan o más extremo que el observado en los datos originales.

Por otra parte, calculó también el p -valor asintótico dado por la distribución t de Student, el cual fue:

$$p\text{-valor } t \text{ de Student} = 0,0497$$

Fisher concluyó que el método de permutación "Constituía una posibilidad de revisión de los métodos (clásicos) más expeditos", por cuanto que la diferencia entre ambos p -valores no fue muy grande.

Posteriormente perdió interés en el tema, quizá tal vez por lo tedioso de los cálculos manuales de los métodos de permutación, (Ludbrook y Dudley, 1998).

Ludbrook y Dudley, (1998); comentan además que Eden y Yates, (1933); analizaron un experimento agrícola conducido bajo un diseño de bloques aleatorizados con réplicas con el propósito de defender al ANAVAR de las acusaciones de que este método es muy sensible a desviaciones de los datos de la normalidad.

Tomaron una muestra aleatoria de 1000 experimentos de $(24)^8$ posibles arreglos de los datos y concluyeron que la prueba F puede ser aplicada en

distribuciones asimétricas, por cuanto que los p -valores obtenidos por ambos métodos fueron similares.

Ludbrook y Dudley, (1998); citan que Pitman, (1937), Box y Anderson, (1955); también empezaron a aplicar pruebas de permutación en diferentes disciplinas.

Por otra parte, Good, (2000); expone que:

- En los años 50 y principios de los 60, los principales esfuerzos se dirigieron a probar la equivalencia asintótica entre las pruebas clásicas y las pruebas de permutación. Entre los principales investigadores, Good, (2000); menciona a: Freeman y Halton, (1951); Kempthorne, (1952); Cornfield y Tukey, (1956).
- En los años 60 se introdujeron las pruebas no paramétricas de rangos, las cuales son pruebas de permutación aplicadas a rangos, Good, (2000); cita como investigadores principales a: Siegel, (1956); Lehmann, (1963 y 1975); Cox y Kempthorne, (1963); Hayek, (1967).
- En los años 80, 90 y en la actualidad, los esfuerzos están orientados hacia el desarrollo de algoritmos y software que trabajan con pruebas de permutación para aplicarlas a diferentes disciplinas.

Romano y Boik, (1990); citados por Ludbrook y Dudley, (1998); afirman que existen discrepancias entre los resultados de las pruebas clásicas y las pruebas de permutación en presencia de varianzas desiguales y muestras pequeñas.

5.1.2. Trabajos previos en datos categorizados

Debido a que buena parte de la investigación llevada a cabo en esta área del conocimiento proviene de los trabajos de Alan Agresti, Cyrus Mehta y Nitin Patel, se considera conveniente dividir esta sección de acuerdo a los aportes de éstos y otros investigadores.

5.1.2.1. Alan Agresti y colaboradores

- Su libro de texto *“Categorical Data Analysis”* constituye una de las bases teóricas más sólidas para el estudio de datos categorizados porque contempla numerosos modelos para el análisis de este tipo de datos tanto por métodos exactos como asintóticos, (Agresti, 2002).
- Agresti y Wackerly, (1977); evaluaron diversas pruebas aplicadas a tablas de contingencia y crearon un programa computacional escrito en lenguaje FORTRAN para realizar pruebas exactas condicionadas.
- Agresti, Wackerly y Boyett, (1979); propusieron el método de remuestreo Monte Carlo para tablas de contingencia.

5.1.2.2. *Cyrus Mehta, Nitin Patel y colaboradores*

- Mehta y Patel, (1980); desarrollaron el algoritmo de red para calcular p -valores exactos en pruebas de permutación aplicadas a datos categorizados.
- Mehta y Patel, (1983); aplicaron el algoritmo de red para calcular p -valores exactos en la prueba de Freeman-Halton para tablas de contingencia de dimensión $R \times C$ no ordenadas.
- Mehta, Patel, Senchaudhuri y Tsiasis, (1994); Mehta, Patel y Wei, (1988); aplicaron tal algoritmo de red para calcular p -valores exactos en pruebas de permutación aplicadas a ensayos clínicos.
- Mehta, Patel y Senchaudhuri, (1988); desarrollaron el ensayo de simulación de Monte Carlo denominado muestreo por importancia, basados en el algoritmo de red y demuestran que este es más eficiente que el Monte Carlo clásico en tablas $2 \times C$.
- Mehta y Hilton, (1993); compararon la potencia de dos pruebas, una condicionada y otra no condicionada, para tablas 2×3 y encontraron que la potencia de ambas es similar e incluso superada por la prueba condicionada y conjeturaron que este comportamiento se acentúa con la dimensión de la tabla.
- Mehta y Patel, (1997); revisaron aspectos teóricos de las pruebas de permutación aplicadas a datos categorizados para el enfoque condicionado.

- Gajjar, Mehta, Patel y Senchaudhuri son los creadores del software StatXact[®], (Cytel Software, 2001 y 2003).

5.1.2.3. Otros aportes importantes

- Barnard G. A. propuso en 1945 una prueba no condicionada para probar independencia en tablas de contingencia 2×2 como una alternativa a la prueba exacta de Fisher, (Agresti, 2002; Cytel Software, 2001 y 2003); introdujo además, el método de simulación Monte Carlo en 1963, (Kelly, 1999).
- Zelen, (1971); propuso una prueba exacta para la homogeneidad en tablas de contingencia 2×2 estratificadas.
- Gail y Mantel, (1977); desarrollaron una metodología para calcular el número de tablas de contingencia $R \times C$ condicionadas y propusieron además una aproximación basada en la distribución normal.
- Hollander y Wolfe, (1999); son autores del libro de texto "*Nonparametric Statistical Methods*", que incluye pruebas de permutación; el cual constituye actualmente una de las contribuciones más sólidas para el estudio de la estadística no paramétrica, porque incluye un gran número de pruebas no paramétricas desarrolladas y aplicadas en la actualidad.

5.2. Generalidades sobre pruebas de permutación

Denominadas también pruebas de aleatorización exactas, es un método que consiste en obtener la distribución de todos los rearrreglos posibles de las observaciones para un estadístico dado, sin reemplazamiento, y a partir de esta distribución realizar inferencias estadísticas con probabilidades exactas, (Kelly, 1999; Good, 2000).

El método se basa en suponer que bajo la hipótesis nula, la distribución subyacente de los datos es la misma para todas las poblaciones consideradas, es decir, una observación y_i puede provenir de cualquiera de estas poblaciones, (Good, 2000).

Por lo general, para la mayoría de las pruebas paramétricas clásicas, se han desarrollado pruebas alternativas, ya sean no paramétricas o de permutación, (Good, 2000).

El procedimiento para llevar a cabo una prueba de permutación consta de los siguientes pasos:

- Elegir un estadístico de prueba adecuado.
- Calcular el valor del estadístico de prueba para los datos observados.
- Encontrar todas las permutaciones posibles del experimento, sin reemplazamiento, y calcular el estadístico de prueba para cada uno de estos rearrreglos.

- Construir la distribución de las permutaciones del estadístico de prueba.
- Comparar el valor observado con la distribución de las permutaciones y rechazar la hipótesis nula si el valor observado se ubica hacia los extremos de la distribución.

Bajo ciertas condiciones, las pruebas de permutación son exactas, insesgadas y tan o más potentes que su equivalente paramétrica, (Good, 2000).

Una prueba es denominada exacta con respecto a una hipótesis alternativa, si la probabilidad de ésta de cometer el error tipo I es α ; además, se dice que una prueba es conservadora si el error Tipo I nunca excede α , (Good, 2000).

Se dice que una prueba es insesgada y de nivel α , siempre que su función de potencia β_θ para el parámetro θ satisface las siguientes condiciones:

- (i) β_θ es conservadora; es decir, $\beta_\theta \leq \alpha$ para todo θ que satisfaga la hipótesis nula.
- (ii) $\beta_\theta \geq \alpha$ para todo θ que sea una alternativa de la hipótesis nula, (Lehmann, 1986; Good, 2000).

Lehmann, (1986); propone que una condición suficiente para que una prueba de permutación sea exacta e insesgada es la intercambiabilidad de las observaciones de la muestra. Las observaciones $\{X, Y, Z\}$ son intercambiables, si

la probabilidad de algún resultado conjunto en particular, digamos $X + Y + Z = 6$, es el mismo sin importar el orden en el cual las observaciones sean consideradas.

Good, (2000); indica que observaciones idéntica e independientemente distribuidas y muestras sin reemplazo de una población finita son intercambiables.

Berger, (2000); comenta que las pruebas de permutación preservan la tasa de error Tipo I a un nivel inferior o igual a α , lo cual no necesariamente ocurre con las aproximaciones asintóticas; sin embargo, no necesariamente constituyen las pruebas más potentes disponibles para una situación dada.

Los siguientes esquemas ilustran los pasos necesarios para la aplicación de pruebas de hipótesis según tres enfoques diferentes: permutación, rangos y paramétricas; los cuales permiten visualizar que las estructuras de tales pruebas, aunque similares, varían según el enfoque, (Good, 2000).

5.2.1. Prueba de permutación

1. Se elige un estadístico de prueba $S(X)$.
2. Se calcula S para el conjunto original de observaciones.
3. Se obtiene la distribución de las permutaciones de S por rearrreglos repetidos de las observaciones y se calcula S para todos los rearrreglos. Los remuestreos se realizan sin reemplazamiento.

4. Se obtiene el cuantil del porcentaje superior α de la distribución de las permutaciones y se acepta o se rechaza la hipótesis nula de acuerdo si S para los valores originales es más pequeño o más grande que ese valor.

Si las observaciones son intercambiables bajo la hipótesis nula, entonces la prueba resultante es exacta e insesgada.

5.2.2. Prueba no paramétrica de rangos

1. Se elige el estadístico de prueba S .
2. Se reemplazan las observaciones originales $\{X_{ij}, i = 1, \dots, I, j = 1, \dots, J\}$ por sus rangos en la muestra combinada $\{R_k, k = 1, \dots, IJ\}$. Por ejemplo, si las observaciones originales son 5,2; 1 y 7, sus rangos son 2, 1 y 3. Se calcula S para el conjunto original de rangos.
3. Se obtiene la distribución de las permutaciones de S por rearreglos repetidos de los rangos y se recalcula el estadístico de prueba para cada rearreglo. O, dado que los rangos siempre toman los mismos valores $1, 2, \dots, IJ$; se puede utilizar la distribución de los rangos previamente tabulada.
4. Se acepta o se rechaza la hipótesis de acuerdo con el cuantil del porcentaje superior α de su distribución de permutación.

En pocas palabras, una prueba de rangos es simplemente una prueba de permutación aplicada a los rangos de las observaciones en vez de a los valores originales. Si las observaciones son intercambiables, entonces la prueba de rangos resultante es exacta e insesgada, aunque, generalmente, es menos potente que una prueba de permutación, (Good, 2000).

5.2.3. Prueba paramétrica

Para llevar a cabo una prueba paramétrica (por ejemplo, una prueba t de Student o una prueba F de Snedecor):

1. Se elige un estadístico de prueba, $S(X)$, cuya distribución F_S puede ser calculada y tabulada independientemente de las observaciones.
2. Se calcula S para las observaciones.
3. No es necesario elaborar la distribución F_S puesto que ya es conocida y está tabulada.
4. Se compara $S(X)$ con el cuantil del porcentaje superior α de F_S y se acepta o se rechaza la hipótesis nula de acuerdo a si $S(X)$ es más grande o más pequeño que este valor.

Si S está distribuido como F_S , entonces la prueba paramétrica es exacta y, muchas veces, la prueba disponible más potente. Para que S tenga distribución F_S , en la mayoría de los casos las observaciones necesitan ser independientes y,

con muestras pequeñas, idénticamente distribuidas con una distribución específica, G_S . Si S realmente tiene alguna otra distribución, entonces la prueba paramétrica puede carecer de potencia y puede no ser conservadora. Con muestras grandes, las pruebas de permutación usualmente son tan potentes como la más potente prueba paramétrica, (Bickel y Van Zwet, 1978) citados por Good, (2000). Si S no está distribuida como F_S , pueden ser más potentes.

Por lo general, para la mayoría de las pruebas paramétricas clásicas, se han desarrollado pruebas alternativas, ya sean no paramétricas o de permutación, (Good, 2000); sin embargo, para casos particulares, se han desarrollado pruebas que carecen de equivalente paramétrico, como por ejemplo la prueba *Convex Hull*, (Berger, 2000).

5.3. Pruebas de permutación aplicadas a tablas de contingencia de dimensión 2×2

El caso más simple de analizar es el de una tabla de contingencia de dimensión 2×2 constituida por dos poblaciones binomiales, en la que se desea probar que ambas poblaciones tienen la misma probabilidad de éxito π . Para llevar a cabo esto se presentan dos metodologías alternativas a saber:

- i. Pruebas condicionadas.
- ii. Pruebas no condicionadas.

Los datos para la tabla de contingencia observada \mathbf{x} de dimensión 2×2 , pueden escribirse de la forma mostrada en la tabla 6.

Tabla 6. Esquema de una tabla de contingencia observada \mathbf{x} de dimensión 2×2 .

Respuesta	Población 1	Población 2	Total de Filas
Éxitos	x_{11}	x_{12}	m_{1+}
Fracasos	x_{21}	x_{22}	m_{2+}
Total de Columnas	n_{+1}	n_{+2}	N

Fuente: Cytel Software, 2001 y 2003.

Para las frecuencias x_{ij} ; $i, j = 1, 2$, de la tabla de contingencia observada \mathbf{x} se cumple que

$$\sum_{j=1}^2 x_{ij} = m_{i+}$$

$$\sum_{i=1}^2 x_{ij} = n_{+j}$$

$$\sum_{i=1}^2 \sum_{j=1}^2 x_{ij} = N$$

En la tabla 6 se asume que las dos columnas provienen de dos poblaciones binomiales independientes. Hay x_{1j} éxitos y x_{2j} fracasos en n_{+j} ensayos de Bernoulli independientes, cada uno con probabilidad π_j de éxitos, $j = 1, 2$. La suma de los éxitos de las dos poblaciones es $m_{1+} = x_{11} + x_{12}$. Los tamaños de muestra n_{+1} y n_{+2} se consideran cantidades fijas. Para el enfoque condicionado,

las marginales m_{j+} se consideran fijas, mientras que para el enfoque no condicionado las marginales m_{i+} son variables aleatorias, (Cytel Software, 2001 y 2003).

Se desarrollarán primero los aspectos teóricos concernientes a las pruebas condicionadas debido a su simplicidad en comparación con las pruebas no condicionadas.

5.3.1. Pruebas de hipótesis exactas condicionadas

Las pruebas exactas que se van a desarrollar están orientadas según la hipótesis nula

$$H_0 : \pi_1 = \pi_2$$

contra las alternativas de una cola de la forma

$$H_1 : \pi_2 > \pi_1$$

$$H_1' : \pi_2 < \pi_1$$

y de dos colas de la forma

$$H_1 : \pi_1 \neq \pi_2$$

Bajo el supuesto de que H_0 es cierta, las probabilidades de éxitos para las dos distribuciones binomiales son $\pi_1 = \pi_2 = \pi$, y la probabilidad de observar los datos en la tabla \mathbf{x} es el producto de dos binomiales, por lo cual

$$f_0(\mathbf{x}) = \binom{n+1}{x_{11}} \binom{n+2}{x_{12}} \pi^{x_{11}+x_{12}} (1-\pi)^{x_{21}+x_{22}} \quad (1)$$

Sea el p -valor la probabilidad, bajo H_0 , de obtener una tabla 2×2 al menos tan extrema como la tabla de contingencia observada \mathbf{x} . El problema se presenta al desear calcular dicho valor, puesto que la distribución en la ecuación (1) depende del parámetro desconocido π . Mientras la distribución de las tablas de contingencia de dimensión 2×2 , dependa de tal parámetro desconocido no es posible realizar inferencias exactas, ya que el p -valor es una suma de muchas de tales tablas, cada una de ellas dependiente del mismo valor del parámetro desconocido π , (Cytel Software, 2001 y 2003)

R. A. Fisher propuso en 1934, eliminar el parámetro desconocido π condicionando el total de las filas a los totales observados en la tabla \mathbf{x} , (Good, 2000).

Los principios en los cuales se basa el condicionamiento son el principio de suficiencia y el principio de estadísticos auxiliares (*ancillarity*).

El primero establece que los totales marginales de la tabla de contingencia son estadísticos suficientes de los parámetros desconocidos, por lo que condicionando la tabla a ellos se tiene una forma conveniente de eliminar dichos parámetros de la función de verosimilitud, puesto que toda la información de la

muestra acerca de los parámetros desconocidos se encuentra resumida en estos estadísticos.

El segundo establece que el principio subyacente para la prueba de hipótesis es comparar lo que fue en realidad observado con lo que podría haber sido observado en repeticiones hipotéticas del experimento original bajo la hipótesis nula. En estas repeticiones hipotéticas es una buena idea mantener todas las condiciones experimentales no relacionadas con la hipótesis nula inalteradas tanto como sea posible. Las marginales de la tabla de contingencia son representativas de los parámetros desconocidos cuyos valores no proveen información acerca de la hipótesis nula de interés; en este sentido ellos son estadísticos auxiliares. Fijarlos en repeticiones hipotéticas es lo más cerca que se puede lograr de fijar los valores de los parámetros mismos en tales repeticiones, dado que estos últimos son desconocidos, (Mehta y Patel, 1997).

Para ilustrar este procedimiento, sea \mathbf{y} una tabla de contingencia genérica de dimensión 2×2 donde los totales de las columnas n_{+j} son fijos, tal y como se presenta en la tabla 7.

Tabla 7. Tabla de contingencia genérica \mathbf{y} de dimensión 2×2 .

<i>Respuesta</i>	<i>Población 1</i>	<i>Población 2</i>	<i>Total de Filas</i>
<i>Éxitos</i>	y_{11}	y_{12}	$y_{11} + y_{12}$
<i>Fracasos</i>	y_{21}	y_{22}	$y_{21} + y_{22}$
<i>Total de Columnas</i>	n_{+1}	n_{+2}	N

Fuente: Cytel Software, 2001 y 2003.

Para las frecuencias y_{ij} ; $i, j = 1, 2$, de la tabla de contingencia genérica \mathbf{y} , se cumple que

$$\sum_{i=1}^2 y_{ij} = n_{+j}$$

$$\sum_{i=1}^2 \sum_{j=1}^2 y_{ij} = N$$

El estadístico suficiente para π es $y_{11} + y_{12}$, la suma de los éxitos de las dos poblaciones. El valor observado en la tabla de contingencia \mathbf{x} es m_{1+} . Así, por el principio de suficiencia, si se condiciona $y_{11} + y_{12} = m_{1+}$, la probabilidad condicionada de las tablas genéricas a sus marginales de filas no dependerá del parámetro π como se ilustrará a continuación.

Sea Γ el conjunto formado por las tablas de contingencia 2×2 tales que,

$$\Gamma = \left\{ \mathbf{y}_{2 \times 2} : \sum_{j=1}^2 y_{ij} = m_{i+}; \sum_{i=1}^2 y_{ij} = n_{+j} \right\}$$

Sea $h(\mathbf{y})$ la probabilidad de observar cualquier tabla $\mathbf{y} \in \Gamma$ bajo la hipótesis nula, entonces

$$h(\mathbf{y}) = \frac{f_0(\mathbf{y})}{\sum_{\mathbf{y} \in \Gamma} f_0(\mathbf{y})}$$

donde

$$f_0(\mathbf{y}) = \binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}} \pi^{y_{11}+y_{12}} (1-\pi)^{y_{21}+y_{22}},$$

Condicionando los totales marginales de las filas a los valores observados en la tabla \mathbf{x} ; es decir, haciendo $y_{11} + y_{12} = m_{1+}$ y $y_{21} + y_{22} = m_{2+}$, esta expresión queda

$$f_0(\mathbf{y} | m_{1+}, m_{2+}) = \binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}} \pi^{m_{1+}} (1-\pi)^{m_{2+}}$$

sustituyendo se obtiene

$$h(\mathbf{y}) = \frac{\binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}} \pi^{m_{1+}} (1-\pi)^{m_{2+}}}{\sum_{\mathbf{y} \in \Gamma} \binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}} \pi^{m_{1+}} (1-\pi)^{m_{2+}}} = \frac{\binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}}}{\sum_{\mathbf{y} \in \Gamma} \binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}}} = \frac{\binom{n_{+1}}{y_{11}} \binom{n_{+2}}{y_{12}}}{\binom{N}{m_{1+}}}$$

la cual es una distribución hipergeométrica libre del parámetro π . El anterior resultado indica que se pueden realizar inferencias exactas limitándose a las tablas de contingencia que se encuentran en el conjunto de referencia Γ .

Para determinar si una tabla \mathbf{y} es al menos tan extrema como una tabla observada \mathbf{x} , se establece una medida de discrepancia D , que mida el alejamiento de una tabla cualquiera de la hipótesis nula, es decir, sea $D: \Gamma \rightarrow \Re$ una función de variable real, tal que $D(\mathbf{y})$ le asigne un número real a cada tabla $\mathbf{y} \in \Gamma$, en ese sentido la tabla \mathbf{y} es al menos tan extrema que la tabla \mathbf{x} siempre que $D(\mathbf{y}) \geq D(\mathbf{x})$, (Cytel Software, 2001 y 2003; Agresti y Wackerly, 1977).

Las medidas de discrepancia o estadísticos de prueba más importantes en el estudio de las tablas de doble entrada lo constituyen:

- El estadístico χ^2 de Pearson, el cual está dado por

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(y_{ij} - m_{i+}n_{+j}/N)^2}{m_{i+}n_{+j}/N}$$

- El estadístico de razón de verosimilitud, dado por

$$G^2(\mathbf{y}) = 2 \sum_{i=1}^2 \sum_{j=1}^2 y_{ij} \log \left(\frac{y_{ij}}{m_{i+}n_{+j}/N} \right)$$

- El estadístico de Fisher, dado por

$$F(\mathbf{y}) = -2 \log h(\mathbf{y}) - \log \left(2.51 N^{-3/2} \sqrt{m_{1+} m_{2+} n_{+1} n_{+2}} \right)$$

Las pruebas de hipótesis desarrolladas a partir de tales medidas de discrepancia son llamadas prueba de Pearson, razón de verosimilitud y Fisher, respectivamente, (Cytel Software, 2001 y 2003).

El p -valor exacto de dos colas, está definido como

$$p_2 = \Pr [D(\mathbf{y}) \geq D(\mathbf{x})] = \sum_{D(\mathbf{y}) \geq D(\mathbf{x})} h(\mathbf{y})$$

lo cual no es más que la suma de las probabilidades hipergeométricas sobre todas las tablas $\mathbf{y} \in \Gamma$ tales que $D(\mathbf{y}) \geq D(\mathbf{x})$. Un valor pequeño de p_2 ofrece evidencia a favor de la alternativa de dos colas.

Se puede definir, además un p -valor exacto de una cola; el cual está basado en el estadístico de prueba

$$D(\mathbf{y}) = y_{11}$$

Este estadístico especifica a toda la tabla de contingencia 2×2 ; por cuanto que los marginales de fila y columna están fijos; al conocer a y_{11} quedan automáticamente definidos todos los y_{ij} restantes en la tabla de contingencia, así la probabilidad exacta de y_{11} es $h(\mathbf{y})$, además, es fácil observar que y_{11} varía desde un valor mínimo de

$$t_{\min} = \max(0, n_{+1} - m_{2+})$$

hasta un máximo de

$$t_{\max} = \min(m_{1+}, n_{+1})$$

Luego, el valor del p -valor exacto de una cola, denotado por p_1 es definido como el área a la derecha o a la izquierda de la distribución de y_{11} a partir del

valor observado x_{11} , basado en la ubicación de x_{11} en relación a $n_{+1}m_{1+}/N$, el valor medio de y_{11} ; es decir

$$\rho_1 = \begin{cases} \sum_{y_{11}=t_{\min}}^{x_{11}} h(\mathbf{y}) & \text{si } x_{11} > n_{+1}m_{1+}/N \\ \sum_{x_{11}}^{y_{11}=t_{\max}} h(\mathbf{y}) & \text{si } x_{11} \leq n_{+1}m_{1+}/N \end{cases}$$

Un valor pequeño de ρ_1 ofrecerá evidencia a favor de $H_1: \pi_2 > \pi_1$ si es calculado hacia la derecha y ofrecerá evidencia a favor de $H_1': \pi_2 < \pi_1$ si es calculado hacia la izquierda, (Cytel Software, 2001 y 2003).

Para muestras grandes las distribuciones de las tres medidas de discrepancia mencionadas convergen a la distribución χ^2 con un grado de libertad, (Cytel Software, 2001 y 2003).

En el caso de muestras grandes, el p -valor asintótico de dos colas, denotado $\tilde{\rho}_2$, está dado por

$$\tilde{\rho}_2 = \Pr[\chi_{(1)}^2 \geq D(\mathbf{x})]$$

Donde $\Pr[\chi_{(1)}^2 \geq D(\mathbf{x})]$ representa la probabilidad de que la distribución χ^2 con un grado de libertad tome un valor al menos tan grande como cualquiera de las tres medidas D citadas anteriormente para la tabla de contingencia observada \mathbf{x} .

Un caso particular muy importante, desarrollado en el anexo 3, lo constituye el estudio de S tablas de contingencia de dimensión 2×2 estratificadas, el cual por poseer diferentes fundamentos teóricos a los contemplados en este estudio y no estar considerado en los objetivos, escapa al alcance de este trabajo, pero que debido a su gran aplicación en ensayos clínicos, jurídicos y sociales el autor considera valioso mencionarlas ya que puede ser tema de futuras investigaciones, (Good, 2000; Cytel Software, 2001 y 2003; Agresti, 2002).

5.3.2. Pruebas de hipótesis exactas no condicionadas

Considérese nuevamente probar la hipótesis nula

$$H_0 : \pi_1 = \pi_2$$

contra las alternativas de una cola de la forma

$$H_1 : \pi_2 > \pi_1$$

$$H_1' : \pi_2 < \pi_1$$

y de dos colas de la forma,

$$H_1 : \pi_1 \neq \pi_2$$

El punto de partida es la tabla de contingencia de dimensión 2×2 observada, \mathbf{x} . La distribución de probabilidad no condicionada está dada por $f_0(\mathbf{x})$ en (1); para calcular un p -valor exacto es necesario definir un conjunto de referencia de las tablas de contingencia 2×2 y sumar las probabilidades de aquellas tablas que

sean tan extremas como la observada \mathbf{x} . En el caso de la inferencia condicionada, el conjunto de referencia utilizado Γ está constituido por las tablas de contingencia \mathbf{y} en las cuales tanto los totales de filas y columnas son fijados en los valores observados en la tabla \mathbf{x} . En la inferencia no condicionada se utiliza un conjunto más grande de tablas de contingencia 2×2 en las cuales solamente se fijan los totales de las columnas, es decir, los tamaños de muestra binomiales. Los totales de las filas son tratados como variables aleatorias. Si Ω denota este conjunto de referencia, entonces

$$\Omega = \{ \mathbf{y}_{2 \times 2} : y_{1j} + y_{2j} = n_{+j}; j = 1, 2 \}$$

y se ordenan las tablas de contingencia $\mathbf{y} \in \Omega$ de acuerdo al estadístico de prueba

$$D(\mathbf{y}) = \frac{\hat{\pi}_2 - \hat{\pi}_1}{\sqrt{\left(\frac{y_{11} + y_{12}}{N} \right) \left(\frac{y_{21} + y_{22}}{N} \right) \left(\frac{1}{n_{+1}} + \frac{1}{n_{+2}} \right)}} \quad (2)$$

donde $\hat{\pi}_j = y_{1j}/n_{+j}$, $j = 1, 2$. Si $y_{11} = y_{12} = 0$, se hace a $D(\mathbf{y}) = 0$. El denominador de (2) es el error estándar de la diferencia de proporciones binomiales observadas bajo la hipótesis nula. Además, el estadístico $D(\mathbf{y})$ tiene media 0 y varianza 1 bajo la hipótesis nula. Un valor extremo positivo del estadístico de prueba observado $D(\mathbf{x})$ proporciona evidencia a favor de la

alternativa H_1 y un valor extremo negativo del estadístico de prueba observado $D(\mathbf{x})$ proporciona evidencia a favor de la alternativa H_1' .

El p -valor exacto es la suma de las probabilidades de las tablas de contingencia $\mathbf{y} \in \Omega$ que son al menos tan extremas que la tabla observada \mathbf{x} con respecto al estadístico de prueba $D(\mathbf{y})$. El problema se presenta en el hecho de que cada una de estas tablas extremas tiene distribución de probabilidad $f_0(\mathbf{y})$, la cual depende del parámetro desconocido π . En la metodología condicionada se eliminó este parámetro usando su estimador suficiente, m_{1+} ; sin embargo, en este caso no es posible hacer esto, puesto que el conjunto de referencia especificado es Ω en vez de Γ . La inferencia exacta no condicionada utiliza un argumento diferente para eliminar el parámetro π . Se consideran todos los posibles valores de π en su espacio parametral y se selecciona aquel valor que produce el p -valor más grande. Esto produce una respuesta conservadora de manera tal que no importa cual pudiera ser el verdadero valor de π , el error tipo I de la prueba no puede exceder su nivel de significancia nominal. Es decir, se utiliza el principio minimax (Freund y Walpole, 1990); del parámetro π el cual selecciona el p -valor correspondiente al peor de los casos posibles para π , de manera tal que si este p -valor ofrece evidencias para rechazar la hipótesis nula, cualquier otro p -valor

correspondiente a cualquier otro valor de π también rechazará la hipótesis nula, (Cytel Software, 2001 y 2003).

Para calcular el p -valor se procede en dos etapas. En la primera etapa se expresa al p -valor como una función de π . En la segunda etapa se obtiene el supremo de esta función sobre todos los valores de $\pi \in (0, 1)$; este supremo se utiliza como p -valor.

Específicamente, el p -valor exacto no condicionado de una cola para un valor dado de π está dado por

$$p_1(\pi) = \min \left\{ \sum_{D(\mathbf{y}) \leq D(\mathbf{x})} f_0(\mathbf{y}); \sum_{D(\mathbf{y}) \geq D(\mathbf{x})} f_0(\mathbf{y}) \right\}$$

y el p -valor exacto no condicionado de dos colas para un valor dado de π está dado por

$$p_2(\pi) = \sum_{|D(\mathbf{y}) - D(\mathbf{x})| \leq \pi} f_0(\mathbf{y})$$

Finalmente se obtiene un p -valor de una o dos colas que es independiente de π tomando un supremo sobre todos los valores de π y se argumenta que aún en el peor de los casos, el verdadero p -valor nunca podría exceder al supremo, esto es

$$p_1 = \sup \{ p_1(\pi) : 0 < \pi < 1 \}$$

y

$$p_2 = \sup \{ p_2(\pi) : 0 < \pi < 1 \}$$

Esta prueba es conocida como prueba de Barnard, llamada así en honor a George Barnard, quien fue el primero en proponer una prueba de este tipo en el año de 1947 como alternativa a la propuesta por R. A. Fisher, (Cytel Software, 2001 y 2003). En el trabajo de McDonald, Davis y Milliken, (1977); se presentan tablas con las regiones críticas para la prueba no condicionada de Barnard para tamaños de muestra de hasta $n_{+1} = n_{+2} = 15$.

Resulta interesante resaltar que en el año de 1949, el mismo Barnard apoyó a las pruebas condicionadas basándose principalmente en el principio de estadísticos auxiliares mencionado anteriormente y reconociendo que Fisher estaba en lo correcto al basar las pruebas condicionadas en repeticiones hipotéticas del mismo experimento, (Cytel Software, 2001 y 2003; Agresti, 2002; Mehta y Hilton, 1993).

Una adaptación de la metodología de Barnard fue propuesta por Berger y Boos, (1994); en la cual el supremo es tomado sobre un intervalo de confianza para π en vez de tomarlo sobre su espacio parametral. Su p -valor no condicionado está dado por

$$p_1 = \sup_{\pi \in C_\gamma} \{ p_1(\pi) \} + \gamma$$

o,

$$\rho_2 = \sup_{\pi \in \mathcal{C}_\gamma} \{\rho_2(\pi)\} + \gamma$$

dependiendo de si la prueba es de una o dos colas; donde \mathcal{C}_γ es un intervalo al $100(1-\gamma)\%$ de confianza para π y γ toma un valor muy pequeño, por ejemplo $\gamma = 0,001$, (Agresti, 2002).

En los manuales del usuario de StatXact[®], (Cytel Software, 2001 y 2003); se comenta que el valor óptimo de γ es materia de futuras investigaciones, así pues, la elección del valor de γ queda por cuenta del usuario, pudiendo variar desde $0 \leq \gamma \leq \alpha$, aunque proponen como valor por defecto a $\gamma = 0,000001$.

El correspondiente p -valor asintótico, se obtiene por el hecho de que, bajo la hipótesis nula, la distribución de $D(\mathbf{y})$ converge a la distribución normal estándar, así

$$\tilde{\rho}_1 = \min\{\Phi[D(\mathbf{x})]; 1 - \Phi[D(\mathbf{x})]\}$$

para el caso de la prueba de una sola cola y

$$\tilde{\rho}_2 = 2\tilde{\rho}_1$$

para el caso de la prueba de dos colas.

Donde $\Phi[D(\mathbf{x})]$ es el valor acumulado de la distribución normal estándar para el valor observado $D(\mathbf{x})$, (Cytel Software, 2001 y 2003).

5.3.3. Comparación entre las pruebas de permutación condicionadas y no condicionadas en tablas de contingencia de dimensión 2×2

La principal ventaja de utilizar Ω es que éste es un conjunto más grande que Γ . En consecuencia la distribución de $D(\mathbf{y})$ tiene más puntos si está definida en Ω que en Γ . Esto reduce el conservacionismo de la prueba, por lo cual es posible construir pruebas de hipótesis exactas cuyos niveles de significancia reales estén más cerca de su nivel de significancia nominal bajo Ω que bajo Γ . La principal desventaja es que el parámetro π solamente puede ser eliminado considerando todas las posibilidades en su espacio parametral y colocarlo en el peor de los casos. Esto incrementa el conservacionismo de la prueba de hipótesis. Hay entonces una compensación entre la ventaja ganada enriqueciendo el conjunto de referencia y la desventaja de colocar al parámetro en el peor de los casos, (Cytel Software, 2001 y 2003).

Mehta y Hilton, (1993); mostraron que si bien el conservacionismo se disminuye notablemente utilizando a Ω en vez de Γ para el caso de las tablas 2×2 , esta ventaja se pierde rápidamente al pasar de tablas 2×2 a 2×3 e incluso en el caso de estas últimas encontraron que para valores de π cercanos a 0,5, la prueba condicionada tenía más potencia que la prueba no condicionada, por el hecho de que la prueba no condicionada toma como p -valor el correspondiente al peor escenario para el parámetro π y aunque la distribución de la prueba no

condicionada tiene un número mayor de puntos, esta prueba parece utilizar información no relevante para probar la hipótesis nula, en cambio en la prueba condicionada la información presente en los datos es la que existiría en repeticiones hipotéticas del mismo experimento, esto trae como consecuencia que el conjunto Γ sea mucho más pequeño que Ω , pero aquel utiliza información relevante para probar H_0 en vez de considerar todos los posibles escenarios.

El razonamiento de considerar repeticiones hipotéticas del mismo experimento fue introducido por R. A. Fisher y ha sido mantenido por los defensores del condicionamiento. Por otra parte, Mehta y Hilton, (1993); conjeturan que el comportamiento diferencial entre ambos enfoques pudiera disminuir en tablas de mayores dimensiones e incluso que el enfoque condicionado puede llegar a ser más potente, aunque no hay estudios que muestren esto.

Por otra parte, operativamente lo que se gana con el no condicionamiento es el hecho de tener un conjunto de referencia Ω más grande, (Cytel Software, 2001 y 2003; Agresti, 2002); lo cual representaría una ventaja en el caso de muestras pequeñas. McDonald, Davis y Milliken, (1977); afirman que en el caso de muestras pequeñas, el nivel de significación real de la prueba condicionada llega a ser hasta de un cuarto del valor del nivel de significancia nominal; sin embargo, en el enfoque condicionado el conjunto Γ crece rápidamente con la dimensión de las tablas y con el tamaño de muestra, lo cual en otros escenarios producirían un

conjunto de referencia Γ con un número muy grande de puntos el cual no adolecería de los problemas de discretitud presentes en tablas de baja dimensión y tamaños de muestra pequeños, (Cytel Software, 2001 y 2003). En el trabajo de Garside y Mack, (1976); se puede observar como el conservacionismo de la prueba de Fisher disminuye en la medida que el tamaño de las muestras binomiales aumenta, así cuando ambos tamaños son $n_{+1} = n_{+2} = 500$, los valores del nivel de significación son muy parecidos a los obtenidos con la aproximación asintótica de χ^2 , aunque ambas pruebas permanecen conservadoras, ellos encontraron también que el conservacionismo se acentuaba con el desbalanceo de las muestras.

Con el advenimiento de mayor poder computacional y el desarrollo paralelo que han tenido las pruebas de permutación y en general, las técnicas de remuestreo, Good, (2000); los softwares que se han desarrollado para datos categorizados, como el StatXact[®], permiten, al menos para el caso 2×2 , desarrollar pruebas condicionadas y no condicionadas, aunque manifiestan que en la medida que el tamaño de muestra crece, los tiempos de procesamiento son muy superiores en el enfoque no condicionado, por tener un conjunto de referencia Ω más grande, (Cytel Software, 2001 y 2003).

Una ventaja de las pruebas no condicionadas es que permiten probar hipótesis acerca de diferencias de proporciones y razones de poblaciones binomiales del tipo, $H_0 : \pi_2 - \pi_1 = \delta$ y $H_0 : \pi_2 / \pi_1 = \rho$, donde $-1 \leq \delta \leq 1$ y

$0 \leq \rho < \infty$ son los valores especificados para la diferencia y razón de los parámetros π_1 y π_2 , respectivamente; dado que en el enfoque condicional, el conjunto de referencia Γ se escogió fijándolo a sus estadísticos suficientes m_{i+} y dado que no existen estadísticos suficientes para las diferencias y razones de proporciones, este tipo de pruebas no pueden llevarse a cabo utilizando el enfoque condicional; en el software StatXact[®] 5.0.3 y 6.0, se presentan algunas pruebas de hipótesis no condicionadas para razones y diferencias de proporciones binomiales para tablas de contingencia 2×2 . Para información adicional con respecto al desarrollo y comportamiento de estas pruebas ver Cytel Software, (2001 y 2003).

En vista de que las distribuciones de los estadísticos de prueba para ambos enfoques son discretas, ambos producen pruebas conservadoras, ya que para un nivel de significancia nominal α de alguna prueba, los dos garantizan que el nivel de significancia real φ es siempre menor o igual que α ; así, ambos enfoques ofrecen protección contra el error tipo I; sin embargo, como se indicó en el caso de las tablas de contingencia 2×2 , el mayor número de puntos de la distribución de los estadísticos de prueba del enfoque no condicionado garantiza, que el nivel de significancia real φ esté más cerca del nivel de significancia nominal α , (Cytel Software, 2001 y 2003; Mehta y Hilton, 1993).

Para ilustrar este comportamiento; sean $S = 0, 1, \dots, m_{1+}, \dots, N$, las posibles realizaciones s de las marginales de los éxitos en la tabla de contingencia

\mathbf{y} , donde para cada realización s de S existirá un valor crítico o umbral $d_\alpha(s)$, entonces hay $N+1$ valores críticos al nivel de significancia nominal α para el enfoque condicional, definidos por

$$d_\alpha(s) = \min \left\{ d \in \delta_s : \Pr_{H_0} [D(\mathbf{y}) \geq d(\mathbf{y}) | s] \leq \alpha \right\}$$

donde δ_s es el conjunto formado por las distribuciones del estadístico de prueba $D(\mathbf{y})$ para cada valor s de S y $\Pr_{H_0} [D(\mathbf{y}) \geq d(\mathbf{y}) | s]$ representa la probabilidad de que el estadístico de prueba $D(\mathbf{y})$ sea al menos tan extremo que $d(\mathbf{y})$ bajo la hipótesis nula; además, el nivel de significancia real $\varphi_{c|s}$ se define como

$$\varphi_{c|s} = \Pr_{H_0} [D(\mathbf{y}) \geq d_\alpha(s) | s]$$

Se observa que para cada realización s de S existirá un valor crítico $d_\alpha(s)$ el cual garantizará que el nivel de significación real de la prueba condicionada $\varphi_{c|s}$ nunca excederá a α . Para una realización específica de $S = m_{1+}$, esta definición garantiza que las pruebas condicionadas ofrecen protección contra el error tipo I y sean conservadoras. El grado de conservacionismo dependerá de la distribución de $D(\mathbf{y})$ para la realización $S = m_{1+}$.

Para el enfoque no condicional, el nivel de significancia real de la prueba φ_U está dado por

$$\varphi_U = \sum_{s=0}^N \Pr_{H_0} [D(\mathbf{y}) \geq d_\alpha(s) | S = s] \Pr_{H_0} [S = s | \pi_0] = \sum_{s=0}^N \varphi_{C|s} \Pr_{H_0} [S = s | \pi_0]$$

donde π_0 es el valor del parámetro bajo la hipótesis nula H_0 .

Evidentemente $\varphi_U \leq \alpha$, puesto que éste es la suma de los niveles de significación reales condicionados $\varphi_{C|s}$, los cuales son menores o iguales que α , ponderados por las probabilidades de las realizaciones s de S , las cuales suman uno independientemente del verdadero valor del parámetro π_0 , (Mehta y Hilton, 1993; Cytel Software, 2001 y 2003).

5.4. Pruebas de permutación aplicadas a tablas de contingencia de dimensión $R \times C$ no ordenadas

Sea la tabla de contingencia observada \mathbf{x} de dimensión $R \times C$, mostrada en tabla 8.

Tabla 8. Esquema de una tabla de contingencia observada \mathbf{x} de dimensión $R \times C$.

<i>Respuesta</i>	<i>Columna 1</i>	<i>Columna 2</i>	...	<i>Columna c</i>	<i>Total</i>
<i>Fila 1</i>	x_{11}	x_{12}	...	x_{1c}	m_{1+}
<i>Fila 2</i>	x_{21}	x_{22}	...	x_{2c}	m_{2+}
⋮	⋮	⋮	⋮	⋮	⋮
<i>Fila r</i>	x_{r1}	x_{r2}	...	x_{rc}	m_{r+}
<i>Total</i>	n_{+1}	n_{+2}	...	n_{+c}	N

Fuente: Cytel Software, 2001 y 2003.

Donde los x_{ij} ; $i = 1, 2, \dots, r$; $j = 1, 2, \dots, c$ representan las frecuencias de las observaciones correspondientes a la categoría de la fila i y a la categoría de la columna j y cumplen que

$$\sum_{j=1}^c x_{ij} = m_{i+}; \quad \sum_{i=1}^r x_{ij} = n_{+j}; \quad \sum_{i=1}^r \sum_{j=1}^c x_{ij} = N.$$

El principal objetivo para este tipo de tablas es probar si existe independencia entre filas y columnas, lo que es equivalente a probar que no existe interacción entre filas y columnas.

Esto es, probar que

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j} \text{ para todo } i, j$$

o equivalentemente,

$$H_0 : \pi_{1j} = \pi_{2j} = \dots = \pi_{rj} \text{ para } j = 1, 2, \dots, c$$

contra la alternativa general de que esta hipótesis no se mantiene. Obsérvese que la hipótesis alternativa no se encuentra direccionada, (Good, 2000; Cytel Software, 2001 y 2003).

Para llevar a cabo esta prueba de hipótesis, la cual representa una extensión natural del caso 2×2 , se procede de la siguiente manera:

- i. Se define un conjunto de referencia de tablas de dimensión $R \times C$ en el cual cada tabla tiene una probabilidad conocida bajo el supuesto de no interacción entre filas y columnas.
- ii. Se ordenan todas estas tablas de acuerdo a una medida de discrepancia o estadístico de prueba, que cuantifique que tanto se desvía una tabla de contingencia de la hipótesis nula.
- iii. Se suman las probabilidades de todas las tablas en el conjunto de referencia que sean al menos tan discrepantes como la tabla observada \mathbf{x} .

El conjunto de referencia Γ es aquel constituido por todas las tablas de contingencia \mathbf{y} de dimensión $R \times C$ tales que los totales marginales de filas y columnas son fijos e iguales a los observados en la tabla \mathbf{x} .

Sea \mathbf{y} una tabla genérica de dimensión $R \times C$ como se presenta en la tabla 9.

Tabla 9. Esquema de una tabla de contingencia genérica \mathbf{y} de dimensión $R \times C$.

Respuesta	Columna 1	Columna 2	...	Columna c	Total
Fila 1	y_{11}	y_{12}	\dots	y_{1c}	m_{1+}
Fila 2	y_{21}	y_{22}	\dots	y_{2c}	m_{2+}
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
Fila r	y_{r1}	y_{r2}	\dots	y_{rc}	m_{r+}
Total	n_{+1}	n_{+2}	\dots	n_{+c}	N

Fuente: Cytel Software, 2001 y 2003.

Donde los y_{ij} ; $i = 1, 2, \dots, r$; $j = 1, 2, \dots, c$ representan las frecuencias de las observaciones correspondientes a la categoría de la fila i y a la categoría de la columna j y cumplen que $\sum_{j=1}^c y_{ij} = m_{i+}$; $\sum_{i=1}^r y_{ij} = n_{+j}$; $\sum_{i=1}^r \sum_{j=1}^c y_{ij} = N$.

Entonces el conjunto de referencia Γ está constituido por las tablas de contingencia \mathbf{y} de dimensión $R \times C$ tales que,

$$\Gamma = \left\{ \mathbf{y}_{r \times c} : \sum_{j=1}^c y_{ij} = m_{i+}; \sum_{i=1}^r y_{ij} = n_{+j} \right\} \text{ para todo } i, j$$

Se puede demostrar que bajo la hipótesis nula de no interacción entre filas y columnas la probabilidad de observar alguna tabla $\mathbf{y} \in \Gamma$ está dada por

$$\Pr(\mathbf{y} | \mathbf{y} \in \Gamma) = \prod_{i=1}^r \prod_{j=1}^c \frac{n_{+j}! m_{i+}!}{N! y_{ij}!}$$

la cual es una distribución hipergeométrica múltiple libre de parámetros desconocidos y que permite calcular probabilidades exactas, (Cytel Software, 2001 y 2003).

Las medidas de discrepancia o estadísticos de prueba más importantes en el estudio de las tablas de contingencia de dimensión $R \times C$ no ordenadas lo constituyen:

- El estadístico χ^2 de Pearson, el cual está dado por

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(y_{ij} - m_{i+} n_{+j} / N)^2}{m_{i+} n_{+j} / N}$$

- El estadístico de razón de verosimilitud, dado por

$$G^2(\mathbf{y}) = 2 \sum_{i=1}^r \sum_{j=1}^c y_{ij} \log \left(\frac{y_{ij}}{m_{i+} n_{+j} / N} \right)$$

- El estadístico de Freeman-Halton, el cual constituye una extensión de la prueba exacta de Fisher a tablas de contingencia $R \times C$, y está dado por

$$F(\mathbf{y}) = -2 \log [\psi \Pr(\mathbf{y} | \mathbf{y} \in \Gamma)]$$

donde ψ viene dado por:

$$\psi = (2\pi)^{(r-1)(c-1)/2} N^{-(rc-1)/2} \prod_{i=1}^r m_{i+}^{(c-1)/2} \prod_{j=1}^c n_{+j}^{(r-1)/2}$$

Las pruebas de hipótesis desarrolladas a partir de tales medidas de discrepancia son llamadas prueba de Pearson, razón de verosimilitud y Freeman-Halton, respectivamente, (Cytel Software, 2001 y 2003).

Tales medidas de discrepancia ordenan las tablas en Γ de forma diferente, las pruebas de Pearson y de la razón de verosimilitud lo hacen de acuerdo a las medidas de discrepancia χ^2 y G^2 , los cuales son una medida de cuan diferentes son los valores observados y_{ij} de sus estimadores de máxima verosimilitud $E(y_{ij}) = m_{i+}n_{+j}/N$; mientras que la prueba de Freeman-Halton lo hace de acuerdo a las probabilidades hipergeométricas $\Pr(\mathbf{y}|\mathbf{y} \in \Gamma)$ de las tablas mismas, (Agresti, 2002; Agresti y Wackerly, 1977).

Radlow y Alf, (1975); citados por Agresti y Wackerly; (1977); comentan que construir una prueba basada en una medida de discrepancia que ordene las tablas en Γ de acuerdo sólo a su probabilidad $\Pr(\mathbf{y}|\mathbf{y} \in \Gamma)$ no es del todo conveniente, puesto que pueden existir tablas \mathbf{y} con probabilidad menor a la tabla observada \mathbf{x} , pero que sin embargo sean menos discrepantes que esta última de la hipótesis nula.

El p -valor exacto de dos colas, está definido como

$$\rho_2 = \Pr [D(\mathbf{y}) \geq D(\mathbf{x})] = \sum_{D(\mathbf{y}) \geq D(\mathbf{x})} \Pr(\mathbf{y} | \mathbf{y} \in \Gamma)$$

lo cual no es más que la suma de las probabilidades sobre todas las tablas $\mathbf{y} \in \Gamma$ tales que $D(\mathbf{y}) \geq D(\mathbf{x})$. Un valor pequeño de ρ_2 ofrece evidencia a favor de la alternativa de dos colas.

Para muestras grandes las distribuciones de los tres medidas de discrepancia estudiadas convergen a la distribución χ^2 con $(r-1)(c-1)$ grados de libertad, (Cytel Software, 2001 y 2003; Agresti, 2002). Aunque Agresti, (2002); señala que el estadístico χ^2 de Pearson converge más rápidamente que el estadístico G^2 de razón de verosimilitud y que la aproximación es pobre para este último si $N/rc < 5$.

En el caso de muestras grandes, el p -valor asintótico de dos colas, denotado por $\tilde{\rho}_2$ está dado por

$$\tilde{\rho}_2 = \Pr [\chi_{(r-1)(c-1)}^2 \geq D(\mathbf{x})]$$

Donde $\Pr [\chi_{(r-1)(c-1)}^2 \geq D(\mathbf{x})]$ representa la probabilidad de que la distribución χ^2 con $(r-1)(c-1)$ grados de libertad tome un valor al menos tan grande como cualquiera de las tres medidas D estudiadas para la tabla de contingencia observada \mathbf{x} .

5.5. Aspectos computacionales en pruebas de permutación para datos categorizados

En esta sección se discuten los aspectos computacionales para llevar a cabo pruebas exactas en tablas de contingencia. Actualmente este tipo de pruebas se desarrollan bajo dos enfoques diferentes, a saber:

- Enumeración completa del conjunto de referencia Γ para el cálculo del p -valor exacto.
- Estimación del p -valor exacto mediante métodos de simulación.

5.5.1. Enumeración completa

El método de enumeración completa encuentra todas las tablas que presenten los mismos totales marginales de filas y columnas que la tabla observada, éstas constituyen el conjunto de referencia Γ . A partir de este conjunto se calcula el p -valor exacto como la suma de todas las probabilidades hipergeométricas de las tablas presentes en Γ cuyo estadístico de prueba o medida de discrepancia sea tan o más extremo que el de la tabla observada.

Inicialmente, los primeros algoritmos desarrollados para realizar pruebas exactas en tablas de contingencia aparecieron en la década de los 70 los cuales fueron escritos en lenguaje FORTRAN y básicamente realizaban la enumeración completa del conjunto de referencia Γ y a partir de ésta construían la distribución del estadístico de prueba con el fin de calcular el p -valor exacto.

Todos estos algoritmos presentaban como desventaja el intenso requerimiento computacional necesario para enumerar todo el conjunto de referencia, lo que traía como consecuencia que su uso estaba limitado a tablas de contingencia de dimensión y tamaños de muestra pequeños, (Thomas, 1971; Baptista y Pike, 1977). Muchas veces los investigadores recomendaban calcular previamente el número total de tablas de contingencia presentes en el conjunto de referencia Γ para verificar si era posible calcular el p -valor por métodos exactos en vez de utilizar aproximaciones asintóticas. Gail y Mantel, (1977); y Good, (1976); citados por Agresti, Wackerly y Boyett, (1979); proponen metodologías para el cálculo exacto y aproximado del número total de tablas del conjunto de referencia Γ . Puesto que el tamaño del conjunto de referencia Γ crece geoméricamente con la dimensión de la tabla y el tamaño de muestra, el uso de tales algoritmos estuvo sumamente limitado, (Agresti, 2002; Cytel Software, 2001 y 2003; Michaelides, 1997).

Mehta y Patel, (1980); proponen esquematizar al conjunto de referencia Γ como una estructura de red, donde todos los posibles caminos dentro de esta red representan a todas las tablas de contingencia cuyos totales marginales de filas y columnas son iguales a los de la tabla de contingencia observada. Si bien inicialmente el algoritmo basado en tal estructura no se aplicó a ningún estadístico de prueba; más tarde, Mehta y Patel, (1983); utilizaron esta estructura de red y el

algoritmo correspondiente para obtener los p -valores exactos de la prueba de Freeman-Halton sin necesidad de calcular el valor del estadístico de prueba para todas las tablas posibles, sino solamente para aquellas tan o más extremas que la observada y que contribuyen al cálculo del p -valor, por cuanto que en la estructura de red, tales tablas se pueden identificar fácilmente; por supuesto, tal innovación redujo considerablemente las necesidades de memoria y procesamiento por parte del computador y al mismo tiempo amplió el rango de aplicación de pruebas exactas a tablas de contingencia tanto de mayores dimensiones como de mayores tamaños de muestra.

Este algoritmo ha sido utilizado para calcular p -valores exactos en diferentes pruebas y esquemas de tablas de contingencia; como ejemplos se pueden citar a:

- Mehta, Patel y Gray, (1985); utilizaron el algoritmo de red con el fin de construir intervalos de confianza exactos para las *odds ratio* en tablas de contingencia de dimensión 2×2 .
- Mehta, Patel y Wei, (1988); calcularon p -valores exactos en ensayos clínicos basados en el algoritmo de red.
- Agresti, Mehta y Patel, (1990); aplicaron el algoritmo de red para calcular p -valores exactos en tablas de contingencia con categorías ordenadas.
- Mehta, Patel, Senchaudhuri y Tsiatis, (1994); aplicaron el algoritmo de red para llevar a cabo pruebas exactas en ensayos clínicos secuenciales.

- Mehta, Patel y Senchaudhuri, (1998); aplicaron el algoritmo de red para calcular p -valores exactos en tablas de contingencia 2×2 estratificadas.
- Booth, Capanu y Heigenhauser, (2003); aplicaron el algoritmo de red para el cálculo de p -valores exactos en tablas de contingencia cuadradas cuasi-simétricas para medir el acuerdo entre pares de individuos.
- El software StatXact[®] utiliza tal algoritmo de red para calcular p -valores exactos en tablas de contingencia, (Cytel Software, 2001 y 2003).

Un enfoque alternativo al algoritmo de red fue propuesto por McDonald, De Roure y Michaelides, (1999); y Michaelides, (1997); el cual consiste en enumerar las tablas del conjunto de referencia Γ mediante métodos iterativos de la manera siguiente: enumerar todas las tablas de dimensión $(R - 1) \times C$ con marginales fijas, luego enumerar las tablas de dimensión $2 \times C$ con marginales fijas y por último, enumerar las tablas 2×2 con marginales fijas.

Baglivo, Olivier y Pagano, (1988); desarrollaron un algoritmo basado en un principio diferente a los citados anteriormente; ellos proponen descomponer a la tabla de contingencia observada en dos subtablas. El p -valor de la primera subtabla se obtiene por métodos exactos y el de la segunda por métodos asintóticos, a partir de éstos se estima el p -valor de la tabla observada como el producto de los p -valores de las dos subtablas, de manera que este algoritmo híbrido requiere menos intensidad de procesamiento computacional que los otras

metodologías, lo que lo hace más expedito, aunque el p -valor obtenido es una aproximación. Ellos muestran que tal enfoque híbrido presenta una potencia similar a la obtenida por métodos exactos.

5.5.2. Simulación

Aunque el algoritmo de red de Mehta y Patel ha extendido el rango de aplicación del cálculo del p -valor exacto mediante el método de enumeración completa a tablas de mayores dimensiones y muestras más grandes; aún no es posible calcularlo cuando el tamaño de la muestra y/o la dimensión de la tabla de contingencia genera un conjunto de referencia muy grande hasta para las computadoras más potentes disponibles en la actualidad. Si se sospecha que los métodos asintóticos no son adecuados para la situación estudiada y que además no es posible calcular el p -valor por métodos exactos, una alternativa la constituye el estimar tal p -valor mediante métodos de simulación y a partir de esta estimación construir intervalos para el p -valor exacto con una amplitud y confianza deseada.

Agresti, Wackerly y Boyett, (1979); propusieron un ensayo de simulación de Monte Carlo denominado Monte Carlo clásico para tablas de contingencia. Este ensayo selecciona tablas aleatoriamente del conjunto de referencia Γ con probabilidad hipergeométrica múltiple. Bajo este esquema de simulación es necesaria una muestra aleatoria de 17000 tablas de contingencia con el objeto de generar un intervalo para el p -valor exacto con 99% de confianza y 0,01 unidades

de amplitud. En este ensayo de Monte Carlo la estimación del p -valor exacto se determina por la razón entre el número de tablas simuladas tan o más extremas que la tabla observada y el tamaño de la muestra aleatoria.

Jones, O'Gorman, Lemke y Woolson, (1989); utilizaron ensayos de Monte Carlo para estudiar la homogeneidad de K tablas de contingencia de dimensión 2×2 ; sin embargo, ellos muestrearon las tablas utilizando distribuciones de probabilidad diferentes a la hipergeométrica múltiple, tales como log-normal, uniforme, exponencial y distribuciones discretas con sólo dos valores. Tales ensayos permitieron comparar diferentes estadísticos de prueba para probar homogeneidad de razones de probabilidad binomial en tablas estratificadas.

Mehta, Patel y Senchaudhuri, (1988); propusieron un esquema de muestreo de Monte Carlo basado en el algoritmo de red, denominado muestreo por importancia. Este ensayo de simulación previamente identifica en la red a aquellos caminos que corresponden a las tablas de contingencia tan o más extremas que la tabla observada, constituyéndose así el conjunto de referencia restringido Γ^* ; posteriormente selecciona tablas de este conjunto Γ^* según su importancia para reducir la varianza del estimador del p -valor exacto. En otras palabras, este esquema de muestreo sólo selecciona tablas en Γ^* y garantiza que la varianza del estimador del p -valor sea mucho menor que la varianza del estimador del ensayo de Monte Carlo clásico. En los resultados comparativos de la aplicación del

muestreo por importancia y del ensayo de Monte Carlo clásico sobre distintas pruebas encuentran que al construir intervalos para el p -valor exacto con una confianza y amplitud dada, el tamaño de muestra requerido con el muestreo por importancia es entre 50 y 12000 veces menor que con el ensayo de Monte Carlo clásico.

Mehta, Patel y Senchaudhuri, (2000); basados en un esquema similar al del muestreo por importancia formularon un ensayo de Monte Carlo específico para la regresión logística condicionada, el cual basándose en el algoritmo de red, sólo selecciona aleatoriamente a aquellas tablas del conjunto de referencia Γ que pertenezcan al conjunto de referencia reducido Γ' , constituido por las tablas de contingencia que además de tener marginales de filas y columnas fijas presenten las mismas frecuencias conjuntas y_{ij} ; $i=1, 2, \dots, r$; $r=c$; en la diagonal principal de la tabla de contingencia observada; este conjunto de referencia reducido Γ' es mucho más pequeño que Γ , tal como lo muestran McDonald, De Roure y Michaelides, (1999). Es importante notar que este algoritmo, luego de identificar al conjunto de referencia reducido Γ' , selecciona las tablas de este conjunto con probabilidad hipergeométrica múltiple.

Senchaudhuri, Mehta y Patel, (1995); propositieron otra metodología de simulación de Monte Carlo para tablas de contingencia basada en la técnica de reducción de varianza de control de variables, obteniendo así el intervalo de

confianza para el p -valor exacto con más precisión que el ensayo de Monte Carlo clásico para un número dado de tablas. El método consiste básicamente, en generar simultáneamente dos variables aleatorias conjuntamente distribuidas y correlacionadas para utilizar la información de la covarianza de la variable control o auxiliar y reducir la varianza de la variable simulada, (Szechtman, 2003).

Las últimas tendencias de simulación para ensayos de Monte Carlo consisten en utilizar algoritmos que combinan métodos bayesianos y cadenas de Markov; esencialmente, este ensayo es un proceso iterativo que parte de una distribución *a priori* asumida inicialmente y la cual se actualiza en cada iteración hasta que la distribución muestral converge a la poblacional, (Kelly, 1999; De Martini y Rapallo, 2002). Un ejemplo de esta metodología se encuentra en Smith, Forster y McDonald, (1996); donde estimaron el p -valor exacto en tablas de dimensión 8×8 orientadas a medir acuerdo entre las opiniones de parejas casadas.

6. RESULTADOS Y DISCUSIÓN

6.1. Estudio comparativo del comportamiento de las metodologías exactas y asintóticas sobre tablas de contingencia tomadas de algunos trabajos publicados en la literatura científica internacional

En esta sección se ilustra la aplicación de las pruebas de permutación en tablas de contingencia de dimensión 2×2 y de dimensión $R \times C$ no ordenadas con su correspondiente equivalente asintótico, con la finalidad de mostrar las diferencias y similitudes entre ambos enfoques, utilizando para ello algunos ejemplos publicados en la literatura científica internacional.

6.1.1. Tablas de contingencia de dimensión 2×2

Ejemplo 1: La tabla de contingencia 10 corresponde a 76 respuestas a una pregunta específica dicotómica (sí o no) clasificadas por género (masculino y femenino) en dos cursos de una escuela primaria.

Tabla 10. Clasificación de las respuestas a una pregunta dicotómica en dos grados distintos clasificados según el género.

GÉNERO				
Femenino			Masculino	
GRADO	Sí	No	Sí	No
Primero	10	4	2	16
Cuarto	6	11	15	12

Fuente: Mielke y Berry, 1996.

En el trabajo de Mielke y Berry, (1996); se prueba que existe interacción entre el género y el grado de los estudiantes por lo que las tablas se pueden analizar según el género; por esta razón sólo se toma en cuenta a las respuestas del género femenino de la tabla 10 como ejemplo ilustrativo y cuyo tamaño de muestra $N = 31$ puede considerarse pequeño y donde además, los totales marginales de las filas son balanceados.

Los resultados de los p -valores para cada una de las pruebas aplicadas en este estudio además de su medida de similitud, se presentan en la tabla 11.

Tabla 11. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo 1.

	PRUEBAS DE SIGNIFICACIÓN			
	Fisher	χ^2	Verosimilitud	Barnard
p-valor exacto	0,07317275	0,04512570	0,04245759	0,04512570
p-valor asintótico	0,04860854	0,07317275	0,07317275	0,05381446
Δ	0,02456421	0,02804705	0,03071516	0,00868876

En la tabla 11 se observa que la prueba de Fisher presentó un p -valor exacto mayor que el p -valor asintótico, mientras que ocurrió lo contrario en las demás pruebas; por otra parte, en las pruebas condicionadas el valor de Δ fue mayor que el umbral establecido en este estudio $\Delta^* = 0,01$; en cambio, en la prueba no condicionada de Barnard se encontró que si bien el p -valor exacto fue

menor que el p -valor asintótico, el valor de Δ para este caso resultó menor que el umbral establecido.

Considerando los valores obtenidos para Δ y el umbral establecido y siguiendo la recomendación de Berger, (2000); para este ejemplo en particular, en lo concerniente a las pruebas condicionadas se deberían utilizar metodologías exactas y para el caso de la prueba no condicionada de Barnard metodologías asintóticas.

Es interesante notar que si se tomara una decisión al nivel de significación del 5% considerando los p -valores exactos y asintóticos, las decisiones dentro de cada prueba serían contrarias según se utilice una u otra metodología, siendo el caso más notable el de la prueba no condicionada de Barnard, en la cual, donde a pesar de que pueden utilizarse metodologías asintóticas, la decisión no coincide con la metodología exacta, situación que no debería ocurrir por cuanto que Δ es inferior al umbral Δ^* ; ante esta situación Berger, (2000); advierte que al utilizar a Δ como regla de decisión el nivel de significación real de la prueba puede inflarse hasta un máximo de $100(\alpha + \Delta^*)\%$; es decir, que el nivel de significación para las decisiones tomadas al 5% se encontrarán en realidad a un nivel de significación comprendido entre $0,05 \leq \alpha \leq 0,06$ por cuanto que el umbral utilizado en este estudio es de $\Delta^* = 0,01$.

Ejemplo 2: La tabla de contingencia 12 provista por Mendenhall y colaboradores, (1984); y citada por Agresti, (2002); muestra los resultados de un estudio retrospectivo en el cual se comparan los tratamientos de cirugía y terapia radioactiva, la respuesta mide el estado del paciente luego de dos años de tratamiento.

Tabla 12. Resultados de un estudio retrospectivo en el tratamiento de cáncer de laringe.

	Cáncer Controlado	Cáncer no Controlado	Total
Cirugía	21	2	23
Terapia Radiactiva	15	3	18
Total	36	5	41

Fuente: Agresti, 2002.

Los p -valores exactos y asintóticos junto con los valores Δ para las pruebas consideradas en este estudio se presentan en la tabla 13.

Tabla 13. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente para el ejemplo 2.

	PRUEBAS DE SIGNIFICACIÓN			
	Fisher	χ^2	Verosimilitud	Barnard
p -valor exacto	0,63842578	0,63842578	0,63842578	0,58935137
p -valor asintótico	0,41569295	0,43890077	0,44058417	0,43890077
Δ	0,22273283	0,19952501	0,19784161	0,15045060

La tabla de contingencia 12 es relativamente balanceada y con tamaño de muestra pequeña $N = 41$; en la tabla 13 se observa que para las cuatro pruebas aplicadas el valor de las diferencias entre los p -valores exactos y asintóticos Δ es muy superior al umbral establecido Δ^* , a pesar de que en la tabla los totales marginales de filas son balanceados; pudiera entonces pensarse que estas diferencias entre los p -valores exactos y asintóticos se deben al tamaño de muestra, el cual en este caso se puede considerar pequeño; pudiendo además conjeturarse que en la medida que el tamaño de muestra N es más pequeño estas diferencias Δ serán mayores.

Ejemplo 3: Las siguientes tablas de contingencia mostradas en la tabla 14 reportan los resultados de los despidos en una compañía, para cuatro períodos de despidos diferentes, clasificados según la edad del empleado. (Kadane, 1990).

En este estudio sólo se consideraron las tablas de contingencia correspondientes al primero y cuarto período de despidos por ser ejemplos ilustrativos e interesantes.

La tabla de contingencia correspondiente al primer período de despidos fue seleccionada por presentar un tamaño de muestra grande $N = 249$; y además es balanceada en sus totales marginales de filas. Los p -valores exactos y asintóticos junto con los valores Δ para las pruebas consideradas en este estudio se presentan en la tabla 15.

Tabla 14. Empleados despedidos y retenidos por cierta empresa para cuatro períodos de despidos clasificados según la edad.

PRIMER PERÍODO (30 / 06 / 1982)			
	Despedidos	No Despedidos	Total
≥ 40 años	18	129	147
< 40 años	0	102	102
Total	18	231	249
SEGUNDO PERÍODO (30 / 11 / 1982)			
	Despedidos	No Despedidos	Total
≥ 40 años	26	105	131
< 40 años	10	83	93
Total	36	188	224
TERCER PERÍODO (31 / 05 / 1983)			
	Despedidos	No Despedidos	Total
≥ 40 años	13	92	105
< 40 años	14	66	80
Total	27	158	185
CUARTO PERÍODO (28 / 06 / 1984)			
	Despedidos	No Despedidos	Total
≥ 40 años	13	81	94
< 40 años	2	52	54
Total	15	133	148

Fuente: Kadane, 1990.

Tabla 15. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente para el ejemplo 3, primer período.

PRUEBAS DE SIGNIFICACIÓN				
	Fisher	χ^2	Verosimilitud	Barnard
p -valor exacto	0,00006748	0,00022801	0,00004941	0,00545816
p -valor asintótico	0,00004496	0,00024331	0,00000800	0,00024331
Δ	0,00002252	0,00001530	0,00004141	0,00521485

Por otra parte en el caso de la tabla de contingencia correspondiente al cuarto período, cuyo tamaño de muestra es grande $N = 148$; y con totales marginales de fila desbalanceados, los p -valores exactos y asintóticos junto con los valores Δ para las pruebas consideradas en este estudio se presentan en la tabla 16.

Tabla 16. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente para el ejemplo 3, cuarto período.

PRUEBAS DE SIGNIFICACIÓN				
	Fisher	χ^2	Verosimilitud	Barnard
p -valor exacto	0,05326488	0,08620740	0,05326488	0,05041118
p -valor asintótico	0,04950657	0,04941467	0,03506218	0,04941467
Δ	0,00375831	0,03679273	0,01820270	0,00099651

En la tabla 15 se observa que para las cuatro pruebas aplicadas las diferencias entre los p -valores exactos y asintóticos Δ fueron mucho menores que el umbral Δ^* establecido y las decisiones a tomar al nivel de significación $\alpha = 0,05$ son iguales para las cuatro pruebas; en este caso, tal decisión sería rechazar la hipótesis nula de independencia entre la edad del empleado y su condición de retenido o despedido por la empresa; también se observa que los p -valores fueron muy pequeños para todas las pruebas y por supuesto también sus diferencias Δ ; lo que conlleva a conjeturar que en la medida que la hipótesis nula de independencia tienda a ser falsa va a ser más frecuente que Δ sea inferior al Δ^* establecido.

En la tabla 16 se observa que en dos de las pruebas: la prueba de χ^2 y de razón de verosimilitud, las diferencias entre los p -valores exactos y asintóticos fueron superiores al umbral Δ^* y que en la prueba exacta de Fisher y en la prueba no condicionada de Barnard ocurrió lo contrario; es curioso notar que para las cuatro pruebas consideradas las decisiones a tomar al nivel de significación $\alpha = 0,05$ hubieran sido contrarias si se tomaran los p -valores exactos o asintóticos.

En el caso de elegir p -valores exactos, en las cuatro pruebas no se hubiera podido rechazar la hipótesis nula de independencia y de haberse elegido los p -valores asintóticos, en las cuatro pruebas las decisiones hubieran sido exactamente las contrarias.

Si se sigue la recomendación de Berger, (2000); en la prueba exacta de Fisher y en la prueba no condicionada de Barnard se tomarían los p -valores asintóticos y se rechazaría la hipótesis nula de independencia; y para las pruebas de χ^2 y razón de verosimilitud se deberían tomar los p -valores exactos y no se podría rechazar la hipótesis nula de independencia; si bien en este trabajo se estableció como umbral de decisión a $\Delta^* = 0,01$ porque se piensa que tal valor no inflará de manera desmesurada el nivel de significación real de la prueba y se espera que el comportamiento de los métodos exactos y asintóticos sea equivalente siempre que $\Delta < \Delta^*$; por estas razones es que se sugiere el uso de los p -valores asintóticos en este caso; sin embargo, esta situación no ocurre en el presente ejemplo; es decir, siempre existe el riesgo de tomar decisiones contrarias para una prueba dada aún siendo $\Delta < \Delta^*$; por esta razón, el valor óptimo para Δ^* puede ser motivo de futuras investigaciones donde se desearía encontrar un valor para Δ^* tal que:

- (i) No infle excesivamente al tamaño de la muestra N ; es decir, que permita establecer un límite razonable para decidir entre el uso de uno u otro método.
- (ii) Que minimice el riesgo de tomar decisiones contrarias entre los métodos una vez que $\Delta < \Delta^*$.

Ejemplo 4: Berger y Boos, (1994); utilizaron la tabla de contingencia 17 y provista por Taylor *et al*, (1982); para ilustrar la modificación del procedimiento propuesto por ellos en el cálculo del p -valor exacto en la prueba no condicionada de Barnard, los autores no reportan la naturaleza de las variables medidas en tal tabla de contingencia.

Tabla 17. *Tabla de contingencia utilizada por Berger y Boos para ilustrar la modificación propuesta por ellos para el cálculo del p -valor exacto de la prueba no condicionada de Barnard.*

	Éxito	Fracaso	Total
Muestra 1	14	33	47
Muestra 2	48	235	283
Total	62	268	330

Fuente: Berger y Boos, 1994.

Los p -valores exactos y asintóticos junto con los valores Δ para las pruebas consideradas en este estudio se presentan en la tabla 18.

Tabla 18. *Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo de Berger y Boos.*

	PRUEBAS DE SIGNIFICACIÓN			
	Fisher	χ^2	Verosimilitud	Barnard
p -valor exacto	0,04441980	0, 04441983	0,06758505	0,03731378
p -valor asintótico	0,04081388	0,03710219	0,04731249	0,03710219
Δ	0,00360592	0,00731764	0,02027256	0,00021159

La tabla de contingencia 17 presenta un tamaño de muestra grande, $N = 330$ y es desbalanceada en los totales marginales de las filas; en la tabla 18 se observa que en la prueba exacta de Fisher, la prueba de χ^2 y la prueba no condicionada de Barnard, las diferencias Δ fueron inferiores al umbral Δ^* y en la prueba de razón de verosimilitud ocurrió lo contrario.

Las decisiones a tomarse al nivel de significación $\alpha = 0,05$ fueron iguales en las pruebas de Fisher, χ^2 y Barnard; es decir, si se sigue la recomendación de Berger, (2000); se utilizarían los p -valores asintóticos para tomar las decisión correspondiente, la cual en este ejemplo sería rechazar la hipótesis nula de independencia entre las categorías de filas y columnas. En este ejemplo no hay contradicción alguna entre las decisiones tomadas a partir de los p -valores exactos y asintóticos cuando $\Delta < \Delta^*$.

Para el caso de la prueba de razón de verosimilitud, la decisión a tomar al nivel de significación $\alpha = 0,05$ debe basarse en el p -valor exacto, siguiendo la recomendación de Berger, (2000); y así, para este ejemplo no se rechazaría la hipótesis nula de independencia entre las categorías de las filas y las columnas. Esta discrepancia en la decisión en relación a las otras tres pruebas pudiera deberse al desbalanceo presente en los totales marginales de las filas, lo cual pudiera afectar la sensibilidad de la prueba de la razón de verosimilitud.

Ejemplo 5: La tabla de contingencia 19 y citada por Agresti, (2002); muestra los resultados de un estudio realizado en Suecia sobre la relación entre el uso de aspirina e infarto al miocardio sobre una muestra de 1360 pacientes. La tabla muestra el número de muertes por infarto al miocardio durante los siguientes tres años de tratamiento.

Tabla 19. Resultados del estudio sueco sobre uso de aspirina e infarto al miocardio.

	Infarto	No infarto	Total
Placebo	28	656	684
Aspirina	18	658	676
Total	46	1314	1360

Fuente: Agresti, (2002).

Los resultados de los p -valores para cada una de las pruebas aplicadas en este estudio además de su medida de similitud, se presentan en la tabla 20.

Tabla 20. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al estudio sueco sobre el uso de aspirina e infarto al miocardio.

	PRUEBAS DE SIGNIFICACIÓN			
	Fisher	χ^2	Verosimilitud	Barnard
p -valor exacto	0,17678162	0,17678162	0,17678162	0,14742964
p -valor asintótico	0,14628120	0,14444339	0,14281592	0,14444339
Δ	0,03050042	0,03234772	0,03396570	0,00298625

La tabla de contingencia 19 presenta un tamaño de muestra grande, $N = 1360$; y donde si bien los totales marginales de las filas son balanceados, los totales marginales de columnas están fuertemente desbalanceados; en la tabla 20 se observa que para las tres pruebas no condicionadas: Fisher, χ^2 y razón de verosimilitud; a pesar de que el tamaño de muestra N sea tan grande, las diferencias entre los p -valores exactos y asintóticos Δ fueron superiores al umbral Δ^* ; esta falta de convergencia de los p -valores asintóticos a los exactos pudiera deberse al fuerte desbalanceo presente en los totales marginales de las columnas.

En el caso de la prueba no condicionada de Barnard, la diferencia Δ resultó inferior al umbral Δ^* ; esto pudiera deberse principalmente al hecho que el conjunto de referencia Ω para pruebas no condicionadas es por definición mucho más grande que el conjunto Γ para las pruebas condicionadas, esto trae como consecuencia que el estadístico de la prueba de Barnard presente una distribución de probabilidad con muchos más puntos; es decir, menos discreta que la correspondientes distribuciones de probabilidad de los estadísticos de prueba condicionados; así pues, la distribución del estadístico de prueba de Barnard se parecerá más a la distribución de probabilidad de su equivalente asintótico y por ende, la convergencia de este último al primero se alcanza más rápidamente.

6.1.2. Tablas de contingencia de dimensión $R \times C$ no ordenadas

Ejemplo 6: La tabla de contingencia 21 fue utilizada por Agresti, Wackerly y Boyett, (1979); como ejemplo para ilustrar el comportamiento de la prueba de Freeman-Halton al estimar el p -valor exacto utilizando al ensayo de Monte Carlo para datos categorizados introducido por ellos.

Tabla 21. Ejemplo utilizado por Agresti et al para estimar el p -valor exacto utilizando el método de ensayos de Monte Carlo.

		Variable Y			Total
		Columna 1	Columna 2	Columna 3	
Variable X	Fila 1	10	1	6	17
	Fila 2	3	5	0	8
	Fila 3	5	0	1	6
	Total	18	6	7	31

Fuente: Agresti, Wackerly y Boyett, 1979.

Los resultados de los p -valores para cada una de las pruebas aplicadas en este estudio además de su medida de similitud, se presentan en la tabla 22.

Tabla 22. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo utilizado por Agresti et al.

PRUEBAS DE SIGNIFICACIÓN			
	Freeman-Halton	χ^2	Verosimilitud
p -valor exacto	0,00983589	0,00379704	0,00670007
p -valor asintótico	0,02277185	0,00511163	0,00404239
Δ	0,01293596	0,00131459	0,00265768

La tabla de contingencia 21 de dimensión 3x3 presenta totales marginales de filas desbalanceados y tamaño de muestra pequeño, $N = 31$. En la tabla 22 se observa que las diferencias entre los p -valores exactos y asintóticos Δ en las pruebas de χ^2 y razón de verosimilitud fueron menores que el umbral Δ^* ; y en la prueba de Freeman-Halton ocurrió lo contrario. Las decisiones a tomar al nivel de significación $\alpha = 0,05$ serían las mismas en las tres pruebas aplicadas y sólo en la prueba de Freeman-Halton pudiera pensarse que el desbalanceo y el tamaño de muestra pequeño hayan afectado la convergencia de los p -valores asintóticos a los exactos.

Ejemplo 7: Everitt, (1977); citado por Davis, (1991); muestra los resultados de un estudio que trata la relación entre la ubicación (lóbulo frontal, lóbulo temporal, otros) y el tipo de tumor cerebral (maligno, benigno, otros) de 141 individuos. Los resultados se presentan en la tabla 23.

Tabla 23. Tipo de tumor cerebral y ubicación del mismo.

		Tipo de Tumor			Total
		Benigno	Maligno	Otros	
Ubicación	Lóbulo Frontal	23	9	6	38
	Lóbulo Temporal	21	4	3	28
	Otros	34	24	17	75
	Total	78	37	26	141

Fuente: Davis, 1991.

Los resultados de los p -valores para cada una de las pruebas aplicadas en este estudio además de su medida de similitud, se presentan en la tabla 24.

Tabla 24. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo utilizado por Davis.

PRUEBAS DE SIGNIFICACIÓN			
	Freeman-Halton	χ^2	Verosimilitud
p -valor exacto	0,11229250	0,09703004	0,09856013
p -valor asintótico	0,11426107	0,09745954	0,08813259
Δ	0,00196857	0,00042950	0,01042754

La tabla de contingencia 23 de dimensión 3x3 presenta totales marginales de filas moderadamente desbalanceados y tamaño de muestra grande, $N = 141$. En la tabla 24 se observa que las diferencias entre los p -valores exactos y asintóticos Δ en las pruebas de χ^2 y Freeman-Halton fueron menores que el umbral Δ^* ; y en la prueba de razón de verosimilitud ocurrió lo contrario. Las

decisiones a tomar al nivel de significación $\alpha = 0,05$ fueron las mismas en las tres pruebas aplicadas y sólo en la prueba de razón de verosimilitud se vio afectada la convergencia de los p -valores asintóticos a los exactos, esto indica que el comportamiento de las diferentes pruebas de significación varía según la naturaleza de la tabla de contingencia.

Ejemplo 8: La tabla de contingencia 25 es utilizada por Senchaudhuri, Mehta y Patel, (1995); para ilustrar la comparación de dos métodos de simulación, Monte Carlo clásico y Monte Carlo con control de variables en cuanto al tiempo de procesamiento del computador:

Tabla 25. *Ejemplo utilizado por Senchaudhuri et al para comparar al ensayo de Monte Carlo clásico con el ensayo de Monte Carlo con control de variables.*

	Col 1	Col 2	Col 3	Col 4	Col 5	Total
Fila 1	1	0	0	0	0	1
Fila 2	1	2	3	0	6	12
Fila 3	1	6	8	6	13	34
Fila 4	0	7	11	8	21	47
Total	3	15	22	14	40	94

Fuente: Senchaudhuri, Mehta y Patel, (1995).

Los resultados de los p -valores para cada una de las pruebas aplicadas en este estudio además de su medida de similitud, se presentan en la tabla 26.

Tabla 26. Resultados de los p -valores exactos y asintóticos con el Δ correspondiente al ejemplo utilizado por Senchaudhuri et al.

PRUEBAS DE SIGNIFICACIÓN			
	Freeman-Halton	χ^2	Verosimilitud
p -valor exacto	0,27565999	0,02467506	0,21212955
p -valor asintótico	0,27822365	0,00040374	0,24600379
Δ	0,00256366	0,02427132	0,03387424

La tabla de contingencia 25 de dimensión 4×5 presenta un ejemplo de una tabla desbalanceada y dispersa y tiene la característica particular de que seis de las veinte celdas de la tabla de contingencia presentan frecuencia cero y tres adicionales presentan frecuencia igual a uno.

En la tabla 26 se observa que solo la prueba de Freeman-Halton presenta un valor Δ que es menor que Δ^* ; es decir, que en las pruebas de χ^2 y de razón de verosimilitud no se presenta la convergencia deseada de los p -valores asintóticos a los p -valores exactos. Esto pudiera deberse al hecho de que la tabla presenta fuerte desbalanceo en los totales marginales de las filas y además es dispersa; estos dos factores pueden estar afectando la convergencia en las pruebas de χ^2 y de razón de verosimilitud a pesar de la gran dimensión de la tabla de contingencia.

Resulta particularmente interesante en este ejemplo observar las diferencias tan marcadas que presentaron los p -valores de una prueba a otra; así, en las pruebas de Freeman-Halton y razón de verosimilitud fueron superiores a 20%, mientras que en la prueba de χ^2 fueron inferiores a 2,5%; evidentemente las decisiones a tomar al nivel de significación $\alpha = 0,05$ en la prueba de χ^2 es contraria a las tomadas en las pruebas de Freeman-Halton y razón de verosimilitud; ya que en éstas no se rechaza la hipótesis nula de independencia y en aquella sí. En este ejemplo se evidencia el hecho de que la naturaleza de la tabla de contingencia afecta al comportamiento de las diferentes pruebas; de manera que el investigador debe ser cuidadoso al seleccionar que prueba de significación aplicará a una tabla de contingencia dada.

En la presente sección se pudo observar que las conclusiones a las cuales se llegan en el análisis de las tablas de contingencia ejemplificadas pueden ser similares o diferentes para ambos métodos dependiendo de la situación considerada, esto induce a pensar que las conclusiones obtenidas en muchos trabajos publicados pueden diferir notablemente a las generadas utilizando métodos exactos.

6.2. Estudio del comportamiento de las metodologías exactas y asintóticas a partir de la simulación de tablas de contingencia

A partir de los ejemplos discutidos en la sección anterior, se ha podido ilustrar que el comportamiento de las metodologías exactas y asintóticas en tablas de contingencia depende de los siguientes factores:

- Tamaño de la muestra.
- Balanceo de la tabla de contingencia.
- Tipo de prueba aplicada.
- Dimensión de la tabla de contingencia.

Por estas razones se realizó la simulación de tablas de contingencia con la finalidad de estudiar con más detalle este comportamiento y así recomendar el uso de una u otra metodología en función de los factores anteriores.

Para todas las pruebas consideradas, los p -valores exactos y asintóticos fueron obtenidos utilizando el software StatXact[®] versión 6.0, y para el caso de aquellas tablas de contingencia cuyo tiempo de procesamiento fue superior a los 15 minutos se estimó el p -valor exacto utilizando ensayos de Monte Carlo clásico.

Puesto que los p -valores calculados en este estudio variaron en un amplio rango y resultaron muy disímiles de una tabla a otra, se estableció como medida de similitud al módulo de la diferencia entre el p -valor exacto y el asintótico, denotado por Δ . Tal valor permite establecer una regla de decisión para utilizar

una de las dos metodologías consideradas, exacta y asintótica, como lo propone Berger, (2000); y como se mostró en la sección de materiales y métodos. Estas diferencias Δ dan lugar a 114 tablas con sus respectivos gráficos, los cuales permiten mostrar las tendencias entre las diferencias de los p -valores exactos y asintóticos y constituyen una forma de apreciar la convergencia de los p -valores asintóticos a los exactos para cada una de las pruebas estudiadas y en cada situación considerada.

Debido al gran número de tablas y gráficos generados, éstos se presentan en forma digital en el CD presente en el anexo 4; sin embargo, con fines ilustrativos, la figura 1 muestra uno de tales gráficos.

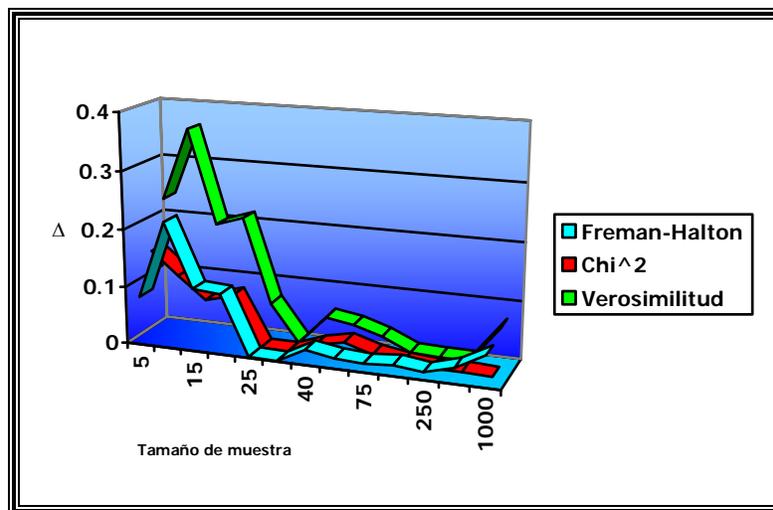


Figura 1. Valores de la medida de similitud Δ según el tamaño de muestra para las pruebas consideradas en tablas de contingencia 3×3 balanceadas y vectores de probabilidades multinomiales uniformes.

En la figura 1 puede observarse para este caso particular, que a diferencia de las pruebas de razón de verosimilitud y de Freeman-Halton, la prueba de χ^2 presenta valores inferiores al umbral $\Delta^* = 0,01$, a partir de $N = 100$; es decir, ambas metodologías presentan un comportamiento equivalente a partir de este tamaño de muestra. Para las otras dos pruebas no se alcanza tal convergencia para los tamaños de muestra considerados en la presente simulación.

Sin embargo, como resulta impráctico basar los análisis en tan gran número de tablas y gráficos, los resultados se resumen en la tabla 27 al final de la sección 6.2.1 para las tablas de contingencia de dimensión 2×2 y en las tablas 28, 29 y 30 al final de la sección 6.2.2 para las tablas de contingencia de dimensión $R \times C$ no ordenadas, donde se presentan los tamaños de muestras a partir de los cuales los p -valores asintóticos convergen a sus contrapartes exactos clasificados según la dimensión, prueba aplicada y tipo de balanceo.

6.2.1. Simulación en tablas de contingencia de dimensión 2×2

En la tabla 27 se observa que para las tablas de contingencia de dimensión 2×2 , la rata de convergencia fue mucho más rápida en todas las pruebas cuando la hipótesis nula es falsa y este comportamiento fue común tanto para las tablas balanceadas como para las desbalanceadas.

Bajo el escenario de hipótesis nula cierta, los tamaños de muestra críticos resultaron mucho más grandes que en la situación anterior, siendo éstos mayores

o iguales que 500 en las tablas desbalanceadas y mayores que 1000 en las tablas balanceadas. El comportamiento que se observa cuando la hipótesis nula es falsa puede deberse a que el conteo del número de tablas de contingencia que constituyen el p -valor tiende a ser menor en esta situación, ya que el conjunto de referencia reducido Γ^* es más pequeño que cuando la hipótesis nula es cierta; dando como resultado que estos p -valores tiendan a ser pequeños y sus diferencias Δ también, lo que permite que la rata de convergencia sea más rápida bajo este escenario por cuanto que esta diferencia se hace menor que el umbral Δ^* para tamaños de muestra menores que en el escenario de hipótesis nula cierta.

Cuando la hipótesis nula es cierta, el comportamiento observado es contrario al esperado, ya que en Cytel Software, (2001 y 2003); se afirma que cuando la tabla de contingencia es balanceada el conjunto de referencia Γ es más grande que en el caso desbalanceado y por ende, menos discreto, lo cual origina una rata de convergencia más rápida de los p -valores asintóticos a los exactos en tablas de contingencia balanceadas. Esta disimetría con la teoría puede deberse probablemente al azar.

En el caso de la prueba de Barnard, tal y como se esperaba, la rata de convergencia se alcanzó mucho más rápido en comparación a las pruebas no condicionadas en tablas de contingencia de dimensión 2×2 lo cual coincide con los argumentos de Mehta y Hilton, (1993); y Cytel Software (2001 y 2003), esto se

debe a que el conjunto de referencia Ω de las pruebas no condicionadas es por definición más grande que el conjunto de referencia Γ de las pruebas condicionadas, por lo que el estadístico de la prueba no condicionada de Barnard presenta una distribución con más puntos que la que pueden presentar los otros estadísticos para las pruebas condicionadas, resultando esto en una tasa de convergencia más rápida del estadístico asintótico a su contraparte exacto. En la tabla 27 se observa que para la prueba de Barnard, en contraposición a las pruebas condicionadas, los resultados coincidieron con la teoría, en el sentido que los tamaños de muestra críticos tienden a ser más pequeños para esta prueba; además, en esta prueba los tamaños de muestra críticos fueron menores en las tablas de contingencia balanceadas que en las tablas de contingencia desbalanceadas.

Tabla 27. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en diferentes pruebas aplicadas para tablas de contingencia de dimensión 2×2 .

Prueba	Balanceadas			Desbalanceadas		
	H ₀ Cierta		H ₀ Falsa	H ₀ Cierta		H ₀ Falsa
	Uniforme	Decreciente		Uniforme	Decreciente	
Fisher	> 1000	> 1000	250	500	1000	250
χ^2	> 1000	> 1000	250	> 1000	500	250
Verosimilitud	> 1000	> 1000	250	> 1000	500	250
Barnard	500	250	250	> 1000	500	30

6.2.2. Simulación en tablas de contingencia de dimensión $R \times C$ no ordenadas

En las tablas de contingencia de dimensión $R \times C$ no ordenadas, el comportamiento de la prueba de Freeman-Halton y de la prueba de razón de verosimilitud fue similar, por cuanto que los resultados no mostraron convergencia de los p -valores asintóticos a los exactos cuando la hipótesis nula es cierta, tanto en las tablas balanceadas como en las desbalanceadas y para el rango de muestreo considerado en este trabajo; en caso de presentarse tal convergencia ocurriría para tamaños de muestras superiores a 1000; ver tablas 28 y 30. Esta situación permite conjeturar que el conjunto de referencia Γ no es lo suficientemente grande como para mitigar la discretitud inherente en la distribución de los estadísticos utilizados para estas pruebas no condicionadas. Además, el comportamiento observado fue errático, dado que un tamaño de muestra crítico es muy grande para una configuración poblacional dada y el otro de la configuración siguiente es menor.

Cuando la hipótesis nula es falsa, si bien los tamaños críticos son mucho más pequeños que cuando la hipótesis nula es cierta y que tienden a serlo aún más en las tablas balanceadas, en este caso se observa una tendencia errática que no permite relacionar el tamaño de muestra crítico con la dimensión de la tabla de contingencia.

Una atención especial lo constituye la prueba de χ^2 para tablas de contingencia de dimensión $R \times C$ no ordenadas, cuyos resultados se muestran en la tabla 29, dicha prueba presentó una tasa de convergencia mucho más rápida que la prueba de Freeman-Halton y la prueba de razón de verosimilitud, tanto para la hipótesis nula cierta como para la hipótesis nula falsa y tanto en tablas balanceadas como en desbalanceadas. Estos resultados conducen a pensar que la prueba de χ^2 pudiera presentar una o más de las características siguientes:

- La distribución del estadístico de la prueba de χ^2 es mucho menos discreta que en la prueba de Freeman-Halton y la prueba de razón de verosimilitud.
- El estadístico de la prueba de χ^2 es más robusto que los estadísticos correspondientes a las otras dos pruebas; es decir, las decisiones tomadas a partir del p -valor asintótico son más parecidas a su equivalente exacto en la prueba de χ^2 independientemente del escenario considerado. Esto no quiere decir que el comportamiento no haya sido errático, sino que la tasa de convergencia de los p -valores fue mucho más rápida en esta prueba que en las otras dos. Esta característica deseable mostrada por la prueba χ^2 permite la recomendación de su aplicación en vez de las otras dos pruebas para tablas de contingencia de dimensión $R \times C$ no ordenadas.

Tabla 28. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en la prueba de Freeman-Halton para tablas de contingencia no ordenadas de dimensión $R \times C$.

Dimensión	Balanceadas				Desbalanceadas			
	H ₀ Cierta			H ₀ Falsa	H ₀ Cierta			H ₀ Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
2×3	> 1000	> 1000	> 1000	250	> 1000	> 1000	> 1000	1000
2×4	> 1000	> 1000	> 1000	100	> 1000	> 1000	> 1000	250
2×5	> 1000	> 1000	1000	50	> 1000	> 1000	> 1000	250
3×3	> 1000	1000	> 1000	75	> 1000	> 1000	> 1000	100
3×4	75	> 1000	> 1000	75	> 1000	1000	> 1000	250
3×5	> 1000	> 1000	> 1000	50	> 1000	> 1000	250	100
4×3	75	> 1000	> 1000	50	> 1000	> 1000	> 1000	250
4×4	75	> 1000	> 1000	75	> 1000	> 1000	> 1000	100
4×5	> 1000	> 1000	> 1000	250	> 1000	> 1000	500	250
5×3	> 1000	> 1000	> 1000	75	> 1000	> 1000	> 1000	250
5×4	> 1000	> 1000	500	250	> 1000	> 1000	> 1000	250
5×5	1000	250	250	250	> 1000	> 1000	> 1000	250

Tabla 29. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en la prueba de χ^2 para tablas de contingencia no ordenadas de dimensión $R \times C$.

Dimensión	Balanceadas				Desbalanceadas			
	H ₀ Cierta			H ₀ Falsa	H ₀ Cierta			H ₀ Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
2×3	1000	250	500	250	500	100	500	250
2×4	250	75	250	250	100	250	250	250
2×5	250	250	250	40	250	250	100	100
3×3	100	75	75	75	50	250	250	100
3×4	100	100	50	75	100	75	500	75
3×5	100	250	75	75	100	50	250	75
4×3	250	250	75	75	50	75	500	75
4×4	100	100	100	30	100	100	100	75
4×5	250	100	75	50	500	250	100	30
5×3	250	100	250	50	100	250	50	100
5×4	500	75	100	50	500	40	> 1000	50
5×5	75	1000	500	75	100	50	100	250

Tabla 30. Tamaños de muestra críticos observados para el umbral $\Delta^* = 0,01$ en la prueba de razón de verosimilitud para tablas de contingencia no ordenadas de dimensión $R \times C$.

Dimensión	Balanceadas				Desbalanceadas			
	H ₀ Cierta			H ₀ Falsa	H ₀ Cierta			H ₀ Falsa
	Uniforme	Simétrica	Decreciente		Uniforme	Simétrica	Decreciente	
2×3	1000	250	500	250	500	250	500	250
2×4	250	75	500	250	100	500	500	250
2×5	250	250	250	250	250	500	1000	250
3×3	> 1000	250	> 1000	75	> 1000	> 1000	> 1000	250
3×4	1000	> 1000	> 1000	100	> 1000	> 1000	> 1000	250
3×5	> 1000	> 1000	> 1000	75	> 1000	> 1000	> 1000	500
4×3	> 1000	> 1000	> 1000	75	> 1000	> 1000	> 1000	250
4×4	> 1000	> 1000	> 1000	75	> 1000	> 1000	> 1000	100
4×5	> 1000	> 1000	> 1000	250	> 1000	> 1000	> 1000	250
5×3	> 1000	> 1000	> 1000	250	> 1000	> 1000	> 1000	250
5×4	> 1000	> 1000	> 1000	75	> 1000	> 1000	> 1000	75
5×5	1000	1000	> 1000	250	> 1000	> 1000	> 1000	250

6.2.3. Tamaños de muestra minimax

A partir de los resultados mostrados en las tablas anteriores, se seleccionaron los tamaños de muestra minimax según la dimensión de la tabla de contingencia y balanceo de los datos, tal como se estableció en la metodología, estos resultados se presentan al final de esta sección en la tabla 31 para las tablas de contingencia de dimensión 2×2 y en la tabla 32 para las tablas de contingencia de dimensión $R \times C$ no ordenadas.

Si bien este estudio puede considerarse exploratorio, las tablas de contingencia consideradas pueden tomarse como punto de referencia para decidir cuando utilizar la metodología asintótica o la exacta.

Se hace evidente que para las pruebas de Fisher, Freeman-Halton y razón de verosimilitud es recomendable utilizar siempre métodos exactos para las tablas de contingencia de cualquier dimensión y clase de balanceo, y que los métodos asintóticos pueden aplicarse para la pruebas de Barnard en las tablas de contingencia de dimensión 2×2 y la prueba de χ^2 en tablas de contingencia de dimensión $R \times C$ a partir de los tamaños minimax determinados en este estudio.

Resulta prudente hacer notar que el comportamiento de las pruebas consideradas en este estudio ha sido desarrollado para tamaños de muestra iguales o menores a $N = 1000$; por lo cual el autor no se aventura a recomendar el uso de una u otra metodología a partir de este tamaño de muestra; es decir, las

discusiones acerca del comportamiento de tales métodos se limitan al rango de muestreo considerado en este estudio.

Tabla 31. *Tamaños de muestra minimax observados para el umbral $\Delta^* = 0,01$ para las distintas pruebas en tablas de contingencia de dimensión 2×2 .*

Prueba	N_{minimax} tablas balanceadas	N_{minimax} tablas desbalanceadas
Fisher	> 1000	> 1000
χ^2	> 1000	> 1000
Verosimilitud	> 1000	> 1000
Barnard	500	> 1000

Tabla 32. Tamaños de muestra minimax observados para el umbral $\Delta^* = 0,01$ en tablas de contingencia de dimensión $R \times C$ no ordenadas para las pruebas consideradas.

Dimensión	Prueba de Freeman-Halton		Prueba de χ^2		Prueba de razón de verosimilitud	
	N_{minimax} tablas balanceadas	N_{minimax} tablas desbalanceadas	N_{minimax} tablas balanceadas	N_{minimax} tablas desbalanceadas	N_{minimax} tablas balanceadas	N_{minimax} tablas desbalanceadas
2×3	> 1000	> 1000	1000	500	1000	500
2×4	> 1000	> 1000	250	250	500	500
2×5	> 1000	> 1000	250	250	250	1000
3×3	> 1000	> 1000	100	250	> 1000	> 1000
3×4	> 1000	> 1000	100	500	> 1000	> 1000
3×5	> 1000	> 1000	250	250	> 1000	> 1000
4×3	> 1000	> 1000	250	500	> 1000	> 1000
4×4	> 1000	> 1000	100	100	> 1000	> 1000
4×5	> 1000	> 1000	250	500	> 1000	> 1000
5×3	> 1000	> 1000	250	250	> 1000	> 1000
5×4	> 1000	> 1000	500	> 1000	> 1000	> 1000
5×5	1000	> 1000	1000	250	> 1000	> 1000

7. CONCLUSIONES Y RECOMENDACIONES

- El comportamiento de las metodologías exacta y asintótica y por ende, su elección, depende del tamaño de muestra, dimensión y balanceo de la tabla de contingencia y prueba aplicada, tal como se evidenció en los ejemplos encontrados en la literatura internacional.
- Para las pruebas condicionadas aplicadas en tablas de contingencia de dimensión 2×2 tanto en el escenario balanceado como en el desbalanceado, se observa que la tasa de convergencia es mucho más rápida y similar cuando la hipótesis nula es falsa, mientras que cuando la hipótesis nula es cierta no se pudo observar convergencia alguna debido a que los tamaños de muestra minimax superan el rango de exploración de este estudio.
- Los tamaños de muestra críticos en la prueba de Barnard resultaron ser menores que los de las pruebas condicionadas, tal y como se establece en la literatura.
- En todas las tablas de dimensión $R \times C$ no ordenadas y para todas las pruebas consideradas, el comportamiento en cuanto a la tasa de convergencia resultó ser mucho más rápido cuando la hipótesis nula es falsa que cuando ésta es cierta.
- Los tamaños de muestra críticos resultaron ser menores para la prueba de χ^2 que para las pruebas de Freeman-Halton y razón de verosimilitud, es

decir, la tasa de convergencia en la prueba de χ^2 se alcanzó más rápido que para las otras dos pruebas.

- En el caso de la prueba de razón de verosimilitud y la prueba de Freeman-Halton para tablas de contingencia de dimensión $R \times C$ no ordenadas y las tres pruebas condicionadas para tablas de contingencia de dimensión 2×2 se recomienda utilizar preferiblemente pruebas exactas, por cuanto que no se observó una tendencia en la tasa de convergencia para tamaños de muestra menores a 1000.
- En el caso de la prueba de χ^2 en las tablas de contingencia de dimensión $R \times C$ no ordenadas y la prueba no condicionada de Barnard en tablas de contingencia de dimensión 2×2 , se pudieran utilizar métodos asintóticos a partir de los tamaños minimax obtenidos en este estudio.
- Si la capacidad en el tiempo de procesamiento del computador lo permite, se sugiere aplicar primero la metodología exacta; de no ser así, aun cuando el tamaño de muestra sea grande utilizar la estimación por ensayo de Monte Carlo y en última instancia utilizar metodologías asintóticas.
- Si se desea encontrar un límite más preciso para el tamaño de muestra minimax en las pruebas anteriores, se puede ampliar el rango de exploración de la simulación en lo referente al tamaño de muestra.

- Se puede evaluar el comportamiento para diferentes valores del umbral Δ^* según la disciplina a la cual se apliquen las pruebas consideradas en este estudio.
- Realizar una simulación similar a la llevada a cabo en este estudio pero con un mayor número de tablas de contingencia simuladas.
- Considerar esquemas de balanceo diferentes y más marcados que los utilizados en este estudio.
- Debido al frecuente uso de análisis de datos categorizados en la investigación en ciencias médicas, biológicas y sociales, y como el Postgrado de Estadística de la Facultad de Agronomía de la Universidad Central de Venezuela incluye la especialización en Estadística Médica, resultaría conveniente la implementación de un curso que contemple la aplicación y el análisis de datos categorizados como un aporte importante en la formación de cuarto nivel para un especialista en el área de la estadística médica.

8. REFERENCIAS BIBIOGRÁFICAS

- AGRESTI, A. 2002. Categorical data analysis. Second edition. Wiley-Interscience. New Jersey. USA. 718 p.
- AGRESTI, A.; MEHTA, C.; PATEL, N. 1990. Exact inference for contingency tables with ordered categories. Journal of the American Statistical Association. Vol. 85, N° 410. June 1990. pp. 453-458.
- AGRESTI, A.; WACKERLY, D. 1977. Some exact conditional tests of independence for $R \times C$ cross-classification tables. Psychometrika. Vol. 42, N° 1. March 1977. pp. 111-125.
- AGRESTI, A.; WACKERLY, D.; BOYETT, J. 1979. Exact conditional tests for cross-classifications: Approximation of attained significance levels. Psychometrika. Vol. 44, N° 1. March 1979. pp. 75-83.
- BAGLIVO, J.; OLIVIER, D., PAGANO, M. 1988. Methods for the analysis of contingency tables with large and small cell counts. Journal of the American Statistical Association. Vol. 83, N° 404. December 1988. pp. 1006-1013.
- BAPTISTA, J.; PIKE, M. 1977. Algorithm AS 115: Exact two-sided confidence limits for the odds ratio in a 2×2 table. Royal Statistical Association C. Vol. 26. pp. 214-220.

- BERGER, R.; BOOS, D. 1994. *P* values maximized over a confidence set for the nuisance parameter. Journal of the American Statistical Association. Vol. 89, N° 427. September 1994. pp. 1012-1016.
- BERGER, V. 2000. Pros and cons of permutation tests in clinical trials. Statistics in medicine. Vol. 19. pp. 1319-1328.
- BOOTH, J.; CAPANU, M.; HEIGENHAUSER, L. 2003. Exact conditional ρ -value calculation for the quasi-symmetry model. University of Florida. Department of Statistics. 17 p. Tomado de: <http://www.stat.ufl.edu/~jbooth/documents/techreports/techreports.html>
Bajado de Internet el día 03/06/2004, hora 11:35 a.m.
- CYTEL SOFTWARE. 2001. StatXact 5[®]. Statistical software for exact nonparametric inference. User manual. Cytel Software. New York. USA. 1123 p.
- CYTEL SOFTWARE. 2003. StatXact 6[®] with Cytel studio. Statistical software for exact nonparametric inference. User manual. Cytel Software. New York. USA. 1398 p.
- DAVIS, C. 1991. A one degree of freedom nominal association model for testing independence in two-way contingency tables. Statistics in medicine. Vol. 10. pp. 1555-1563.

DE MARTINI, D.; RAPALLO, F. 2002. Exact conditional inference on h -sample problems for categorical data using algebraic algorithms. Tomado de:

http://staff.polito.it/giovanni.pistone/preprints_html/rapallo2.pdf

Bajado de Internet el día 01/06/2004, hora 5:30 p.m.

FREUND, J., WALPOLE, R. 1990. Estadística matemática con aplicaciones. Cuarta edición. Prentice-Hall Hispanoamericana, S.A. México. 622 p.

GAIL, M.; MANTEL, N. 1977. Counting the number of $r \times c$ contingency tables with fixed margins. Journal of the American Statistical Association. Vol. 72, N° 360. December 1977. pp. 859-862.

GARSIDE, G.; MACK, C. 1976. Actual type 1 error for various tests in the homogeneity case of the 2×2 contingency table. The American Statistician. Vol. 30, N° 1. February 1976. pp. 18-21.

GOOD, P. 2000. Permutation tests. A practical guide to resampling methods for testing hypotheses. Second edition. Springer-Verlag. New York. USA. 271 p.

HOLLANDER, M., WOLFE, D. 1999. Nonparametric statistical methods. Second edition. John Wiley & Sons INC. New York. USA. 787 p.

JONES, M.; O'GORMAN, T.; LEMKE, J.; WOOLSON, R. 1989. A Monte Carlo investigation of homogeneity tests of the odds ratio under various sample size configurations. *Biometrics*. N° 45. March 1989. pp. 171-181.

KADANE, J. 1990. A statistical analysis of adverse impact of employer decisions. *Journal of the American Statistical Association*. Vol. 85, N° 412. December 1990. pp. 925-933.

KELLY, A. 1999. Overview of computer intensive statistical inference procedures. Tomado de:

<http://www.hsrh.houston.med.va.gov/AdamKelly/resampling.html>

Bajado de Internet el día 10/02/2004, hora 10:15 p.m.

LEHMANN, E. 1986. *Testing statistical hypotheses*. 2nd. Edition. John Wiley & Sons INC. New York. USA.

LUDBROOK, J.; DUDLEY, H. 1998. Why permutations test are superior to t and F tests in biomedical research. *The American Statistician*. Vol. 52, N° 2. May 1998. pp. 127-132.

McDONALD, L.; DAVIS, B.; MILLIKEN, G. 1977. A nonrandomized unconditional test for comparing two proportions en 2×2 contingency tables. *Technometrics*. Vol. 19, N° 2. May 1977. pp. 145-157.

- McDONALD, J.; De ROURE, D.; MICHAELIDES, D. 1999. Exact tests for two-way symmetric contingency tables. *Statistics and Computing*. Vol. 8. pp. 391-399.
- MEHTA, C.; HILTON, J. 1993. Exact power of conditional and unconditional test: Going beyond the 2×2 contingency table. *The American Statistician*. Vol. 47, N° 2. May 1993. pp. 91-98.
- MEHTA, C.; PATEL, N. 1983. A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *Journal of the American Statistical Association*. Vol. 78, N° 382. June 1983. pp. 427-1983.
- MEHTA, C.; PATEL, N. 1980. A network algorithm for the exact treatment of the $2 \times k$ contingency table. *Communications in Statistics B. Simulation. Computation*. Vol. 9. pp. 649-664.
- MEHTA, C.; PATEL, N. 1997. Exact inference for categorical data. 39 p. Tomado de <http://www.cytel.com/papers/sxpaper.pdf>. Bajado de Internet el día 27/12/2003, hora 2:03 p.m.
- MEHTA, C.; PATEL, N.; GRAY, R. 1985. Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables. *Journal of the American Statistical Association*. Vol. 80, N° 382. December 1985. pp. 969-973.

MEHTA, C.; PATEL, N.; SENCHAUDHURI, P. 1998. Approximately exact inference for the common odds ratio in several 2×2 tables. Comment. Journal of the American Statistical Association. Vol. 93, N° 444. December 1998. pp. 1313-1316.

MEHTA, C.; PATEL, N.; SENCHAUDHURI, P. 2000. Efficient Monte Carlo methods for conditional logistic regression. Journal of the American Statistical Association. Vol. 95, N° 449. March 2000. pp. 99-108.

MEHTA, C.; PATEL, N.; SENCHAUDHURI, P. 1988. Importance sampling for estimating exact probabilities in permutational inference. Journal of the American Statistical Association. Vol. 83, N° 404. December 1988. pp. 999-1005.

MEHTA, C.; PATEL, N.; SENCHAUDHURI, P.; TSIATIS, A. 1994. Exact permutational tests for group sequential clinical trials. Biometrics. Vol. 50. December 1994. pp. 1042-1053.

MEHTA, C.; WALSH, S. 1992. Comparison of exact, mid- p , and Mantel-Haenszel confidence interval for the common odds ratio across several 2×2 contingency tables. The American Statistician. Vol. 46, N° 2. May 1992. pp. 146-150.

MEHTA, C.; PATEL, N.; WEI, L. 1988. Constructing exact significance tests with restricted randomization rules. *Biometrika*. Vol. 75, N° 2. pp. 295-302.

MICHAELIDES, D. 1997. Exact tests via complete enumeration: a distributed computing approach. University of Southampton. Faculty of Social Sciences. Social Statistics. Doctoral Thesis. 179 p. Tomado de: <http://eprints.ecs.soton.ac.uk/archive/00000749/03/thesis.pdf>.

Bajado de Internet el día 18/05/2004, hora 4:20 p.m.

MIELKE, P.; BERRY, K. 1996. Exact probabilities for first-order and second-order interactions in $2 \times 2 \times 2$ contingency tables. *Educational and Psychological Measurement*. Vol. 56, N° 5. October 1996. pp. 843-847.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. 2002. Dataplot Version 11/2002. Washington D.C. USA. Tomado de:

<http://www.itl.nist.gov/div898/software/dataplot/>

Bajado de Internet el día 25/07/2004, hora 10:30 a.m.

SENCHAUDHURI, P.; MEHTA, C.; PATEL, N. 1995. Estimating exact p -values by the method of control variates, or Monte Carlo rescue. *Journal of the American Statistical Association*. Vol. 90, N° 430. pp. 640-648.

SMITH, P.; FORSTER, J.; McDONALD, J. 1996. Monte Carlo exact tests for square contingency tables. Journal of the Royal Statistical Society A. Vol. 159, Part 2. pp. 309-321.

SZECHTMAN, R. 2003. Control variates techniques for Monte Carlo simulation. Proceedings of the 2003 Winter Simulation Conference. Tomado de: <http://www.informs-cs.org/wsc03papers/017.pdf> Bajado de Internet el día 01/06/2004, hora 5:15 p.m.

THOMAS, D. 1971. Algorithm AS36: Exact confidence limits for the odds ratio in a 2×2 table. Applied Statistics. Vol. 20, N° 1. pp. 105-110.

ZELEN, M. 1971. The analysis of several 2×2 contingency tables. Biometrika. Vol. 58, N° 1. pp. 129-137.

9. ANEXOS

Anexo 1. *Comandos para generar muestras binomiales independientes utilizando el software Dataplot® versión 11/2002.*

```
let p = data 0.5 0.5
let n = 4
let nevents = 10
seed = 81652272
let m = multinomial random numbers p n nevents
set write decimals 0
write D:\Docume~1\America\prog\81652272\result\U_4.res m
quit
```

Anexo 2. *Comandos para generar muestras multinomiales independientes utilizando el software Dataplot® versión 11/2002.*

```
let p = data 0.4 0.3 0.15 0.1 0.05
let n = 52
let nevents = 10
seed = 12842189
let m = multinomial random numbers p n nevents
set write decimals 0
write D:\Docume~1\America\prog\12842189\result\D_52.res m
quit
```

Anexo 3. Tablas de contingencia estratificadas de dimensión $2 \times 2 \times S$

Un caso particular muy interesante en el estudio de las tablas de contingencia de dimensión 2×2 lo constituye el probar la hipótesis de que el comportamiento de los parámetros de ambas poblaciones binomiales es el mismo a través de S estratos o de S tablas de contingencia medidas para alguna variable concomitante; es decir, probar que no existe interacción entre las tablas de contingencia 2×2 y los S estratos.

Sea $\phi_k = \left(\frac{\pi_{11k}}{1 - \pi_{11k}} \right) / \left(\frac{\pi_{12k}}{1 - \pi_{12k}} \right)$, la *odds ratio*¹ de la tabla k ,

$k = 1, 2, \dots, s$; entonces lo que interesa es derivar una prueba exacta para la hipótesis nula

$$H_0 : \phi_1 = \phi_2 = \dots = \phi_s$$

contra la hipótesis alternativa de que esta igualdad no se mantenga para al menos un par de estratos. Si la hipótesis nula se mantiene, entonces el comportamiento de los parámetros binomiales es el mismo a través de los s estratos puesto que sus cocientes de razones de probabilidad son homogéneos, lo cual implica que no

¹ Las *odds ratio* constituyen, en el caso de las inferencias exactas condicionadas para las tablas de contingencia de dimensión 2×2 una medida muy importante de la homogeneidad de los parámetros binomiales, dicha medida permite elaborar intervalos de confianza exactos para las *odds ratio*. Para un estudio detallado de los aspectos generales sobre este tema ver Agresti, (2002); y Good, (2000); y para un estudio detallado de las inferencias exactas y asintóticas aplicadas tanto a una tabla de contingencia 2×2 como a s tablas de contingencia 2×2 ver Cytel Software, (2001); Baptista y Pike, (1977); Mehta, Patel y Gray, (1985); Jones, O'Gorman, Lemke y Woolson, (1989); Mehta y Walsh, (1992); Mehta, Patel y Senchaudhuri, (1998).

existe interacción entre las tablas de contingencia 2×2 y los estratos. Tal prueba, desarrollada por Zelen, (1971); se conoce como prueba de homogeneidad de razones de cocientes de probabilidad. Para ilustrar el comportamiento de esta prueba, sea el conjunto de referencia Υ constituido por las tablas $\mathbf{y} \in \Upsilon$, denominadas τ , tales que

$$\Upsilon = \left\{ \tau : \mathbf{y}_{2 \times 2 \times s} : y_{i1k} + y_{i2k} = m_{i+k}; y_{1jk} + y_{2jk} = n_{+jk}; i, j = 1, 2; k = 1, 2, \dots, s \right\}$$

además, sea Υ_τ , el conjunto de referencia más restringido, tal que

$$\Upsilon_\tau = \left\{ \tau \in \Upsilon : y_{111} + y_{112} + \dots + y_{11s} = t \right\}$$

Zelen, (1971); demuestra que la probabilidad de observar algún arreglo de tablas τ específico, condicionado al conjunto de referencia Υ_τ no depende de los parámetros desconocidos ϕ_k ; $k = 1, 2, \dots, s$; y se puede expresar como

$$\Pr(\tau|t) = \frac{\prod_{k=1}^s \binom{n_{+1k}}{y_{11k}} \binom{n_{+2k}}{y_{12k}} / \binom{N_k}{m_{1+k}}}{\sum_{\tau \in \Upsilon_\tau} \left[\prod_{k=1}^s \binom{n_{+1k}}{y_{11k}} \binom{n_{+2k}}{y_{12k}} / \binom{N_k}{m_{1+k}} \right]}$$

Sea τ_0 la realización de τ observada, y sea Υ_τ^* el conjunto definido por

$$\Upsilon_\tau^* = \left\{ \tau \in \Upsilon_\tau : \Pr(\tau|t) \leq \Pr(\tau_0|t) \right\}$$

entonces, este conjunto define la región crítica de la prueba.

Luego, el p -valor exacto para esta prueba está dado por

$$p_2 = \sum_{\tau \in Y_\tau^*} \Pr(\tau|t)$$

y se rechazará H_0 para valores pequeños de p_2 .

El trabajo de Zelen presenta una distribución asintótica para este estadístico, pero desafortunadamente, tal aproximación es incorrecta; más tarde, Breslow y Day, (1980); citados por Cytel Software, (2001 y 2003); desarrollaron una distribución asintótica para esta prueba; su propuesta considera que para muestras grandes, el estadístico

$$\chi_{BD}^2 = \sum_{k=1}^s \frac{[X_{11k} - A_k(\hat{\phi})]^2}{\text{Var}(X_{11k}|\hat{\phi})}$$

tiene distribución χ^2 con $(s-1)$ grados de libertad.

Donde $A_k(\hat{\phi})$ es la raíz cuadrada positiva de la solución de la ecuación cuadrática

$$\frac{A_k(N_k - m_{1+k} - n_{+1k} + A_k)}{(m_{1+k} - A_k)(n_{+1k} - A_k)} = \hat{\phi}$$

y $\hat{\phi}$ está dado por

$$\hat{\phi} = \frac{\sum_{k=1}^s x_{11k} (N_k - m_{1+k} - n_{+1k} + x_{11k}) / N_k}{\sum_{k=1}^s (m_{1+k} - x_{11k})(n_{+1k} - x_{11k})}$$

Por otra parte, la varianza estimada de $X_{11k}|\hat{\phi}$ es

$$\text{Var}(X_{11k}|\hat{\phi}) = \left[\frac{1}{A_k(\hat{\phi})} + \frac{1}{m_{1+k} - A_k(\hat{\phi})} + \frac{1}{n_{+1k} - A_k(\hat{\phi})} + \frac{1}{N_k - m_{1+k} - n_{+1k} + A_k(\hat{\phi})} \right]^{-1}$$

Así, en el caso de muestras grandes, el p -valor asintótico, denotado por \tilde{p}_2 está dado por

$$\tilde{p}_2 = \Pr \left\{ \chi_{(s-1)}^2 \geq \sum_{k=1}^s \frac{[x_{11k} - A_k(\hat{\phi})]^2}{\text{Var}(X_{11k}|\hat{\phi})} \right\}$$

Donde $\tilde{p}_2 = \Pr \left\{ \chi_{(s-1)}^2 \geq \sum_{k=1}^s \frac{[x_{11k} - A_k(\hat{\phi})]^2}{\text{Var}(X_{11k}|\hat{\phi})} \right\}$ representa la probabilidad de

que la distribución χ^2 con $(s-1)$ grados de libertad tome un valor al menos tan grande como el estadístico de Breslow-Day calculado sobre las tablas $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s$ observadas.

Anexo 4. *Valores de la medida de similitud Δ según el tamaño de muestra para las pruebas consideradas en tablas de contingencia de dimensión 2×2 y $R \times C$ no ordenadas.*